A dissection of Kekkon5 and its role in mediating epithelial junction architecture

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Biology and Biotechnology

By

Christina L Ernst

APPROVED:

Joseph Duffy, PhD Major Advisor WPI	Tanja Dominko, DVM, PhD Committee Member WPI	Eric Overström, PhD Committee Member WPI

Table of Contents

Figures

Tables

Acknowledgements

Without the love and support of my family, friends, and mentors, the past two years would never have been the incredible experience that it was.

To my fiancé, *Scott Kolodzy* (and the rest of our furry family), for being there for me every step of the way. You've kept me sane (and often entertained) through the past several years, and I'd be lost without your unfaltering love and support.

To my advisor, *Joe Duffy*, for pushing me to grow, as both a person and a scientist. The journey certainly wasn't easy, but seeing where I am now, I realize it's been exactly what I've needed all along.

To my Duffy Lab colleagues and friends, *Harita Haridas, Prachi Gupta, Edie Plada, Michelle Arata, Debbie Afezolli,* and *Annie Rankova*, for making the most of our time in and out of the lab. The past two years simply wouldn't have been possible without you all, and I hope that we stay in touch throughout the years.

To my committee, *Tanja Dominko* and *Eric Overström*, for providing both support and criticism, and encouraging me to think beyond my own project.

To those I have worked for as a TA, for the experiences and perspectives I've gained as both teacher and student.

To the BBT faculty, for teaching and supporting me throughout my undergraduate and graduate studies.

To the BBT administrative assistants past and present, *Carol Butler, Eileen Dagostino,* and *Alex Rivera*, for singlehandedly holding the department together.

To my parents, for being unconditionally supportive all these years. It means so much to know how proud I've made you.

Abstract

The acquisition of cellular adhesion machinery likely represented a key factor in the evolutionary transition from unicellular to multicellular organisms. Within metazoa, cellular adhesion is an integral aspect of organismal integrity through its regulation of a wide range of processes, including tissue patterning, cellular proliferation, and migration. As such, dysregulation of adhesion has been linked to diverse pathologies including cancers and neurodegenerative diseases. At the molecular level, adhesion is mediated by specific transmembrane cell adhesion molecules (CAMs) and intracellular complexes that create a dynamic link between the extracellular milieu and the intracellular cytoskeleton. At the sequence level, immunoglobulin domains act to mediate homo- and heterophilic interactions among CAMs and thus adhesion between neighboring cells. LIGs, a family of Ig-containing proteins that contain Leucine-rich repeats, represent candidates for novel CAMs with functions in axonal regeneration and synaptic pathfinding – all of which are highly dependent on cellular adhesion.

In *Drosophila*, two LIG family members, Kekkon1 (Kek1) and Kekkon5 (Kek5) have been been implicated in EGF signaling, and Bone Morphogenetic Protein signaling as well as cellular adhesion, respectively. To investigate the putative role of Kek5 as a CAM, characterization of Kek5 activity was carried out at the cellular and molecular level. From this it was discovered that Kek5 is able to induce a dramatic upregulation of the adherens junction component Armadillo, in addition to epithelial extrusion and cell enlargement. Together, the studies presented within support a model in which Kek5 acts in a homophilic fashion to upregulate Arm and that this activity is functionally separable from other observed effects (epithelial extrusion and cell enlargement).

Introduction

Cellular adhesion is far more than just the 'glue' that holds our cells together. In addition to serving as the physical force that attaches cells together, cellular junctions allow intercellular communication and regulate polarity, growth, proliferation, and migration, all of which are crucial developmental processes. Many types of molecules – intracellular, transmembrane, and secreted alike – participate in junction formation and maintenance, forming dynamic complexes that modulate both physical adhesion as well as essential regulatory signaling. The absence or misregulation of these crucial molecules may lead to devastating diseases, including cancers and degenerative disorders; thus, a deeper understanding of their roles and mechanisms will lead to more effective therapeutics and potentially even cures for countless ailments.

The LIG superfamily

LIGs, or Leucine-rich repeat (LRR) and Immunoglobulin (Ig) domain-containing proteins are a small, novel class of transmembrane proteins, likely to be involved in crucial developmental processes and signaling pathways. While its core components, Leucine-rich repeats and Immunoglobulin domains, are two of the most common sequence elements in the metazoan proteome (343 LRR molecules in humans and 117 in *Drosophila*; 894 Ig molecules in humans and 314 in *Drosophila*; UniProtKB), there are very few molecules containing both, with only 36 in humans (Appendix A) and 9 in *Drosophila*. The little that is known about this family is intriguing, with suggested roles in nervous system development and maintenance, as well as anti-oncogenic abilities, warranting further investigation.

Leucine-rich repeats

Found in thousands of proteins across the prokaryotic and eukaryotic proteomes, leucine-rich repeats (LRRs) are conserved 20-30 amino acid motifs that are rich in leucine or other hydrophobic residues. The most highly conserved portion of each repeat is a stretch of 11 residues with the core consensus sequence LxxLxLxxN/CxL, followed by a more variable region (Wei et al., 2008). These repeats, each containing a single β sheet and α helix, are connected in tandem, creating a horseshoe-like tertiary structure that often functions as a binding pocket for various ligands (Fig 1). Cysteine-rich regions, thought to help protect the exposed terminal repeats' hydrophobic cores, often flank extracellular LRRs. In most cases, LRRs facilitate protein-protein interactions through their horseshoe-like interface, and function in a wide variety of roles, including innate immunity, synaptic targeting, and extracellular matrix (ECM) architecture (Kobe and Kajava, 2001). LRR mutations have also been demonstrated to play a role in diseases and disorders such as schizophrenia, epilepsy, night blindness, and ovarian dysgenesis (Matsushima et al., 2005).

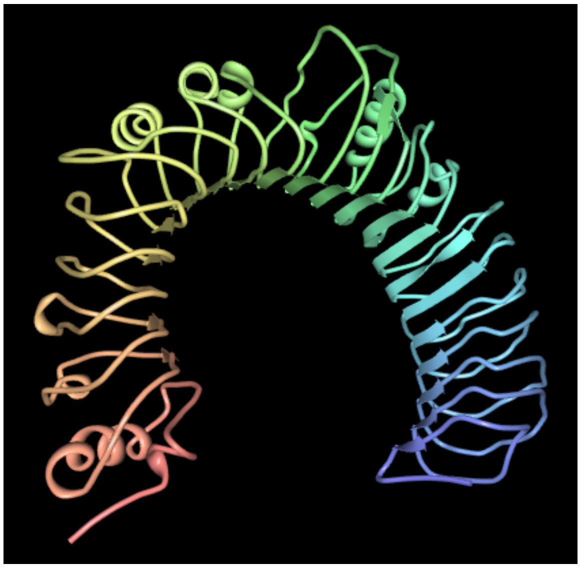


Figure 1: Leucine-rich repeats of Toll-like receptor 2. Each LRR contains a single α-helix and β-sheet, arranged in tandem in a horseshoe-like structure. (PDB ID: 3A7C)

One crucial role of leucine-rich repeat-containing proteins is pathogen recognition. LRRs are found in Toll-like receptors (TLRs) and NOD-like receptors (NLRs), which are transmembrane and intracellular pattern recognition receptors, respectively, composing a large part of our first line of defense against pathogens (Palsson-McDermott and O'Neill, 2007). Additionally, plants contain nucleotide-binding LRRs (NB-LRRs), hybrids of mammalian TLRs and NRLs that "recognize specific pathogen effector proteins," rather than the traditional pattern-based recognition (Padmanabhan et al., 2009). Interestingly, bacteria and parasites have also been found to contain LRR proteins as virulence factors that facilitate cellular attachment and invasion (Kedzierski et al., 2004). As such, it appears that LRRs have maintained a conserved function of ligand recognition and binding.

LRR proteins have also been implicated in synaptic targeting and cellular adhesion. There are several known families of synaptic adhesion molecules, including netrins, Netrin-G ligands (NGLs) and synaptic adhesion-like molecules (SALMs). These protein families are expressed primarily in the brain, and have several roles regarding nervous system development, especially synapse formation and differentiation. Although not yet determined, they may also play a role in adhesion, as there are several isoforms within each family, some of which are expressed in both dendrites and axons (Ko and Kim, 2007). Additionally, fibronectin-leucine-rich transmembrane (FLRT) proteins induce cell sorting/clustering when overexpressed both in cell culture and *Xenopus* embryos, an effect demonstrated to be dependent upon the extracellular leucine-rich repeats. In support of a role in cell sorting and adhesion, FLRTs have also been shown to directly interact with themselves and each other (Karaulanov et al., 2006). Further examination of these and other LRR proteins in the context of binding specificity and adhesion may help further understand the overall role of LIGs and neuronal LRRs.

Immunoglobulin domains

Figure 2: Ig domain 1 of Robo. Each Ig domain consists of a conserved core of four β-sheets surrounded by another 3-5 variable β-sheets. (PDB ID: 2V9T)

Another of the most common sequence elements in proteins is the Immunoglobulin (Ig) domain, most commonly associated with immune response molecules. Ig domains are comprised of 7-9 β sheets in the form of a conserved β sandwich core surrounded by several variable strands (Fig 2) (Bork et al., 1994). While Ig domains, like LRRs, are involved with innate immune response proteins, they also constitute a large part of the adaptive immune system, evidenced by their presence in antibodies and T-cell receptors. Particularly well suited to this role, Ig domains exhibit extensive variability and highly specific binding partners within the family (Palsson-McDermott and O'Neill, 2007).

Immunoglobulin domains are also quite prevalent in cellular adhesion molecules (CAMs). In such roles, the Ig domain often serves as a binding interface for molecules on adjacent cells, and, to a lesser extent, for molecules on the same cell, and can impart varying degrees of specificity. A number of Ig-containing CAMs are found in the nervous system, and are involved in morphogenesis, neurite outgrowth, cell migration, and proper development of the central nervous system (Crossin and Krushel, 2000). One fascinating example of the specificity of Ig domains is the *Drosophila* neural CAM Down's syndrome cell adhesion molecule (Dscam). Containing three variable Ig domains, the Dscam transcript can be spliced into over 19,000 ectodomains. Interactions between isoforms are highly specific, exhibiting almost exclusively homophilic binding, requiring all three variable Ig domains to match. This extensive specificity allows for neuronal self-avoidance and complex tiling arrangements with the variable splicing of a single transmembrane protein (Wojtowicz et al., 2007). As such, it is quite probable that the Ig domains of the LIG family play a key role in the specificity of protein interactions, cellular adhesion, and signaling.

LIG subfamilies

Although the LIG family is quite small, with only 36 members in humans, relatively little is known about what these proteins do. To date, a few members, including the Trk receptors, LRIG1, and LINGO1 have been investigated (Fig 3), demonstrating various roles in nervous system development and maintenance, as well as putatively functioning as proto-oncogenes (Chen et al., 2006). One large-scale analysis of LIG expression has been undertaken in mouse embryos, demonstrating a variety of expression patterns, including specific neuronal and non-neuronal profiles for varying subfamilies (Homma et al., 2009). Although this does not give significant insight into the roles of specific LIGs, it may aid in the challenge of identifying ligands and other molecular interactors.

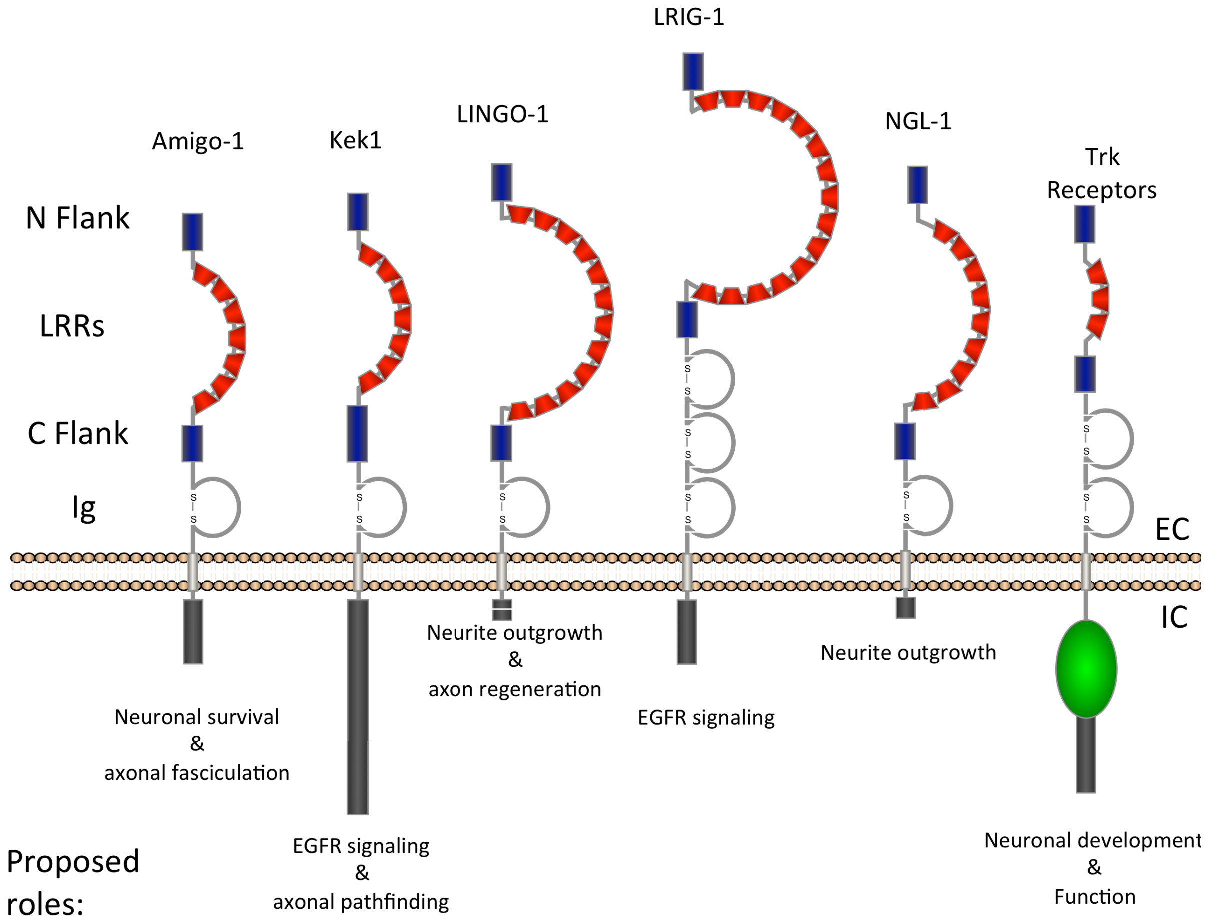


Figure 3: Representative members of vertebrate and *Drosophila* LIG subfamilies. Each LIG is a transmembrane molecule containing extracellular leucine-rich repeats and one or more immunoglobulin domains, as well as a short non-catalytic (with the exception of the Trk receptors) intracellular domain.

Kekkons

The Kekkon (Kek) family in *Drosophila* consists of six members, each of which contains seven LRRs and a single Ig domain (Fig 4). The founding member, Kek1, has been demonstrated to inhibit the Epidermal Growth Factor Receptor (EGFR) through direct binding (Alvarado et al., 2004b; Ghiglione et al., 2003; Ghiglione et al., 1999). The mechanistic details of this interaction have yet to be fully elucidated, as the LRRs have been demonstrated to be sufficient for binding *in vitro*, but the highly conserved transmembrane domain is necessary for inhibition *in vivo*. Inhibition of EGFR is highly specific to Kek1, leaving the role(s) of the remainder of the family in question (Alvarado et al., 2004a). However, an analysis of Keks 1, 2, 5, and 6 demonstrate both homo- and heterophilic associations *in vitro*, reminiscent of cell adhesion molecules. Also of note is the fact that five of the six Kek molecules each contain a C-terminal PDZ domain-binding site (MacLaren et al., 2004). PDZ domains are short conserved motifs found throughout the metazoan proteome that often confer protein-protein interactions and subcellular localization. At least three different types of PDZ domains exist, each with its own respective consensus binding site (Table 1) (Hung and Sheng, 2002). PDZ domains and/or binding sites are found in many cell adhesion and junction-related molecules, furthering the possibility of the Kek family's involvement in cellular adhesion.

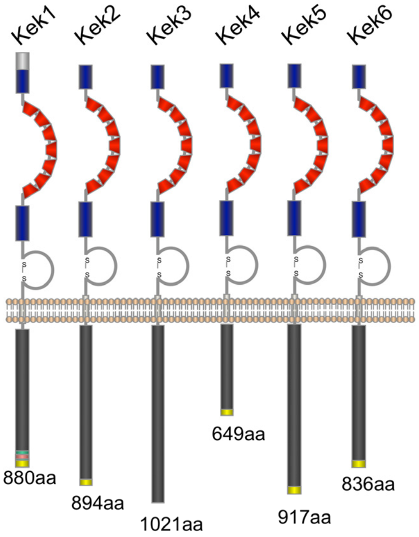


Figure 4: The *Drosophila* Kekkon (Kek) family. Each of the six members contains seven LRRs followed by a single Ig domain. All members except Kek3 contain a C-terminal PDZ domain-binding site.

Table 1: C-terminal PDZ domain-binding site consensus sequences. 'X' indicates any residue; 'Φ' indicates a hydrophobic residue (Hung and Sheng, 2002).

Class	Consensus
Type I	-X-S/T-X-Φ
Type II	-X-Φ-X-Φ
Type III	-X-D/E-X-Φ

Trk receptors

Highly expressed in the nervous system, the Tropomyosin-related kinase (Trk) receptor family is one of the most well characterized LIG subfamilies (Huang and Reichardt, 2003). Each of its three members (TrkA, B, C) contains three LRRs and two Ig domains, as well as an intracellular tyrosine kinase domain, making them the only LIGs to have an identifiable intracellular catalytic domain (Fig 3). These receptors facilitate neurotrophin signaling and respond to at least four ligands: Nerve Growth Factor (NGF), Brain-Derived Neurotrophic Factor (BDNF), and Neurotrophins 3 and 4. Upon ligand binding, most of which occurs through the Ig domains, the receptors dimerize and activate via autophosphorylation. Downstream signaling through a kinase cascade regulates several crucial developmental processes, including neurite outgrowth, cytoskeleton modeling, and synaptic targeting and plasticity (Huang and Reichardt, 2003). The Trk receptors are also known to be oncogenes, as mutation of the Ig domains can cause constitutive activation and thus misregulation of crucial processes, including differentiation and proliferation (Nakagawara, 2001).

LRIGs

An additional LIG family that has been somewhat well studied is the Leucine-rich repeats and Immunoglobulin-like domains (LRIG) family. Each of its three members contains fifteen LRRs and three Ig domains (Fig 3). Although LRIG1's function is similar to that of Kek1 in inhibiting the EGFR (or ErbB) family, it is less direct; as opposed to direct binding and signal inhibition, it binds ErbBs and recruits ubiquitin ligase E3, resulting in subsequent degradation (Rubin et al., 2005). Dysregulation of LRIG1 has been shown to result in several epithelial cancers, including breast, cervix, colon, and lung cancers. It has also been suggested that LRIG1, expressed in epidermal follicle cells "corresponding to the foci of epidermal stem cells," may be a negative regulator of stem cell proliferation (Hedman and Henriksson, 2007). Although it has been suggested that LRIGs contribute to the dysregulation of EGF signaling in cancers, a more rigorous investigation of binding partners would prove beneficial in elucidating these and broader roles of the family.

LINGOs

Another intriguing LIG subfamily is the LINGO family, the four members of which each contains 12 LRRs and a single Ig domain (Fig 3). LINGO1 functions as a part of the Nogo Receptor (NgR1)-p75 Neurotrophin Receptor ($p75^{NTR}$) tripartite signaling complex (Mi et al., 2004), activation of which results in downstream RhoA signaling (Mi et al., 2005), inhibiting axonal regeneration upon injury (Mi et al., 2007). Recent studies have highlighted LINGO1 as a potential therapeutic target for a number of nervous system disorders and ailments. Blocking LINGO1 function with a soluble variant results in axonal regeneration and recovery from paralysis after spinal cord injury in rats. Knocking down LINGO1 in Parkinson's disease models results in increased survival of dopaminergic neurons (Inoue et al., 2007), as well as oligodendrocyte remyelination in Multiple Sclerosis models (Mi et al., 2007). With these recent developments, there appears to be significant promise for therapeutics for disorders for which there has only been symptom relief, rather than a cure.

Cellular adhesion and epithelial junctions

Serving far more roles than simply holding cells together, cellular adhesion is an absolutely crucial aspect of organismal development. These intercellular junctions are some of the most conserved cellular structures and are responsible for cell polarity, shape, paracellular barrier establishment, and intracellular signaling (Franke, 2009). Without appropriate junctions established, development will likely proceed abnormally, and even minor junctional defects can later result in diseases such as cancer, autoimmune diseases, and cell barrier and permeability disorders.

Junction architecture

Epithelial adhesion requires several different types of junctions, some of which vary between vertebrates and invertebrates (Fig 5). One of the most conserved junction types is the adherens junction (AJ), generally the first to form upon cell contact. Contact is initiated by intercellular homotypic binding by transmembrane proteins known as nectins and cadherins, components of integral adhesive protein complexes at the AJ (Cavey and Lecuit, 2009). Localized apically in both vertebrate and invertebrate epithelial cells, adherens junctions serve as the primary connective link between cells, holding the epithelium together as a cohesive sheet through both the extracellular calcium-dependent cadherins interactions, as well as an intracellular link to the actin cytoskeleton (Ebnet, 2008). Despite partaking in such an adhesive role, AJs are also quite developmentally dynamic, showing evidence of constant reorganization and remodeling throughout epithelial morphogenesis (Wirtz-Peitz and Zallen, 2009).

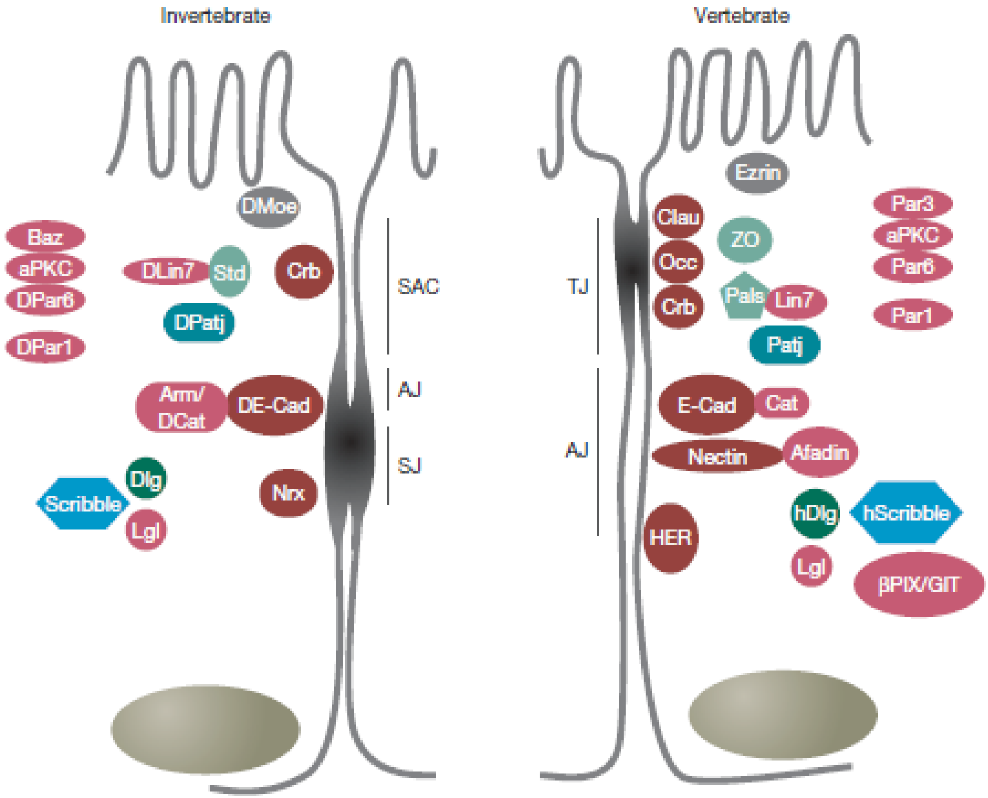


Figure 5: Architecture of invertebrate and vertebrate apical epithelial cellular junctions. Although the junctions vary between invertebrates and vertebrates, many of the core components, such as the Crumbs, Cadherin, Scribble complexes are highly conserved and serve similar developmental functions (Janssens and Chavrier, 2004).

Tight junctions (TJs) are another type of junction found in vertebrates that serve a somewhat different role. While AJs maintain the connective strength between cells, the tight junction is a “belt-like structure” around the apicolateral region of the cells with outbranching strands, providing a boundary between apical and lateral membrane domains, serving to regulate paracellular permeability and maintain apicobasal polarity. The primary molecules at TJs are the claudins and occludins, quadruple-pass transmembrane proteins, able to form homo- and heterotypic interactions with molecules on both adjacent and self-strands (Ebnet, 2008).

While invertebrates such as *Drosophila* do not contain tight junctions per se, they do possess somewhat homologous structures. Septate junctions (SJs), unlike TJs, are found basal to adherens junctions but serve essentially the same role in regard to regulation of paracellular transport. SJs are also somewhat different in structure, and instead of the classical claudins and occludins, contain similar Ig-containing transmembrane adhesion molecules (Furuse and Tsukita, 2006). Additionally, invertebrates harbor a membrane domain known as the sub-apical complex (SAC), which is located in the apical region homologous to vertebrate TJs and retains some orthologous junctional protein complexes (Janssens and Chavrier, 2004).

Additionally, many vertebrate and invertebrate cell types contain gap junctions and desmosomes. Gap junctions are essentially channels between cells that facilitate the transport of molecular and electrochemical signals, and can be homeostatically regulated to be in open or closed conformations. Desmosomes, on the other hand, support the connectivity of AJs by reinforcing contacts between cells and within the cytoskeleton, adding an extra degree of strength to the epithelial layer (Satir and Gilula, 1973).

Major classes of adhesion molecules

As previously mentioned, there are several major classes of adhesion molecules, each having specific roles at one or more cellular junction types. Many of these adhesive molecules are transmembrane proteins that facilitate binding in both *cis* and *trans*, mediated by Ig or Ig-like domains.

Cadherins

One such family is the cadherins, which contains approximately 80 members and several subfamilies, including classical (type I), atypical (type II), desmosomal, protocadherins, and Fat-like cadherins, among others. Including epidermal and neuronal cadherins, the classical subfamily is found primarily at the adherens junction. Classical cadherins contain five Ig-like calcium-activated ectodomains that facilitate primarily homotypic binding with adjacent cells. Type II cadherins are very similar in structure, but lack a core sequence in the first ectodomain believed to be crucial for the tight adhesion of AJs, thus resulting in its role in more “loosely associated cells” (Gooding et al., 2004). Cadherins also participate in complex formation with intracellular signaling and structural molecules, a subset of which facilitate a connection to the cytoskeleton.

Catenins

Catenins are a small family of cytoplasmic Armadillo repeat-containing molecules. Armadillo repeats, each containing three α-helicies, occur in tandem, forming a highly integrated superhelix (Coates, 2003). Several members of the catenin family, including β-catenin (Armadillo in *Drosophila*), α-catenin, and the p120 subfamily, localize to the AJ and interact with classical cadherins through their Arm repeats. Plakophilins, members of the p120 catenin family, localize to the desmosomes by binding desmosomal cadherins, and can also associate with cytoskeletal actin and tubulin filaments (Hatzfeld, 2005).

Claudins

Localized primarily at TJs, claudins are another crucial transmembrane cell adhesion family. Claudins are tetraspan molecules with two fairly conserved extracellular loops that facilitate homo- and heterotypic interactions, potentially in the form of tetramers, as well as an intracellular PDZ-domain binding site. Cytoplasmic scaffolding proteins such as the Zonula Occludens (ZO) family and membrane-associated guanylate kinases (MAGUKs) are known to bind claudins directly through this PDZ-domain binding site (Itoh et al., 1999). Functionally, claudins have been shown to be crucial for appropriate establishment of paracellular permeability, and certain members have been demonstrated to have atypical expression profiles in certain cancers (Lal-Nag and Morin, 2009).

JAMs

Also localized to the TJ, the junctional adhesion molecule (JAM) family is another subset of the Ig superfamily. JAMs, which contain a conserved structure consisting of two extracellular Ig domains, a single transmembrane domain, and an intracellular PDZ-domain binding site, are localized primarily at TJs and along the lateral membrane of epithelial cells and have been shown to facilitate both homo- and heterotypic interactions (Arrate et al., 2001; Kostrewa et al., 2001). In addition to their involvement in TJ assembly, JAMs have been implicated in several other adhesive roles, including platelet activation, angiogenesis, and as viral receptors (Mandell and Parkos, 2005).

Armadillo and Adherens Junction architecture

Adherens junctions, serving as a key adhesive force in cell-cell attachment, contain a number of transmembrane and intracellular complexes that modulate the intricate dynamics of extracellular associations and the intracellular cytoskeleton, many of which are highly conserved between vertebrates and invertebrates. These components include the core cadherin-catenin complex, as well as less well-studied molecules such as *Drosophila* Polychaetoid, Echinoid, and Canoe.

Armadillo/β-catenin: Adhesion and Wnt signaling

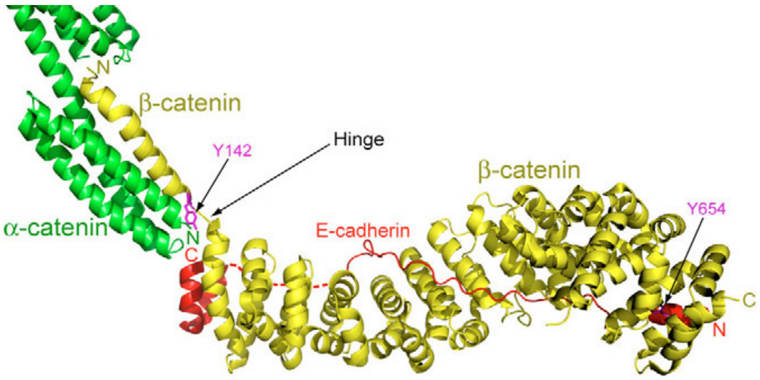


Figure 6: Crystal structure of the interfaces between E-cadherin, β-catenin, and α-catenin. Critical tyrosine residues, Y142 and Y654, undergo phosphorylation by Fer and Src kinases to cause dissociation of β-catenin from the membrane-bound cadherin complex (Xu and Kimelman, 2007).

β-catenin, originally discovered as Armadillo in *Drosophila*, is an intracellular molecule that serves crucial roles in both adherens junction-mediated cell adhesion and gene regulation involved with patterning and differentiation. Containing twelve Armadillo repeats, β-catenin's rigid core scaffold serves as a binding site for many partners (Fig 6) (Xu and Kimelman, 2007). At the cell membrane, β-catenin participates in the adherens junction cadherin-catenin complex, binding E-cadherin (DE-cadherin or Shotgun in *Drosophila*) and α-catenin, while free in the cytoplasm, acts as a transcription factor in response to extracellular Wnt signaling (Brembeck et al., 2006). The equilibrium between membrane-bound and cytoplasmic β-catenin is regulated largely through phosphorylation. Phosphorylation of key tyrosine residues by Fer and Src cause dissociation from the cadherin-catenin complex, while serine phosphorylation by casein kinase I (CKI) and glycogen synthase 3β (GSK3β) promote membrane association. Dysregulation of these kinases is a known cause of tumorigenesis, as it often inappropriately decreases cellular adhesion while increasing Wnt signaling (Brembeck et al., 2006). β-catenin's roles in adhesion and signaling are highly conserved between invertebrates and vertebrates, making *Drosophila* a particularly useful model for studying such cellular activities.

While at the membrane, β-catenin functions in a crucial link at the adherens junctions between extracellular cell-cell association and intracellular cytoskeletal components. Upon cell-cell contact, E-cadherin is stabilized in the membrane, immediately recruiting catenins (Nelson, 2008). In addition to functioning as a link within the complex, β-catenin binding serves to prevent proteolysis of E-cadherin, which is responsible for establishing and strengthening *trans* cellular contacts with like cadherins. As the contact strengthens, cadherin-catenin complexes accumulate, leading to an increased opportunity for α-catenin dimerization. These dimers can then bind actin, as well as regulate other actin remodeling molecules, including Actin-related proteins (Arps) and Rho-related proteins, which are involved in modulating membrane dynamics and cell migration. An additional related catenin, p120, which binds to a separate intracellular region of E-cadherin, has been suggested to further stabilize these cadherin-catenin complexes by preventing endocytosis through the modulation of Rho signaling (Nelson, 2008).

Also found freely in the cytoplasm, β-catenin is a key component of the Wnt signaling pathway, an important modulator of cell differentiation and proliferation (Nelson and Nusse, 2004). In the absence of Wnt ligand, GSK3β and CKI, in complex with axin and Adenomatous Polyposis Coli (APC), sequester and phosphorylate β-catenin, tagging it for ubiquitination and subsequent degradation by the proteasome. However, upon Wnt signaling through the transmembrane receptors Frizzled and Lipoprotein Receptor-related Proteins (LRPs), Disheveled is activated, which inhibits the GSK3β/CKI complex, allowing the accumulation of cytoplasmic β-catenin (Fig 7). Cytoplasmic β-catenin can then be translocated into the nucleus and associate with transcriptional cofactor T-Cell Factor/Lymphoid Enhancer Factor (TCF/LEF), causing the transcription of target genes (Nelson and Nusse, 2004). As a key developmental pathway in both invertebrates and vertebrates, improper regulation of any number of these steps can lead to cancers, stem cell ablation, and other cellular defects.

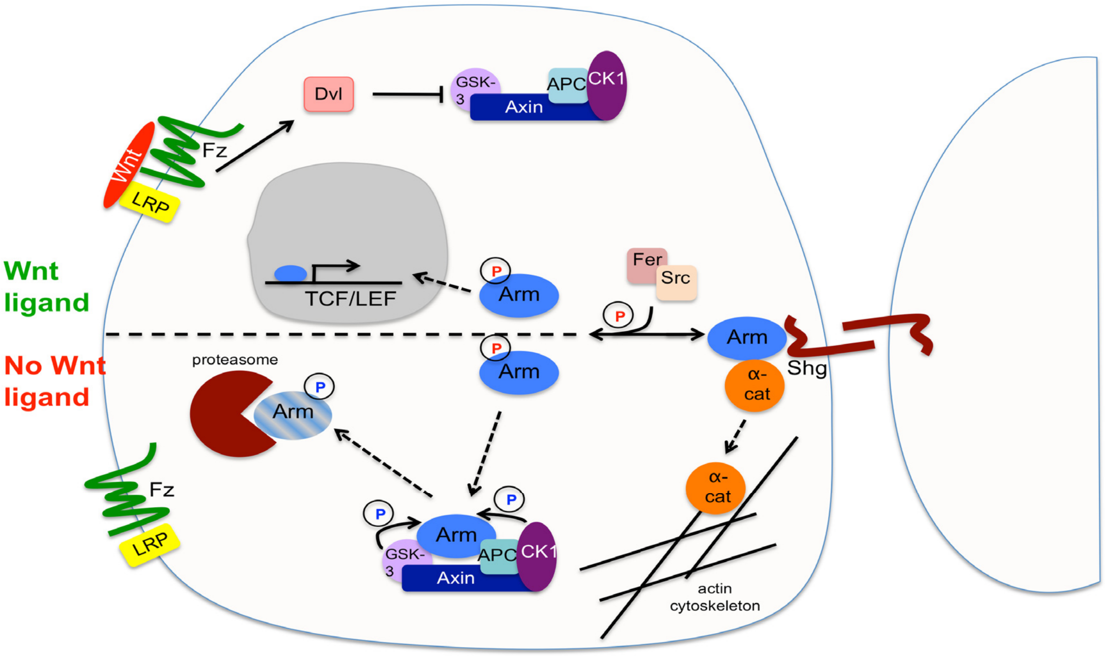


Figure 7: Dynamic regulation of membrane-bound and cytoplasmic Armadillo. Tyrosine phosphorylation regulates the equilibrium between cadherin-associated and free cytoplasmic Arm. Cytoplasmic Arm is subject to one of two fates: Wnt ligand-induced translocation into the nucleus to act as a transcription factor, or in the absence of Wnt signaling, proteasomal degradation induced by serine phosphorylation by the catenin destruction complex.

Polychaetoid

Polychaetoid (Pyd), *Drosophila* ortholog of vertebrate Zonula Occludens-1 (ZO-1), is an intracellular MAGUK protein found at not only adherens junctions, but throughout the lateral membrane as well (Wei and Ellis, 2001). Facilitating its role as a protein-scaffolding molecule, Pyd contains three N-terminal PDZ domains, an SH3 domain, as well as a proline-rich C-terminal domain, thought to interact with the actin cytoskeleton. The transcript is alternatively spliced into at least two isoforms, the most predominant of which localizes primarily to the adherens junctions, in contrast to its vertebrate ortholog, found almost exclusively at tight junctions (Wei and Ellis, 2001). Pyd has been demonstrated to regulate several developmental processes including dorsal closure (Wei and Ellis, 2001), tracheal morphogenesis (Jung et al., 2006), and retinal patterning (Seppa et al., 2008), and may act in concert with Bone Morphogenetic Protein (BMP) and Jun N-terminal Kinase (JNK) signaling in a number of these developmental roles. While the viability of null mutants is subject to debate, several alleles exhibit ectopic thoracic and scutellar bristles and rough eyes, and tissue-specific loss of function has been shown to modulate levels of a number of AJ components, including Arm, Shg, α-catenin, and Roughest (Rst), further indicative of its role in adhesion and junction regulation (Seppa et al., 2008).

Canoe

Drosophila ortholog of afadin/AF6, Canoe (Cno) is another likely intracellular protein scaffold, linking to the actin cytoskeleton (Matsuo et al., 1999; Miyamoto et al., 1995). In vertebrates, afadin associates with the transmembrane Ig-containing protein nectin via its PDZ domain, and colocalizes with E-cadherin at the adherens junctions, providing additional support in establishing cell-cell contacts. While not a direct component of the cadherin-catenin complex, knockdown of afadin does affect recruitment of E-cadherin to AJs (Sawyer et al., 2009). In addition to its PDZ domain, Cno contains several other protein-binding elements, including two Ras association domains, as well as myosin, kinesin, and actin binding motifs (Takahashi et al., 2002). While Cno's role in AJ maintenance is poorly understood, it has been demonstrated to bind the *Drosophila* nectin Echinoid and has been implicated in several developmental pathways and processes, including dorsal closure and asymmetric cell division, as well as JNK and Wnt signaling (Carmena et al., 2006; Sawyer et al., 2009; Speicher et al., 2008). Cno has also been demonstrated to interact with Pyd, furthering its putative role in AJ-actin cytoskeleton linkage (Sawyer et al., 2009). In addition to demonstrated genetic interactions with Dsh, Cno shares homology with Arm's GSK-3β phosphorylation motif, supporting its role in Wnt-related developmental regulation. While preliminary studies have indicated that Cno is not phosphorylated by GSK-3β, it is phosphorylated by related kinases Cdc2 and Cdk5 *in vitro* (Takahashi et al., 2002). Together, these data indicate a more complex, perhaps broader role for Cno in adhesion, junction maintenance, and development.

Echinoid

Echinoid (Ed), an invertebrate transmembrane Ig-containing nectin, colocalizes with E-cadherin at epithelial adherens junctions (Wei et al., 2005). Like E-cadherin, Ed participates in physical attachment at the junctions, homo- or heterotypically associating with itself or Neuroglian on adjacent cells (Islam et al., 2003). Ed contains a C-terminal PDZ domain-binding site, which facilitates binding with both Cno, predominantly serving to bind the actin cytoskeleton, and Bazooka (Baz, also known as PAR-3), an additional AJ component that also associates with Arm (Wei et al., 2005). As with E-cadherin, Ed-deficient epithelial cells demonstrate a sorting behavior in which the mutant clones will form smooth, round contours and associate only with each other, causing depletion of both E-cadherin and Ed at the interface of wild type cells, as well as upregulation of E-cadherin within the clone, implicating that while adhesion and polarity are still intact, Ed is essential for a coherent adhesive network throughout the epithelium (Wei et al., 2005). More recently, Ed has also been implicated in planar cell polarity (PCP) as a regulator of

Flamingo (Fmi) endocytosis, an atypical cadherin that cooperates with Wnt signaling to establish an axis of polarity perpendicular to the apicobasal plane (Ho et al., 2010). As such, although not yet extensively characterized, Ed serves as yet another integral CAM reinforcing both the physical stability of AJs as well as the signaling mechanisms involved in regulating polarity and patterning.

Drosophila wing development

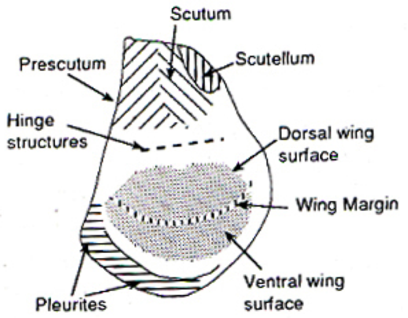


Figure 8: Anatomy of third instar wing imaginal disc. The disc proper (shaded grey) will become the wing blade, while the notum (upper hatched) will contribute to the thorax (Fristrom and Fristrom, 1993).

The *Drosophila* wing is a well-studied model for many cellular processes, including patterning, proliferation, signaling, and adhesion. In larval stages, the wing imaginal disc begins as a columnar epithelial monolayer with a basal lamina and a squamous peripodial membrane apposing the apical surface. Disc patterning begins early, and by the end of the larval stages, several committed regions have already been established, including the wing pouch, hinge regions, and the notum, which will contribute to the thorax (Fig 8) (Cohen, 1993).

Upon pupariation, the wing imaginal disc begins its metamorphosis into an adult tissue. The disc first expands and folds back upon itself into a bilayer along the wing margin, such that the dorsal and ventral regions of the wing pouch are now apposing each other (Fig 9)(Fristrom and Fristrom, 1993). Throughout prepupal and pupal development, the disc proceeds through two series of four stages each: apposition, adhesion, expansion, and separation. By approximately 11 hours after pupal formation (APF), the disc has progressed through the first developmental series, and is ready to complete its final mitotic cycle. By 60 hours AFP, the final separation has finished, and the wing is essentially ready for eclosion at 96 hours APF (Fristrom et al., 1993).

During wing development, patterning and axis formation are tightly regulated by several crucial patterning genes and morphogens. For example, anterior-posterior (A/P) axis formation relies heavily on Engrailed, expressed only in the posterior compartment, and the morphogens Hedgehog (Hh) and Decapentaplegic (Dpp), expressed in the posterior compartment and along the A/P boundary, respectively. The dynamics regulating such patterning are highly reliant upon cellular adhesion and intercellular signaling, as the A/P boundary acts much like an adhesive "fence," preventing inappropriate cell migration (McNeill, 2000). Later in pupal wing development, morphogens such as Dpp further contribute to patterning and development, such as vein/intervein patterning. Other adhesive transmembrane molecules known as integrins are also differentially regulated and expressed in the developing wing, with certain isoforms expressed only in the dorsal compartment, some only in the ventral compartment, and others in both (Brower et al., 1995). This regulatory complexity allows for a relatively small number of molecules to accomplish a wide range of tasks, efficiently governing patterning and development.

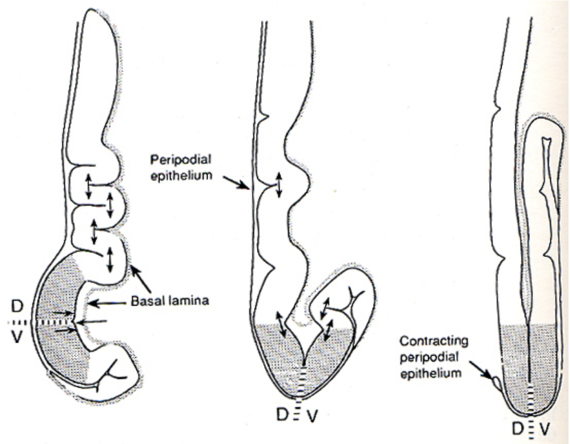


Figure 9: Pupal eversion of the wing disc. During pupal development, the flat disc proper (monolayer) folds outward upon the D/V boundary to form a bilayer that will eventually become the adult wing (Fristrom and Fristrom, 1993).

Kekkon5, a novel *Drosophila* Cell Adhesion Molecule?

Kekkon5 (Kek5), one of the six *Drosophila* Kekkon molecules, has been putatively implicated in a number of developmental and signaling-related activities. As with the rest of the Kek members, Kek5 contains seven LRRs, a single Ig domain, an intracellular PDZ domain-binding site, as well as five additional conserved intracellular motifs. Null mutants display wing growth and patterning defects, as well as reduced viability, indicating its necessity *in vivo*. Gain-of-function studies support this claim, as when misexpressed in the wing, Kek5 again induces defects in growth, patterning, and adhesion. These phenotypes are generally indicative of defects in BMP signaling and integrin-mediated adhesion (Evans, 2006). It has since been published that Kek5 does in fact inhibit BMP signaling, but the mechanism is yet unknown (Evans et al., 2009).

In order to further dissect the role of Kek5, a structural approach was taken. Several variants lacking various domains or motifs were generated in order to identify which sequence elements are critical for Kek5 activity (Evans, 2006). Prior work noted the necessity of motifs in the extracellular (EC) and intracellular (IC) domains, LRRs and the PDZ domain-binding site, respectively, as well as some portion of the remainder of the IC domain. This led to the generation of additional variants lacking the N-terminal and C-terminal halves of the IC domain, misexpression of which further demonstrated the necessity of particular IC motifs for Kek5 activity (Ernst, 2008). This work further examines Kek5 in light of a putative role at apical epithelial junctions, likely distinct from its other functions.

Materials and Methods

Molecular Biology

To generate full-length LIG expression constructs, glycerol stocks of *E. coli* (DH10B) transformed with AMIGO1-3, LRIG1-3 and LINGO1 cDNAs were obtained from OpenBiosystems. Stocks were grown up on LB-agar plates supplemented with the appropriate antibiotic (50μg/mL ampicillin for all AMIGOs, LRIG1, and LINGO1; 50μg/mL kanamycin for LRIG2 and 3). Colonies were cultured in liquid LB-amp or LB-kan and small-scale DNA extraction was performed with QIAGEN's QIAprep Spin Miniprep kit. Diagnostic digests were used to confirm the identity of the cDNAs.

One unit of Vent polymerase was used with 1-10ng of template and 1μg of each primer to generate PCR products of AMIGO1-3 and LINGO1. Primers were generated containing modified attB1 and B2 Gateway recombination sites (B1.1 and B2.1; see Appendix B). Primers used are included in Table 2 (primer sequences in Appendix C).

Table 2: PCR Primers used to generate full-length LIG expression constructs.

Gene	5' Primer	3' Primer
AMIGO1	W104	W105
AMIGO2	W106	W107
AMIGO3	W108	W109
LINGO1	W110	W111

PCR products were checked for size on a 0.8% agarose gel in TAE buffer, and were band purified with QIAGEN's QIAquick Gel Extraction kit. Using Invitrogen's Gateway recombination-based cloning system (summarized in Fig 10), 150ng of each purified product was cloned into 150ng of *pDONR* with BP Clonase and incubated overnight at 25°C. 4μL of each recombinant construct was then transformed into 50μL of Max-Efficiency DH5α *E. coli* cells from Invitrogen and plated on LB-agar supplemented with 50μg/mL

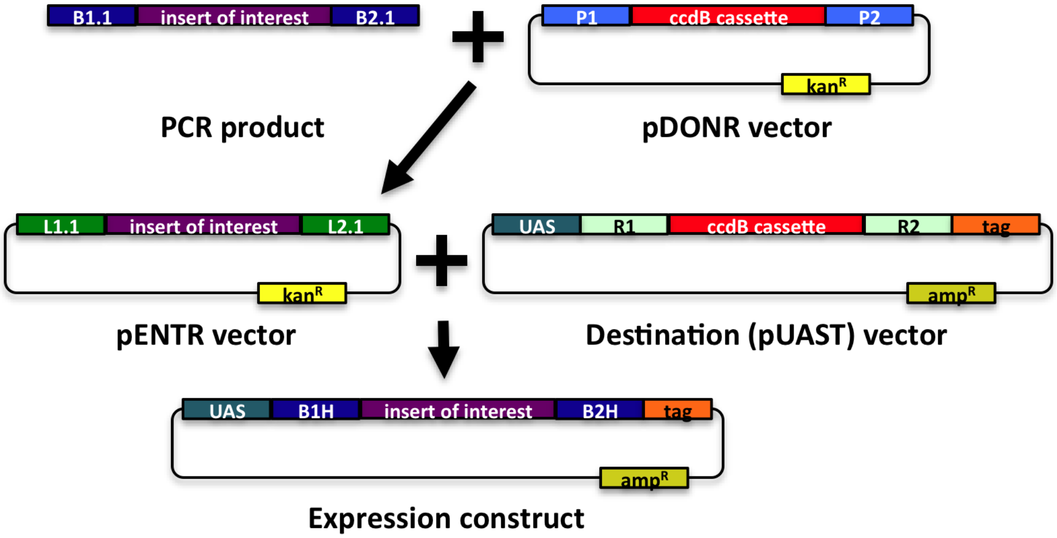


Figure 10: Summary of Gateway cloning procedure. Insert of interest is first generated via PCR with flanking attB recombination sites. The PCR product is recombined via a "BP" reaction into a *pDONR* vector containing attP sites to generate a *pENTR* vector (resulting in hybrid recombination sites, now termed attL sites). The *pENTR* vector can then be used to shuttle the insert of interest into a destination vector containing attR sites ("LR" reaction), resulting in the final expression construct (now again with attB sites).

kanamycin. Colonies were cultured in liquid LB+kan, miniprepped, and confirmed with diagnostic digests.

Each positive recombinant (now termed *pENTR* vector) was again recombined with the *pUAST-a-V5/6XHis* destination vector (150ng of each) with LR Clonase, and incubated at 25°C overnight. 4μL of each recombinant construct was transformed into 50μL of Subcloning-Efficiency DH5α *E. coli* cells (Invitrogen) and plated on LB-agar supplemented with 50μg/mL ampicillin. Colonies were cultured in liquid LB+amp, miniprepped, and confirmed with diagnostic digests. A single positive expression clone of each construct was selected for mass production with QIAGEN's Plasmid Maxi kit. Final expression construct sequences are included in Appendix D.

Secreted versions of dEGFR (DER), Kek1, and Kek2 were generated in a similar fashion. Internal 3' primers were designed just 5' of the transmembrane region of each gene. All primers included modified attB sites. PCR products were generated as described above. Templates and primers are included in Table 3 (primer sequences in Appendix C):

Purified PCR products were cloned into *pDONR* and subsequently *pUAST-a-GFP* and *pUAST-a-V5/6XHis* as described above. Final expression construct sequences are included in Appendix D.

Table 3: PCR primers and templates used to generate secreted Keks and dEGFR.

Gene	5' Primer	3' Primer	PCR Template
sDER	W141	W142	*pUAST-DER-GFP* (K. Cook)
sKek1	W143	W144	*pUAST-Kek1-V5/6XHis* (D. Alvarado)
sKek2	W145	W146	*pUAST-Kek2-V5/6XHis* (D. Alvarado)

Genetics

Fly stocks were maintained at room temperature (approximately 24°C) on standard yellow media. All experimental crosses were kept at 28°C on standard media. The experiments carried out in this work rely upon an expression system known as the GAL4-UAS system (Brand and Perrimon, 1993). Using this system, two strains may be mated to express a protein of interest in a specific spatial and temporal pattern. The first strain is known as the "driver," which expresses the yeast protein GAL4 with a particular promoter that expresses in a desired pattern, whether it be in the embryo, wing, eye, etc. The second strain is the "responder," which consists of a gene of interest and a number of upstream activating sequence (UAS) sites, which promote transcription upon GAL4 binding (Fig 11).

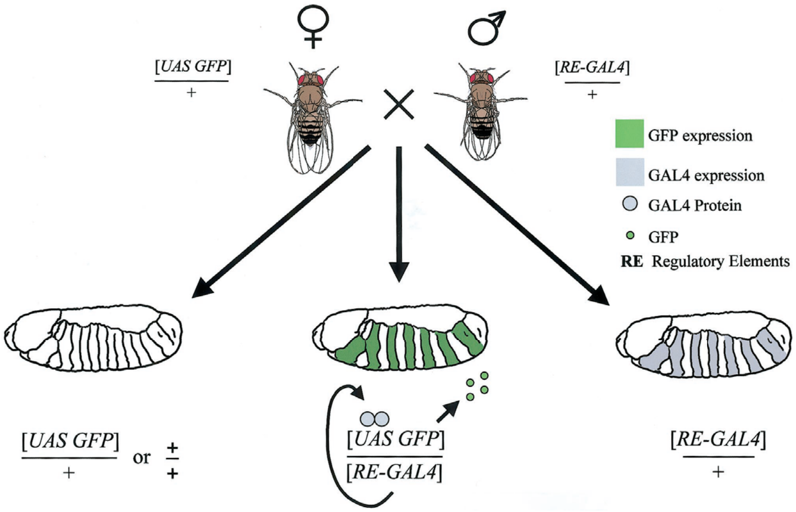


Figure 11: GAL4-UAS expression system in *Drosophila*. To express a given protein in a particular spatial and temporal pattern, a "driver" containing promoter-driven GAL4 and "responder" containing upstream activating sequence sites and the protein of interest must be mated (Duffy, 2002).

The following driver lines were used (with accompanying Bloomington Stock Center numbers):

- *A9-GAL4* – wing, dorsal (8761)
- *ap-GAL4* – wing, dorsal (3041)
- *CY2-GAL4* – ovarian follicle cells (not available; insertion *p[GawB]CY2*)
- *en-GAL4* – wing, posterior (6356)
- *ptc-GAL4* – wing, A/P boundary (2017)

All Kek lines were generated in the Duffy Lab. *UAS-Pyd.RFP* and *UAS-pyd RNAi* lines were obtained from R. Cagan. *UAS-mCD8.GFP* and *UAS-mCD8.RFP* can be found in the Bloomington collection as well (stock numbers 5137 and 27391, respectively). All transgenic stocks generated in the Duffy lab are distinguished with individual line numbers, such as Kek5$^{16\ \mathrm{II}}$ (where '16' indicates the transgenic insertion number and 'II' indicates the chromosome in which it was inserted).

All experimental crosses were set up *de novo* (generally with GAL4 virgin females and UAS males), with a few exceptions. Crosses requiring two different UAS lines in generation F$_1$ were set up with pre-established *ptc-GAL4>UAS-Kek5.GFP*$^{16\ II}$ recombinant virgin females, which were mated by additional UAS males (for example, *UAS-mCD8.RFP*). For most experiments, *UAS-mCD8.GFP* was used as a wild-type control with the appropriate driver, although in some of the earlier experiments, *w*^{1118} was used.

The FLP-FRT system was used to generate random clones misexpressing Kek5.GFP. This was accomplished by first mating heat shock-induced FLP recombinase (virgin females; *p[hsFLP]1, y*^{1} *w*; Sp*^{1}*/CyO; B14-A/TM6B, Tb*^{1}*, brk-lacZ*) by *Actin-GAL4* interrupted by an FRT-flanked CD2 cassette (*Act5C-GAL4(FRT.CD2)*, stock 4780). Virgin F$_1$ females containing both *hsFLP* and *Act-GAL4 (FRT.CD2)* were then mated to *UAS-Kek5*$^{16\ II}$ males. This F$_1$ cross was kept at 25°C and females were allowed to lay eggs for 24 hours before being flipped to a new vial. The cleared vial remained at 25°C for an additional 72 or 96 hours (to allow for approximately 48 or 24 hours of protein expression, respectively, before dissection of 3rd instar larvae). The vial was then heat shocked in a circulating 37°C water bath for

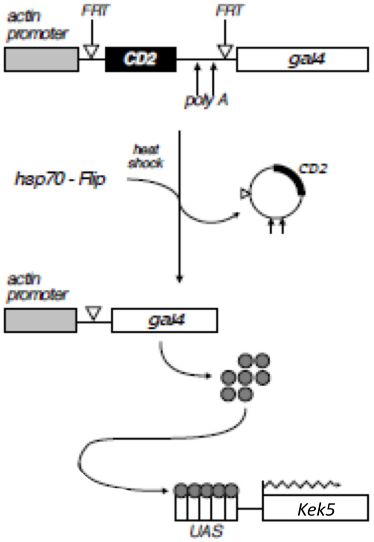


Figure 12: hsFLP-FRT random clone induction. Upon heat shock, FLP recombinase is activated, which excises the FRT-flanked CD2 cassette and allows for GAL4 production (adapted from (Pignoni and Zipursky, 1997)).

10 minutes, causing excision of the FRT-flanked CD2 cassette and allowing GAL4 production (summarized in Fig 12). Vials were then kept at 28°C until dissection to maximize protein production.

Imaging and Immunohistochemistry

All images (except adult bristles) were acquired on a Zeiss Imager.Z1 with a Zeiss Axiocam (and ApoTome for fluorescence imaging). Images were captured and processed with Zeiss's Axiovision software. All larval disc fluorescence images are compilations of 10-30 slices processed with a Maximum Intensity Projection algorithm. Images were further adjusted with Adobe Photoshop CS3 or Microsoft PowerPoint. Adult bristle images were acquired on a Zeiss Discovery.V12.

Adult wings were dissected in 100% ethanol and mounted in GMM (20% Canada Balsam, 80% methyl salicylate). Slides were cured overnight on a heat block at 65°C. For bristle imaging, adults were frozen overnight and mounted on double-sided tape. The light source was supplemented with Zeiss KL1500 fiberoptic lighting from above.

Wing imaginal discs were dissected from 3rd instar larvae in PBS and fixed in PEMP (0.1M PIPES, 2mM $MgSO_4$, 1mM EDTA, 0.5% NP-40) + 3.7% formaldehyde for 15 minutes. Discs were washed several times with antibody wash (100mM TrisCl, 150mM NaCl, 1mg/mL BSA, 0.1% NP-40) and incubated in antibody block (antibody wash + 5mg/mL BSA) for at least an hour at room temperature. Discs were incubated overnight at 4°C in primary antibody, washed several times in antibody wash, and incubated for 2 hours at room temperature in secondary antibody. Both primary and secondary antibodies were diluted in antibody wash. After several additional washes, discs were stored and mounted in 50% glycerol in PBS + 5-10μL of Slowfade (Invitrogen).

For BrdU stainings (protocol adapted from (Rogulja et al., 2008)), discs were dissected in Ringer's solution (130mM NaCl, 5mM KCl, 1.5mM $MgCl_2$) and incubated in Schneider's medium with 0.1mg/mL BrdU for 30 minutes. Discs were rinsed in cold PBT (PBS + 0.1% Tween-20) and fixed in PBT + 3.7% formaldehyde for 20 minutes. After several washes in PBT, discs were incubated in 400μL DNase mix (PBS + 20μL DNase + 40μL reaction buffer; Promega) for 1.5 hours. After several additional washes in PBT, discs were stained with anti-BrdU as above, with the exception that antibody dilutions and washes were performed with PBT instead of antibody wash.

Table 4: Antibodies and dilutions used for IHC and ICC. DSHB antibodies are monoclonal supernatants from the Developmental Studies Hybridoma Bank

Antibody	Dilution	Host	Source
Armadillo	1:100	mouse	DSHB (N27A1)
BrdU	1:200	mouse	DSHB (G3G4)
Canoe	1:500	rabbit	D. Yamamoto
Echinoid	1:250	rat	J. Hsu
dEGFR	1:2000	rabbit	J. Duffy
p120catenin	1:25	mouse	DSHB (p4B2)
Polychaetoid	1:1000	rabbit	R. Cagan
Shotgun	1:250	rat	DSHB (DCAD2)

Ovaries were dissected from adult females in PBT and fixed in PBT + 3.7% formaldehyde for 15 minutes. After several washes in PBT, ovaries were stored and mounted in 50% glycerol in PBS.

Primary antibodies were used at dilutions as indicated in Table 4. All secondary antibodies were from the Alexa Fluor series (405, 488, 568; Invitrogen) and were used at 1:500.

Cell Culture and Protein Expression

S3 cells (embryonic, courtesy of N. Perrimon) were maintained at 24°C in Schneider's medium (Gibco) + 12.5% FBS (Valley Biomedical). Clone8 cells (3rd instar wing imaginal disc, courtesy of N. Perrimon) were maintained at 24°C in M3 medium supplemented with 2% FBS, 5μg/mL insulin, and 2.5% fly extract (complete medium was supplied with cells;

Clone8 cells will not survive in Schneider's-based medium). [Fly extract can be made by homogenizing 1g of frozen flies in 6.8mL M3 medium on ice. Extract is spun at 1500 x g at 4˚C for 15 minutes. The supernatant is transferred to a fresh tube and incubated at 60˚C for 5 minutes to inactivate tyrosinase. The inactivated extract is again spun at 1500 x g at 4˚C for 90 minutes and the supernatant is filter-sterilized with a 0.22μm filter. The final extract is used at a concentration of 2.5mL per 100mL final media (remainder can be stored at -20˚C).] Cells were split 1:10 every 3-4 days, when confluent. Additional culture maintenance details can be found at the *Drosophila* Genomics Resource Center.

For transfections, when 40-60% confluent (~1-2 days after splitting), cells were seeded in 6-well plates at 5-8 x 10^5 cells per well, in 2mL of media. When 40-80% confluent (~1-2 days), cells were cotransfected with 200ng each of *Arm-GAL4* and the appropriate expression construct using QIAGEN's Effectene reagent. Transfected (GFP positive) cells were visible within 24 hours, and transfections were allowed to continue for 3-4 days before cell staining or Western blotting.

For staining of transfected cells, multiwell slides were treated with 1mg/mL poly-L-lysine (Sigma) for 15 minutes and rinsed thoroughly. 10μL of cells were added to each well and allowed to settle for 15 minutes. An equal volume of fixative (3.7% formaldehyde in PBS; final concentration 1.8%) was added and allowed to set for 15 minutes. Cells were then gently rinsed several times with PBS and incubated in primary antibody for 1-2 hours at room temperature with shaking. Cells were again rinsed with PBS and incubated in secondary antibody for 1 hour at room temperature with shaking. Antibodies were diluted in PBS + 1% NGS + 0.1% saponin. After rinsing with PBS, 50% glycerol in PBS was added and coverslip was sealed.

For cells grown on a coverslip, 18x18mm coverslips were treated with poly-L-lysine as above and sterilized in 70% ethanol. Treated coverslips were placed in empty wells in a 6-well plate. 10μL of resuspended transfected cells were added and immediately diluted to surface-tension capacity with culture media to spread the cells. After allowing cells to settle for 30 minutes, 2mL of media was added and cells were allowed to grow for another 1-2 days. Cells were fixed in PBS + 1.8% formaldehyde and stained as above. After staining, coverslips were mounted in PBS + 50% glycerol and sealed.

For detection with a Western blot, 50-100μL of resuspended cells were spun down at 3000 x g and supernatant was aspirated. Pelleted cells were taken up in 20μL of sample buffer (12mM TrisCl, 5% glycerol, 0.4% SDS, 2.9mM βME, 0.02% bromophenol blue) and stored at -20°C. Prior to electrophoresis, cell harvests were lysed by boiling for 5 minutes. Samples were run on an 8% resolving SDS-PAGE gel with a 4% stacking gel at 20mA for 1-1.5 hours (until appropriately resolved) in electrophoresis buffer (25mM Tris, 192mM glycine, 0.1% SDS). Proteins were transferred to a nitrocellulose membrane at 100V for 1 hour in transfer buffer (15.6mM Tris, 120mM glycine). Membrane was Ponceau stained and gel was Coomassie stained to check for completeness of transfer. The membrane was blocked for 1 hour in 5% non-fat dry milk (NFDM) in TBST (100mM Tris, 150mM NaCl, 0.1% Tween-20). Primary antibody incubation was performed at 4° overnight. GFP-tagged constructs were detected with monoclonal anti-GFP (JL-8, Clontech; 1:1000 in TBST), and V5-tagged constructs were detected with monoclonal anti-V5 (Invitrogen; 1:1000 in TBST + 1% NFDM). After several washes in TBST, membrane was incubated for 2 hours in HRP-conjugated goat anti-mouse (Jackson Immuno) at a final dilution of 1:20,000 in TBST + 5% NFDM (lyophilized antibody stock was first taken up in 2mL dH_2O and diluted 1:2 in glycerol). After several washes in TBST, membrane was incubated in HRP substrate for up to 10 minutes (Clontech His Kit substrate; 1:1:2 luminol:peroxide:dH_2O), exposed to Kodak MR-1 film and developed on an Xomat.

Bioinformatics

All *Drosophila* PDZ domain-containing molecules were identified through FlyBase.org by searching with the term *PDZ* (see Appendix E). With these results, under "HitList Conversion Tools," the Protein Data tool was selected, yielding a list of PubMed and UniProt accession numbers for each putative PDZ domain-containing protein. On the UniProt page for each molecule, the InterPro graphical view was selected to retrieve the protein structure annotated with all known domains, which were compiled in Figure 35.

Results

Prior work has implicated Kek5 in a number of roles, including BMP- and adhesion-related activities. Structural dissections using variants lacking the LRRs, Ig domain, and intracellular domain (with and without the PDZ domain-binding site) revealed the necessity of the extracellular LRRs and intracellular PDZ domain-binding site for full activity, both likely necessary for proper folding, stability, and localization (Evans, 2006). Recent investigation has confirmed these results, in addition to reassessing the contribution of the IC domain (Ernst, 2008). Although it had been previously noted that the PDZ domain-binding site was sufficient for both localization and activity, this recent work indicated that one or more additional conserved IC motifs are necessary for Kek5 activity. To this effect, three new IC deletion variants were generated: Kek5Δ123, Kek5Δ45, and Kek5ΔIC+PC (adding an additional six C-terminal residues to the putative PDZ domain-binding site) (Ernst, 2008). This work goes on to further assess Kek5-mediated effects on epithelial junction architecture through dissections of structure-function relationships, family specificity, and co-modulation of additional molecules. It also aims to commence a broader *in vitro* investigation of the vertebrate LIG family through the development of a high-throughput screening platform for putative binding partners.

Kek5 affects wing development

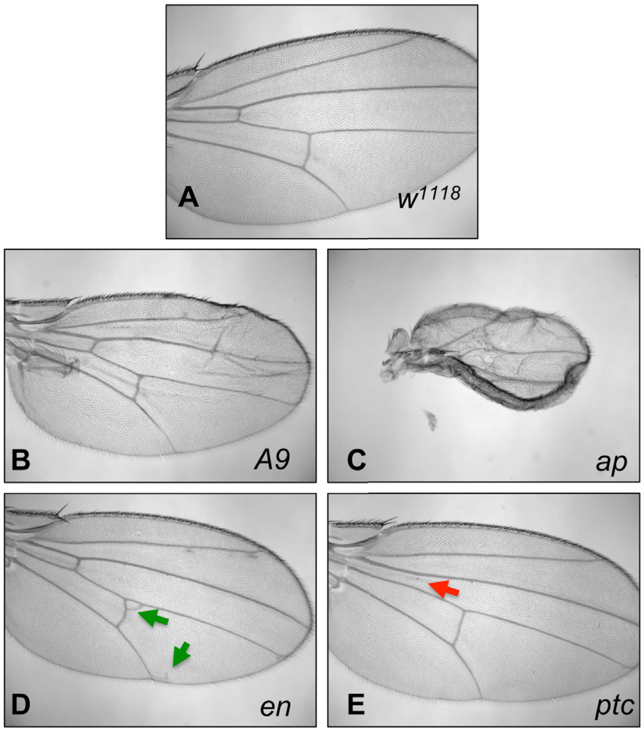


Figure 13: Kek5 misexpression in the adult wing. Misexpression of Kek5 in the adult wing causes a number of morphological defects. (A) Wild type. Expression by *A9-GAL4* (B) and *ap-GAL4* (C) induces curling and blistering. In contrast, expression by *en-GAL4* results in ectopic veins (D; green arrows) and growth defects, while expression by *ptc-GAL4* causes loss of crossvein (E; red arrow). Bright field micrographs of adult wings acquired with 5X dry objective.

Much of the prior work on Kek5 has focused on misexpression-induced adult phenotypes. Upon Kek5 misexpression in the wing, a number of different BMP- and adhesion-related phenotypes result, depending on which *GAL4* driver is used. *A9- and ap-GAL4* both express in the dorsal compartment (becomes the dorsal surface of the adult wing), *en-GAL4* expresses in the posterior compartment, and *ptc-GAL4* expresses along the anterior-posterior boundary of the wing. Phenotypes resulting from Kek5 misexpression by these wing drivers vary from blistering and upward curling (adhesion-related; Fig 13B and C) to crossvein and growth defects (BMP-related; Fig 13D and E).

To confirm and round out prior structure-function analyses, all Kek5 deletion variants on-hand (Fig 14) were misexpressed with *A9-GAL4* and *ap-GAL4*, both of which express in the dorsal surface of the wing, and analyzed for adult wing phenotypes. Misexpression of wild type Kek5 by *A9* and *ap* result in upward wing curling and severe blistering/curling, respectively, making them ideal for analyzing enhancement or suppression of activity. As reported previously, LRR, IC motifs 4-5, and PDZ domain-binding site deletion variants demonstrated little to no activity with respect to wild type Kek5. In contrast, deletion of IC motifs 1-3 showed increased/novel activity, possibly indicative of an auto-inhibitory function. Deletion of the Ig domain, however, does not affect Kek5 activity in these assays (summarized in Table 5). Together, these results suggest that the LRRs and one or more IC motifs are crucial for Kek5 function in the adult wing.

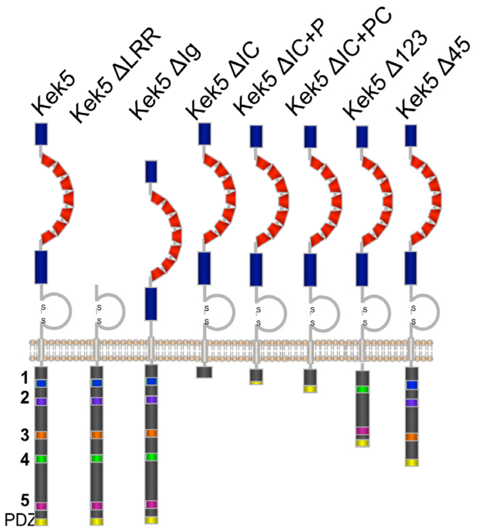


Figure 14: Kek5 variants. Several deletion variants of Kek5 have been generated, including those lacking the extracellular LRRs, Ig domain, and various permutations of intracellular motifs.

Table 5: Structure-function dissection of Kek5-induced adult wing phenotypes. All Kek5 deletion variants were expressed by the indicated GAL4 driver. Phenotypes range from wild type (-) to severe (+++) or lethal (L).

	A9-GAL4	**n (wings)**	***ap-GAL4***	**n (wings)**
Kek5$^{16\ \mathrm{II}}$	++	110	+++	40
Kek5$^{16,52\ \mathrm{II}}$	+++	103	L	N/A
Kek5 ΔL$^{16\ \mathrm{III}}$	-	102	+	135
Kek5 ΔL$^{43\text{-}1,16\ \mathrm{III}}$	-	114	++	114
Kek5 ΔI$^{13\text{-}2\ \mathrm{II}}$	+	80	+++	22
Kek5 ΔI$^{27,13\text{-}2\ \mathrm{II}}$	++	116	L	N/A
Kek5 ΔIC$^{11\ \mathrm{III}}$	-	112	-	94
Kek5 ΔIC+PC$^{11\ \mathrm{I}}$	-	>50**	-	158
Kek5 Δ123$^{16\ \mathrm{II}}$	+++*	>50**	L	N/A
Kek5 Δ45$^{25\ \mathrm{II}}$	+	>50**	++	102

* Misexpression of Kek5 Δ123 by *A9-GAL4* also induces downward curling, a novel phenotype.
** Although n values were not recorded, the indicated lines were extensively tested upon transgenic stock generation (Ernst, 2008).

Kek5 misexpression leads to altered cellular morphology and epithelial extrusion

Defects in adult tissues are often the result of distinct changes in cellular morphology and signaling during development. In order to better understand the effects of Kek5 misexpression on adult tissues, 3rd instar wing imaginal discs were examined at the cellular level. Each Kek5 variant is tagged with GFP, allowing for simple fluorescent visualization of its localization. In nearly all of the following experiments, Kek5 misexpression was driven by the *ptc-GAL4* driver, which expresses in a stripe along the anterior-posterior boundary of the wing disc. This allows for easier analysis of wild type versus mutant tissue, as it is a relatively small region passing through the center of the disc proper, and has a distinct boundary on the posterior side of the stripe (as compared to the other wing drivers, each of which expresses in an entire compartment). When misexpressed, Kek5 demonstrates tight membrane-localized apico-lateral and, to a lesser degree, lateral expression (Fig 15). Current data strongly suggests that endogenous Kek5 localizes apically as well, implying that perhaps misexpressed Kek5 first saturates its normal apico-lateral domain, and the remainder is subsequently localized more basally along the lateral membrane (H. Haridas, personal communication).

Although localization in the disc proper appears predominantly membrane bound, it is worth noting that the *ptc-GAL4* expression domain extends around the disc into the peripodial layer, and a number of Kek5-misexpressing cells in this region appear to have neuron-like extensions protruding beyond the cell body. Unfortunately, demonstration that these extensions are a result of Kek5 misexpression has been unsuccessful, as misexpression of a membrane-bound negative control (mouse surface receptor CD8, or mCD8) does not colocalize with the GFP-labeled extensions (Fig 16). Additional membrane or cytoskeletal markers may aid in further characterizing these extensions.

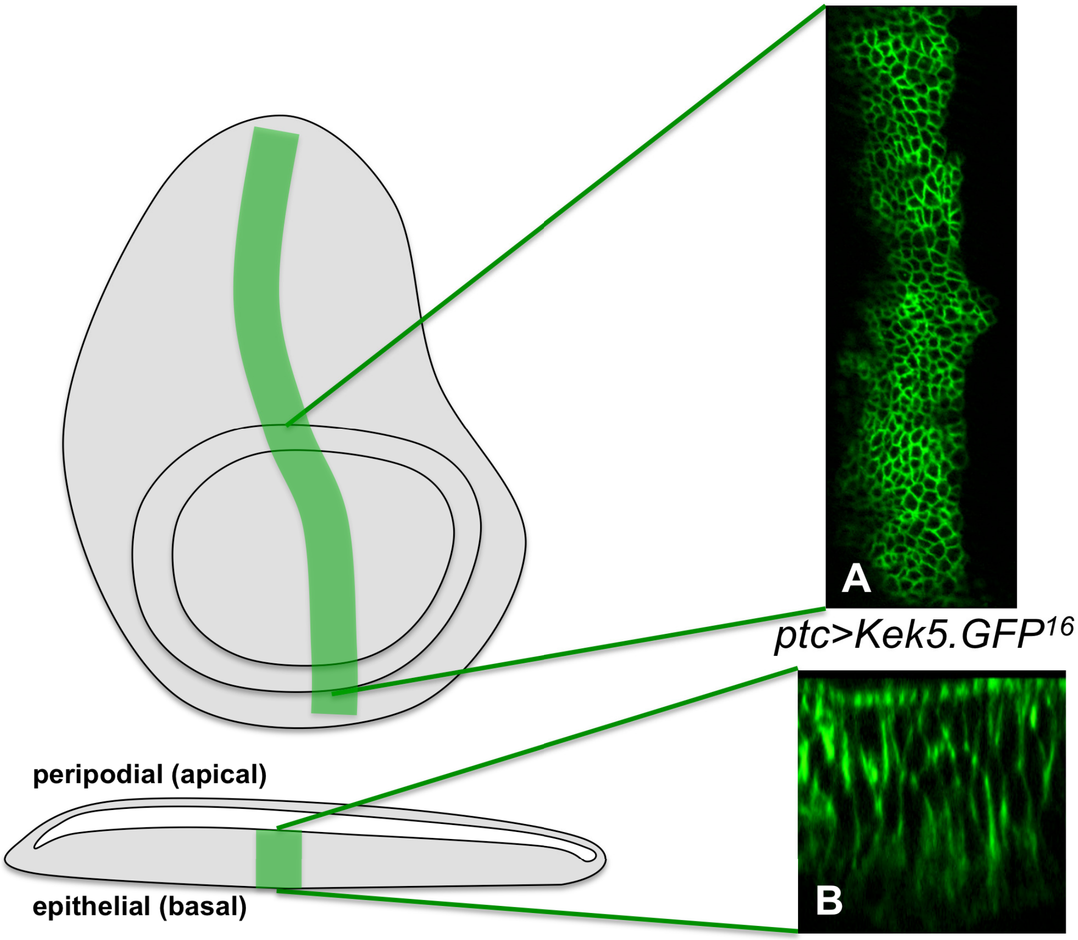
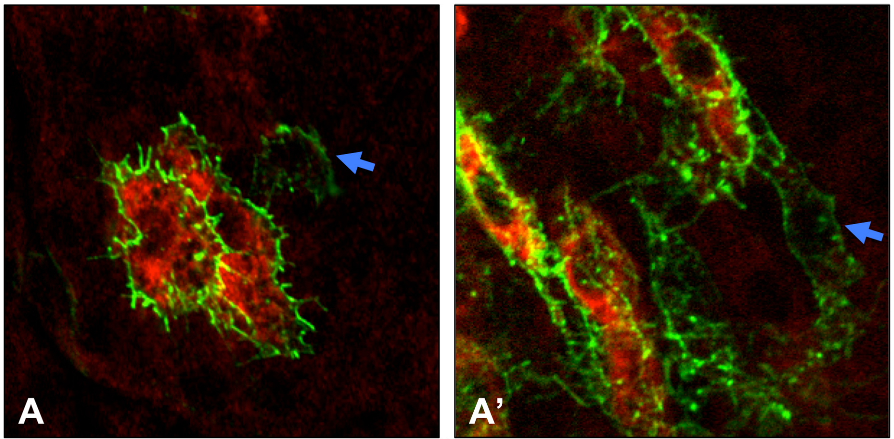
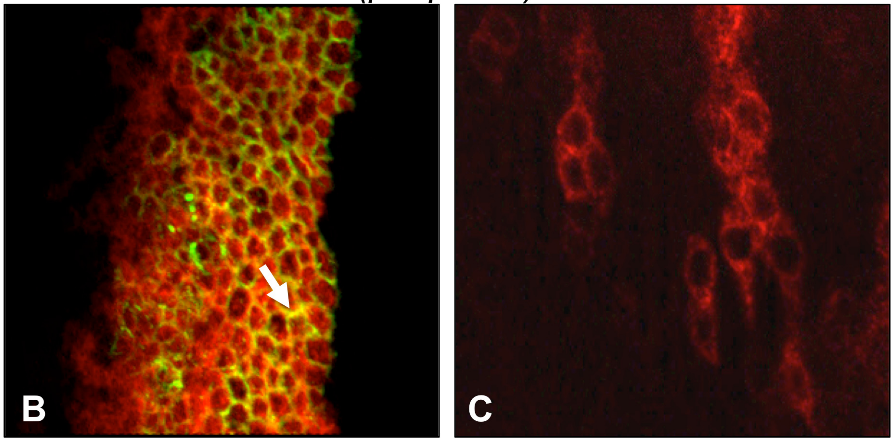


Figure 15: Misexpressed Kek5 is membrane-localized. When expressed with *ptc-GAL4*, Kek5.GFP is tightly membrane-localized near the apical surface (A), as well as more basally along the lateral membrane (B). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective. Cross-section processed with Axiovision.



ptc>Kek5.GFP^{16} + mCD8.RFP
(peripodial)



ptc>Kek5.GFP^{16} + mCD8.RFP (proper, apical)

ptc>mCD8.RFP (peripodial)

Figure 16: Kek5 misexpression may induce neuron-like cellular extensions. When co-misexpressed with mCD8.RFP, Kek5.GFP induces neuron-like extensions that do not colocalize, especially in the peripodial membrane (A, A’; blue arrows). In contrast, mCD8.RFP and Kek5.GFP do exhibit some membrane colocalization in the disc proper (B, white arrow). mCD8.RFP is shown alone for comparison with wild type peripodial cells (C). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective.

Upon examination of the wing disc morphology, it was noted that a large number of cells extruded out of the basal membrane along the region of Kek5 misexpression (Fig 17A and B). Similar cellular extrusion has been previously linked with BMP-related effects, further suggestive of Kek5's role in BMP signaling. However, it is currently unclear what the cellular basis for BMP-mediated cellular extrusion is. One simple model is that cellular adhesion is altered, thereby leading to extrusion. To investigate this, the discs were stained for a component of the cellular adhesion machinery, Shotgun (Shg, E-cadherin), a membrane marker, which is localized at adherens junctions in the apico-lateral region of the cells. Although localization and levels of Shg appear unaffected throughout the disc, cells in the region of Kek5 misexpression appear to be larger than their wild type counterparts (Fig 17C and D). Whether this enlargement is a true growth or adhesion defect, or simply a shift in cell proportions, is unknown.

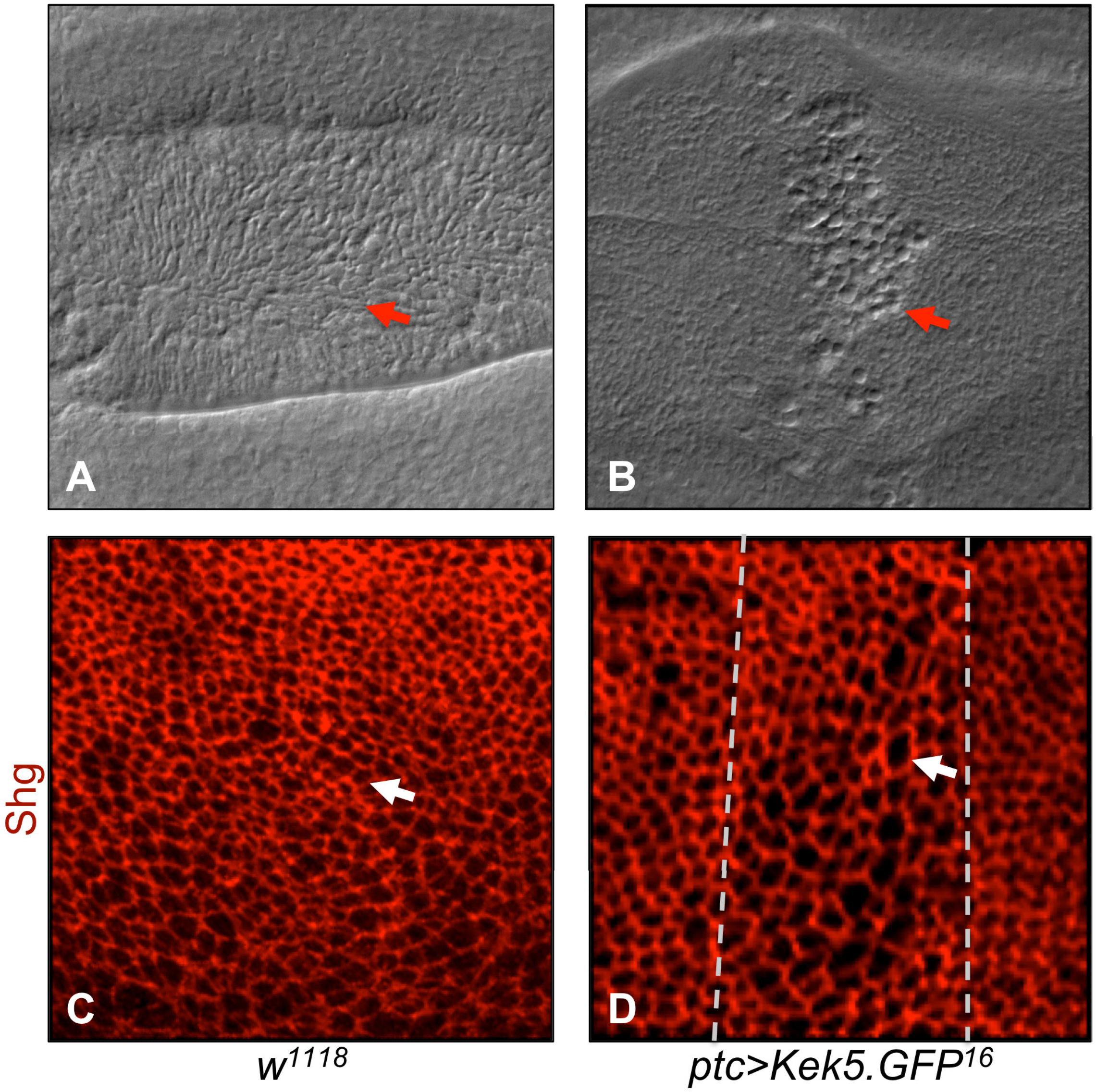


Figure 17: Kek5 misexpression induces epithelial extrusion and cell enlargement. Misexpression of Kek5 by *ptc-GAL4* causes cells within the region of misexpression to extrude from the basal surface (A, B; compare red arrows). When stained with an apical membrane marker (Shg, red), cell enlargement is also visible (C, D; compare white arrows). DIC and fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective.

Many developmental signals in the wing originate from either the A/P or D/V boundary and diffuse outward to form a gradient that is capable of eliciting distinct signals dependent upon the concentration of the given molecule. These molecules are known as morphogens and are crucial to a number of developmental signaling pathways, including the BMP pathway. Because Kek5 has been implicated in BMP signaling, it is possible that the observed cellular effects may exhibit regional specificity throughout the wing disc. To address this possibility, Kek5-expressing clones were randomly generated using heat shock-activated FLP recombinase (*hsFLP*) to excise an FRT-flanked CD2 cassette inserted between the Actin promoter and GAL4 (*Act-GAL4 (FRT.CD2)*). Removal of the FRT cassette by the FLP recombinase then juxtaposes the Actin promoter with GAL4 leading to misexpression of Kek5 in cells where FLP has been induced. Kek5-misexpressing clones confirmed the correlation between Kek5 misexpression, extrusion, and enlarged cells, and demonstrated that these effects are ubiquitous throughout the wing disc proper (Fig 18). Thus, Kek5-induced cellular defects are likely not location-dependent and unlikely to be critically dependent upon regionally distributed factors such as morphogens. In addition,

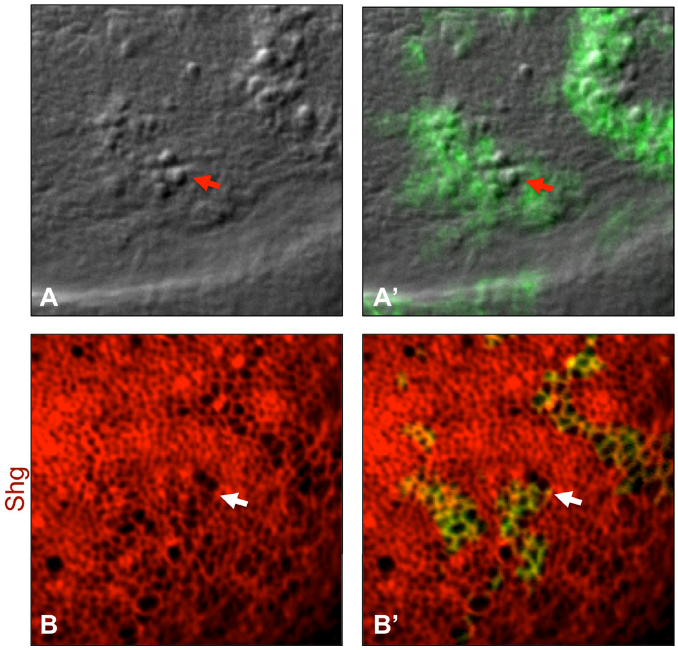


Figure 18: Clonal analysis of Kek5 misexpression. When expressed in random clones using *hsFLP* and *Act-GAL4 (FRT-CD2)*, Kek5.GFP induces both extrusion (A, A' overlay; red arrows) and cell enlargement (B, B' overlay; white arrows) as visualized with Shg (red) with no regard to location in the disc proper. DIC and fluorescent micrographs of 3rd instar wing imaginal discs acquired with 20X dry objective.

the strict correlation between Kek5 misexpression and cellular effects strongly suggests a cell autonomous mechanism for Kek5 function.

It has been noted that distinctly enlarged cells in the wing disc epithelium are often mitotic (Marygold and Vincent, 2003), and that morphogens such as the *Drosophila* BMP ligand Dpp can influence cell proliferation (Rogulja and Irvine, 2005). Given that Kek5 has the ability to affect at least one such developmental signaling pathway, it is possible that it could also be disrupting tightly coordinated cell cycle regulatory mechanisms, perhaps causing the cells to remain arrested in mitosis. To determine whether Kek5 misexpression induces cell cycle defects, discs were stained with BrdU, which is actively incorporated during DNA replication (Fig 19). However, there were no overt differences between wild type and Kek5-misexpressing tissue, implying that the cell enlargement is not likely a result of aberrant cell cycling. These data suggest that the extrusion and cell enlargement observed under Kek5 misexpression are likely not mediated through morphogens or defects in cell cycle regulation, but may rather be a result of changes in cellular adhesion and/or cytoskeletal dynamics.

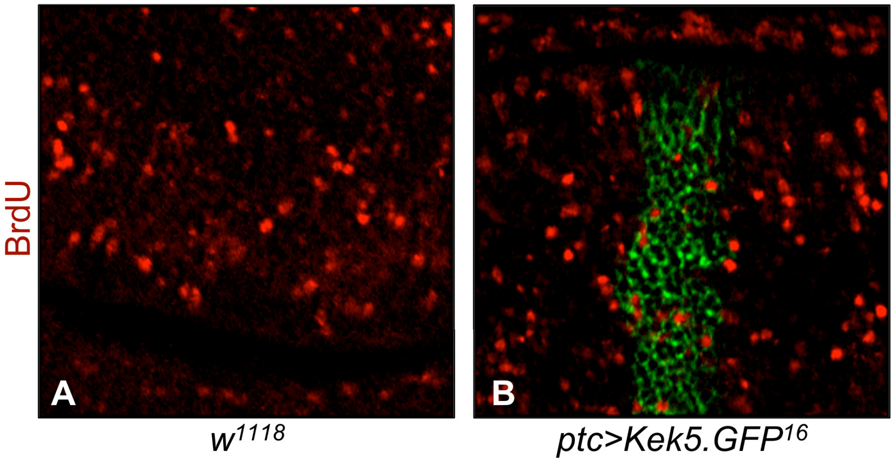


Figure 19: Kek5-induced cell enlargement is likely not related to cell cycle defects. When assessed for BrdU uptake (red) during DNA replication, *ptc-GAL4* driven Kek5.GFP does not appear to confer defects in cell cycle regulation (A, B). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 20X dry objective.

Structural dissection of Kek5 correlates cellular defects with adult phenotypes

If Kek5 has the ability to modulate cellular adhesion, it follows that particular structural elements might be essential for this function. To assess the contribution of each motif to Kek5's ability to confer such cellular defects, each deletion variant was assessed for GFP localization, basal extrusion, and cell enlargement (Table 6). Variants lacking the LRRs and entire IC domain show low levels of GFP expression and are not well-localized to the membrane, likely accounting for their respective lack of activity with respect to wild type Kek5 activity in the adult and larval wing tissues. Again, the Ig domain does not appear to be crucial for Kek5's ability to cause defects such as cellular extrusion and enlargement. Of interest is the PDZ domain-binding site, which is sufficient to confer localization, but together with the extracellular domain, is not sufficient to induce such cellular phenotypes. As before, the additional IC deletions do not give way to a straightforward interpretation of the contribution of the IC motifs to Kek5 function, yet confirm their significance in some capacity. Together, these structural dissections point to the LRRs and PDZ domain binding

Table 6: Structure-function dissection of Kek5 localization and cellular phenotypes. All Kek5 deletion variants were misexpressed with *ptc-GAL4* and assessed for GFP localization, basal extrusion, and apical cell enlargement. Localization is scored on a binary basis, while extrusion and cell enlargement are ranked from wild type (-) to severe (+++).

[*ptc-GAL4*] x	**Localization**	**Extrusion**	**Enlarged Cells**
Kek5$^{16\ \mathrm{II}}$	+	++	++
Kek5$^{16,52\ \mathrm{II}}$	+	+++	+++
Kek5 ΔL$^{16\ \mathrm{III}}$	-	-	-
Kek5 ΔL$^{43\text{-}1,16\ \mathrm{III}}$	-	+	+
Kek5 ΔI$^{13\text{-}2\ \mathrm{II}}$	+	++	++
Kek5 ΔI$^{27,13\text{-}2\ \mathrm{II}}$	+	+++	+++
Kek5 ΔIC$^{11\ \mathrm{III}}$	-	-	-
Kek5 ΔIC+PC$^{11\ \mathrm{I}}$	+	-	-
Kek5 Δ123$^{6\ \mathrm{II}}$	+	+	-
Kek5 Δ45$^{25\ \mathrm{II}}$	+	+	+

site as crucial for protein stability and localization, while additional IC motifs likely play a role in Kek5's ability to modulate cellular adhesion.

Kek5 misexpression triggers Armadillo upregulation

Although overt defects in Shg expression were not observed upon Kek5 misexpression, additional markers for junctional morphology were assayed. It is possible that Kek5 does not alter expression of a specific transmembrane adhesion molecule, but rather functions through interactions with cytoplasmic junction components that are critical in forming dynamic connections between the transmembrane molecules and the actin cytoskeleton. As another core component of the cadherin-catenin complex, which plays an integral role in linking extracellular adhesion to the intracellular cytoskeleton, Armadillo represented another molecule of interest. Surprisingly, despite the fact that Arm localization appears normal, Kek5-misexpressing cells demonstrate a dramatic increase in membrane-localized Arm levels (Fig 20). It is also worth noting that when viewed as a cross-section, it appears as though Kek5 and Arm do not colocalize, and the excess Arm accumulates apically to the region of normal localization (Fig 20C' inset). Importantly, this implies that Kek5 does not

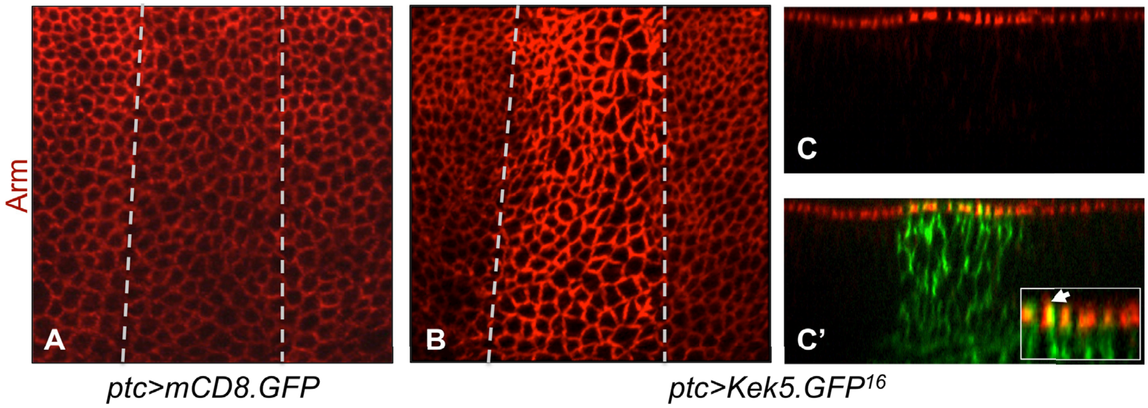


Figure 20: Kek5 misexpression induces an increase in Arm levels. Misexpression of Kek5.GFP with *ptc-GAL4* causes an increase of apical membrane-bound Arm (red; A, B). Arm does not colocalize with Kek5.GFP (C, C'), and appears to accumulate apically in the region of Kek5 misexpression (C' inset). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective

cause Arm accumulation through direct binding, but rather by some mechanism of indirect recruitment. This effect was also confirmed in other 3rd instar imaginal discs (Fig 21).

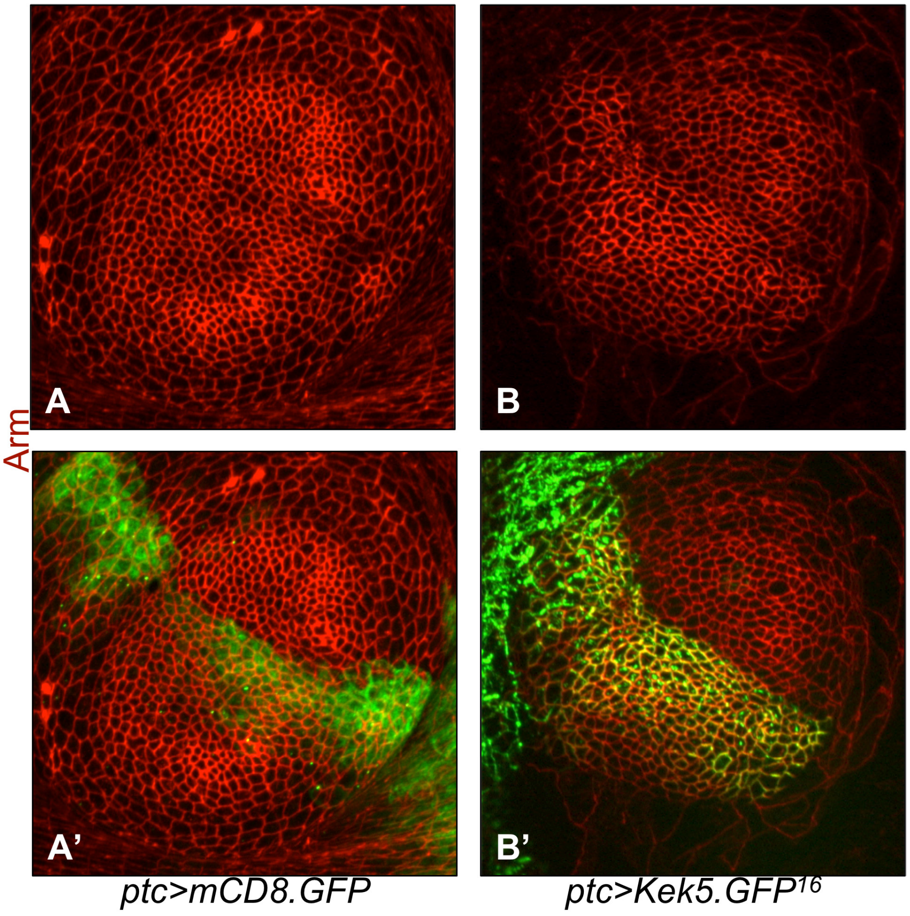


Figure 21: Kek5 misexpression induces Arm upregulation in other larval tissues. When misexpressed with *ptc-GAL4* in leg discs, Kek5.GFP triggers an increase in Arm levels (red; A, B). Fluorescent micrographs of 3rd instar leg imaginal discs acquired with 40X oil objective.

Although Kek5 mutants do not appear to exhibit phenotypes nearly as strong as those observed in gain-of-function assays, loss-of-function analysis was conducted with respect to Arm levels. Neither the *kek5* null mutant (*kek5*^{fe148}) nor *ptc-GAL4*-driven *kek5* RNAi demonstrated obvious alteration of Arm levels, although *kek5* RNAi may induce very mild upregulation (Fig 22A and B). It was also noted that a number of RNAi-knockdown discs exhibited two consistent morphological indentations in the middle of the disc proper on the apical surface (Fig 22C and D). Although this was not observed in the *kek5* null, one possibility is that the effect is due to the juxtaposition of Kek5 expressing and nonexpressing cells in the RNAi background. It was confirmed that *kek5* RNAi does knock down misexpressed Kek5.GFP; however, in the peripodial region of Kek5 misexpression, a number of cells with the previously described neuron-like extensions remained GFP positive (Fig 22E-F'), suggesting that perhaps these cells have somehow migrated away from the region of *ptc-GAL4* expression. Despite the lack of a definitive loss-of-function effect, the correlation between elevated Kek5 levels and increased membrane-bound Arm is a significant step in honing in on Kek5's functions and interactors.

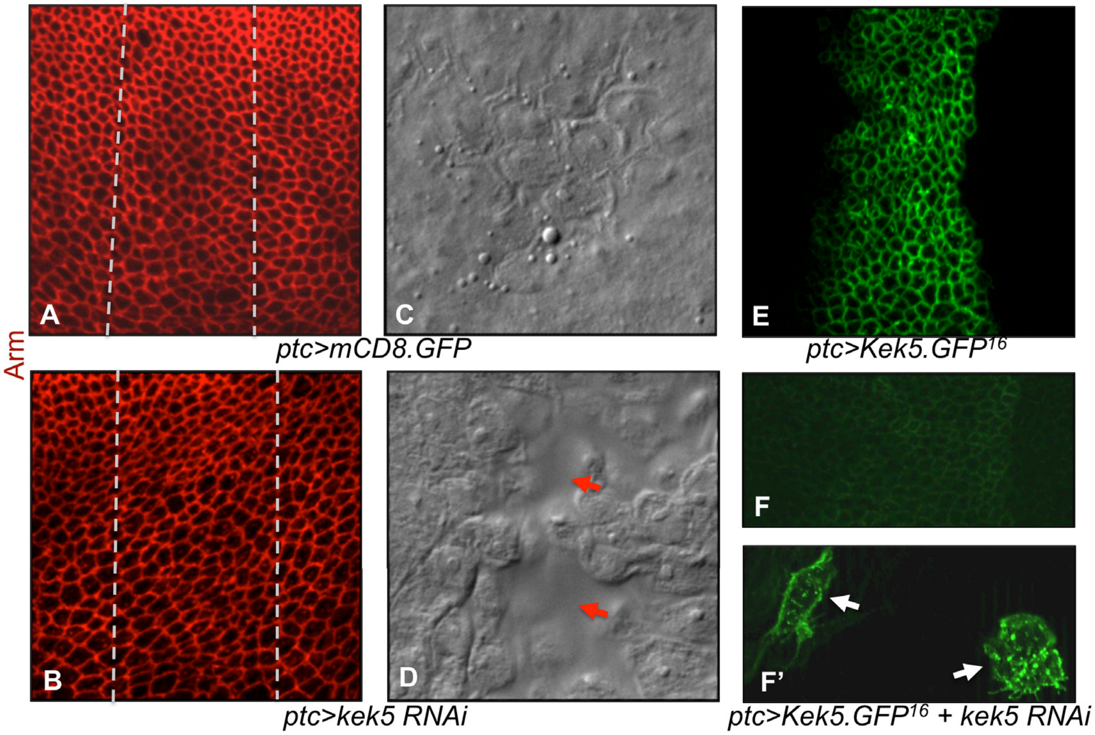


Figure 22: Kek5 loss of function does not affect Arm levels, but may cause morphological defects. Expression of *kek5* RNAi with *ptc-GAL4* does not appear to modulate Arm levels (red; A, B). However, *kek5* RNAi may induce morphological defects on the apical surface of the disc (C, D; red arrows). Additionally, *kek5* RNAi efficacy is shown by nearly complete knockdown of Kek5.GFP (E, F). Interestingly, some peripodial cells remain with unusually high levels of GFP (F'; white arrows). Fluorescent and DIC micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective. Cross-sections processed with Axiovision.

Kek5's Ig domain and PDZ domain-binding site are essential for Arm upregulation

If Kek5 has the ability to indirectly recruit Arm to the apical membrane, it is likely that particular domains and motifs would be responsible for this function. As before, a structure-function analysis was carried out by expressing all Kek5 variants with *ptc-GAL4* and Arm levels were assessed (Table 7). Surprisingly, the Ig domain appears to be crucial for Arm upregulation, as its deletion abrogates Kek5's ability to increase Arm levels at the membrane. Similarly striking was that with respect to the IC domain, the PDZ domain-binding site alone is sufficient for Arm upregulation (Fig 23). These results are in contrast to previous phenotypes observed in terms of the contribution of the Ig and IC domains. The Ig domain is necessary for Arm upregulation, but is not implicated in any of the other cellular effects. With respect to the IC domain, the PDZ domain-binding site alone is sufficient to confer Arm upregulation, while in the case of cellular phenotypes, such as extrusion and cell enlargement, additional IC motifs are required. This suggests that the Kek5-mediated increase in Arm levels occurs through an entirely different mechanism from those leading to extrusion, enlarged cells, and adult phenotypes.

Table 7: Structure-function dissection of Kek5-induced Arm upregulation. Each Kek5 deletion variant was assessed for its ability to confer an increase in Arm levels. Variants were ranked either upregulated (+) or no effect (NE).

[*ptc-GAL4*] x	**Arm**
Kek5$^{16\ \mathrm{II}}$	+
Kek5$^{16,52\ \mathrm{II}}$	+
Kek5 ΔL$^{16\ \mathrm{III}}$	NE
Kek5 ΔI$^{13\text{-}2\ \mathrm{II}}$	NE
Kek5 ΔI$^{27,13\text{-}2\ \mathrm{II}}$	NE
Kek5 ΔIC$^{11\ \mathrm{III}}$	NE
Kek5 ΔIC+PC$^{11\ \mathrm{I}}$	+
Kek5 Δ123$^{6\ \mathrm{II}}$	+
Kek5 Δ45$^{25\ \mathrm{II}}$	+

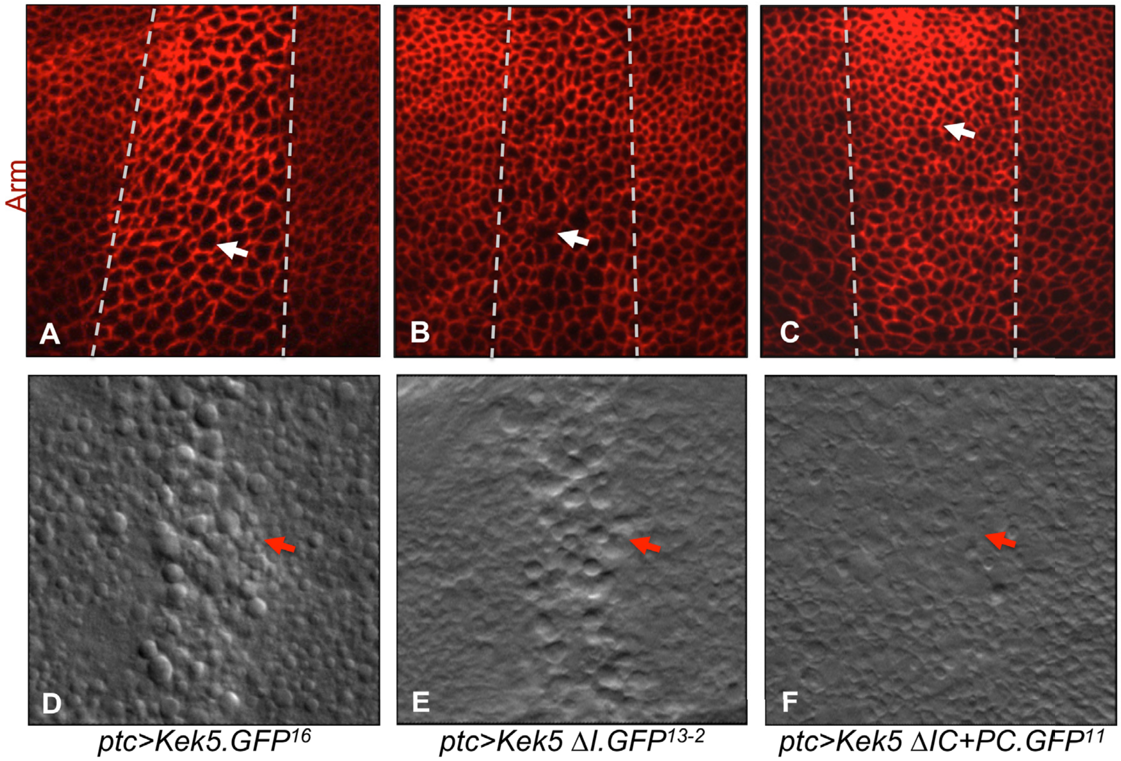


Figure 23: Kek5's ability to upregulate Arm is separable from its effects on cellular morphology. Full-length *ptc-GAL4*-driven Kek5.GFP induces Arm upregulation (red; A), cell enlargement (A; white arrow), and basal extrusion (D; red arrow). The Ig domain is crucial for Arm upregulation (B), but is not necessary for cell enlargement (B; white arrow) or basal extrusion (E; red arrow). Conversely, the PDZ domain-binding site is sufficient for Arm upregulation (C), but not for cell enlargement (C; white arrow) or basal extrusion (F; red arrow). Fluorescent and DIC micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective.

Adjacent cells must both misexpress Kek5 to confer Arm upregulation

Given that the Ig domain is critical for Kek5's ability upregulate Arm, and that Ig domains within specific adhesion molecules have been reported to mediate homophilic interactions between adjacent cells, the next aim was to determine whether Arm upregulation requires Kek5 expression in adjoining cells. To this end, Kek5-misexpressing clones were randomly generated as previously described to produce regions in the wing disc in which Kek5 is expressed in either single cells or small clusters of cells. Larvae were heat shocked either 24 or 48 hours prior to 3rd instar dissection to generate clones ranging from one to more than twenty cells in size. If homophilic adhesion does not play a role in Arm upregulation, it would be expected that Kek5-misexpressing cells would exhibit Arm upregulation around the entire membrane (Fig 24A). Conversely, if homophilic adhesion is crucial, only the interface between Kek5-misexpressing cells will show an increase in Arm levels (Fig 24B). Regardless of clone size, it appears that Kek5 is required to be in two adjacent cells in order to confer an increase in Arm levels (Fig 24C and D). However, it also appears that throughout the clone, Arm upregulation is not necessarily consistent, with some interfaces demonstrating strong upregulation and others yielding little to no change in Arm levels. As seen previously, clones generated in the peripodial membrane show extensive networks of neuron-like extensions, which again are not mirrored by visualization of the cell membrane (in this case, Arm staining; Fig 24E and E'). Together, the clonal data supports a mechanism that requires interaction between Kek5 on adjoining cells for upregulation of Arm, possibly through Ig domain mediated homophilic adhesion.

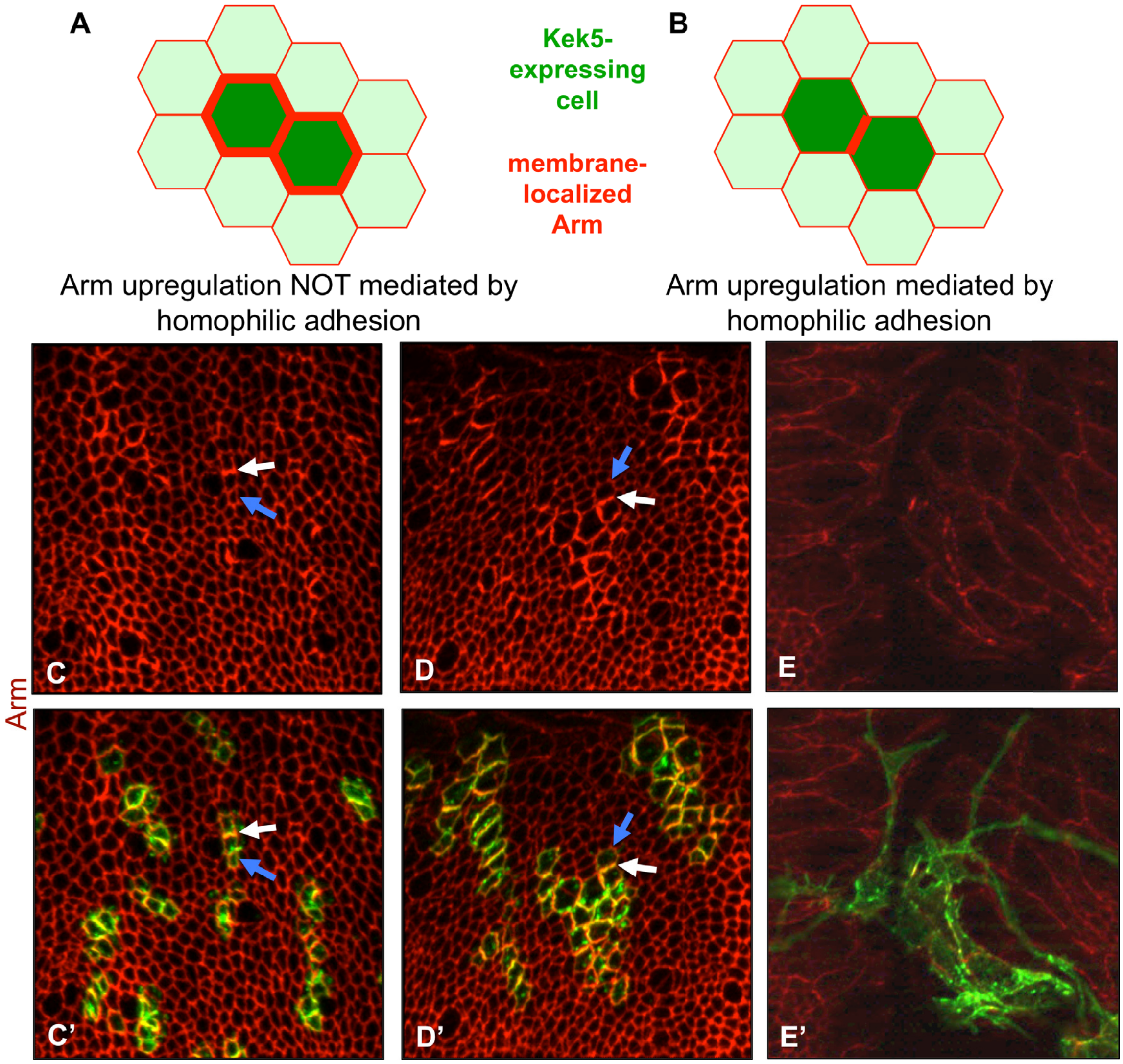


Figure 24: Arm upregulation occurs only at the interface between two Kek5-misexpressing cells. If homophilic adhesion is not critical, Arm upregulation should occur around the entire membrane of Kek5-misexpressing cells (A). If homophilic adhesion is necessary, only the interface between Kek5-misexpressing cells will exhibit Arm upregulation (B). Randomly induced clones using *hsFLP* and *Act-GAL4 (FRT-CD2)* demonstrate that expression of Kek5.GFP is necessary in two adjacent cells to induce Arm upregulation (red). The outer cell membranes of clones (blue arrows) never exhibit Arm upregulation, while a moderate percentage of membranes within the clones (white arrows) show elevated Arm levels. Clones were induced 24 (C, C') and 48 (D, D') hours prior to dissection to vary cell number. Although not obvious with Arm staining alone (E), peripodial clones exhibited an extensive network of neuron-like protrusions (E', 48 hour induction). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective.

Arm upregulation is specific to Kek5

Despite conservation within the family, Kek5 is the only member to exhibit such adhesion-related phenotypes (note that UAS-Kek3 transgenics have not yet been generated). If Kek5's ability to modulate Arm levels is unique within the family, it would provide additional insight into the contribution of its various sequence elements, with particular regard to the IC motifs (many of which are conserved among members of the Kek family). As expected, although all Keks except Kek6 localize in a similar manner to Kek5 under *ptc-GAL4* expression, Kek5 was the only family member able to induce an increase in Arm levels in the wing disc (Table 8).

Table 8: Specificity of Arm upregulation within the Kek family. All Kek family members were evaluated for GFP localization and modulation of membrane-localized Arm. Members were ranked either upregulated (+) or no effect (NE).

[*ptc-GAL4*] x	**Localization**	**Arm**
$Kek1^{59\ II}$	+	NE
$Kek2^{JDH2\ II}$	+	NE
$Kek4^{38\text{-}5\ II}$	+	NE
$Kek6^{12\ II}$	-	NE

Kek4/Kek5 swaps had been generated by Evans et al., as Kek4 is the most closely related member to Kek5 in terms of evolutionary divergence. These variants were all based on a Kek5 backbone and included the LRRs and/or the Ig of Kek4 to replace those of Kek5 (Fig 25). *Ptc-GAL4*-driven Kek5 K4L did not exhibit Arm upregulation despite appropriate localization, while the two other variants displayed poor localization and minimal effects on Arm levels (Table 9). Interestingly, the lack of Arm upregulation in Kek5 K4L indicates that the LRRs may play a role in mediating this effect. However, it is also possible that the LRRs of Kek4 simply prevent the native Ig domain of Kek5 from acting in its normal role.

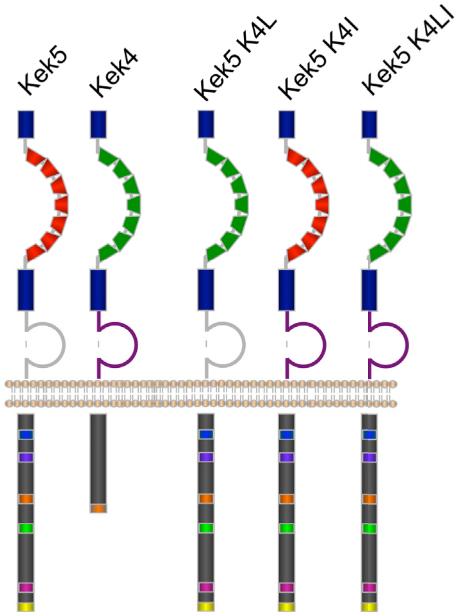


Figure 25: Kek4/Kek5 domain swaps. Swaps were generated on a Kek5 backbone, replacing the LRRs and/or Ig domain of Kek5 with those of Kek4.

Table 9: Specificity of Arm upregulation with regard to Kek4/Kek5 domain swaps. Swaps utilizing a Kek5 backbone and the LRRs and/or Ig domain of Kek4 were assessed in terms of GFP localization and ability to upregulate Arm. Variants were ranked either upregulated (+) or no effect (NE). Parentheses indicate a subtle effect/uncertainty.

[*ptc-GAL4*] x	**Localization**	**Arm**
Kek5 K4L$^{47\ I}$	+	NE
Kek5 K4I$^{5\text{-}1III}$	(+)	(+)
Kek5 K4LI^{57III}	-	NE

The poor localization of the two latter swaps is likely due to altering essential structural elements (i.e. LRRs and Igs), which causes the molecule in question to be misfolded. To better assess this, the Kek4/Kek5 swap of principal interest, Kek5 K4I, was misexpressed in ovarian follicle cells by *CY2-GAL4* to more closely ascertain localization. Kek5 K4I did not localize to the membrane as faithfully as wild type Kek5, but was predominantly membrane localized, indicating that the minimal effects on Arm levels in wing discs were not simply due to protein misfolding (Fig 26). At this stage, further studies will be necessary to determine the specific role of Kek5's Ig domain in Arm upregulation.

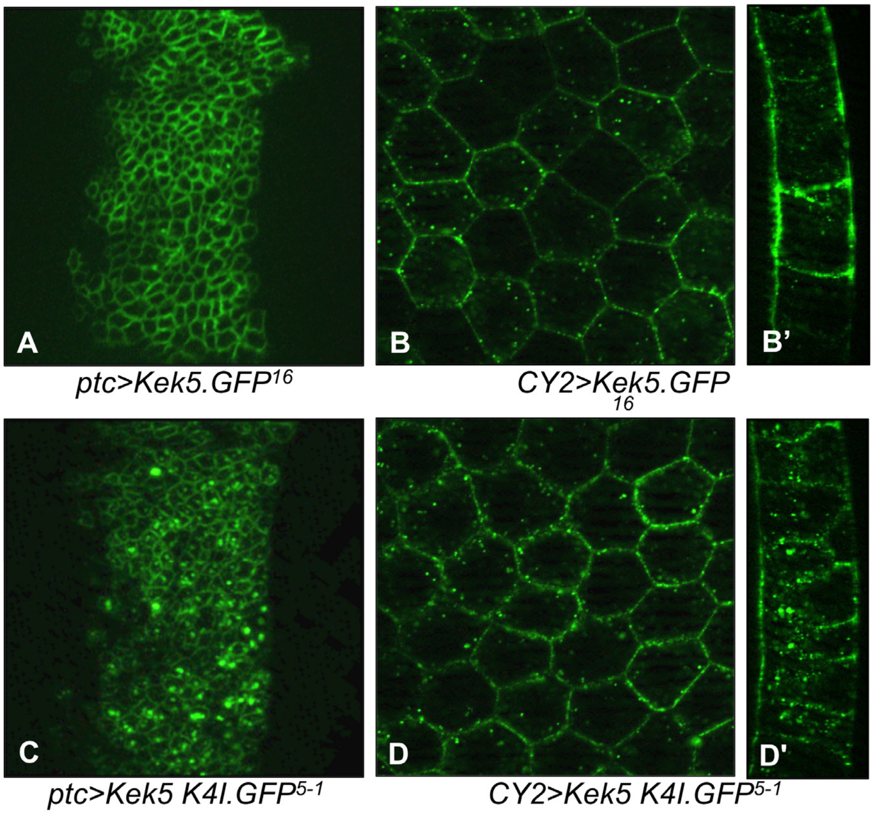


Figure 26: Kek5 K4I localizes poorly but is not misfolded. Full-length Kek5.GFP localizes to the membrane in both wing discs (A) and in the follicle cells of egg chambers (B, B'). Kek5 K4I.GFP is largely vesicular in wing discs (C), but is relatively well membrane-localized in egg chambers (D, D'). B and D are surface views, while B' and D' are lateral cross-sections of the same egg chamber. Fluorescent micrographs of 3rd instar wing imaginal discs and stage 10 egg chambers acquired with 40X oil objective.

Arm upregulation is not mediated through Pyd

Given that the PDZ domain-binding site of Kek5 is crucial for elevated Arm levels at the membrane and that the mechanism is likely indirect, it is probable that it is mediated through a molecule containing a PDZ domain. As such, the next objective was to identify a PDZ domain-containing protein in apical epithelial junctions that might link Kek5 to the modulation of Arm levels. Polychaetoid (Pyd) was the first molecule of interest for several key factors: (1) Pyd has PDZ domains and an SH3 domain (which could interact with Kek5's proline-rich IC motif 5); (2) Pyd is localized to apical epithelial junctions; (3) loss of Pyd has been linked to Arm upregulation in the eye; and (4) a genomic region containing Pyd was identified as a putative enhancer of *ptc-GAL4*-driven Kek5 activity with respect to thoracic bristle duplication (H. Haridas, personal communication).

Together these results support a model in which Kek5 and Pyd may interact, as phenotypes induced by misexpressed Kek5 resemble those resulting from reduced Pyd levels. For example, misexpression of Kek5 could titrate Pyd away from its normal activity through direct association via the PDZ binding interaction. In this model, Kek5's effects would be dependent on Pyd activity. This model makes two predictions: (1) loss of Pyd in the wing disc would result in Arm upregulation, and (2) increased expression of Pyd would suppress Kek5 mediated Arm upregulation. To test this model *UAS-pyd RNAi* and *UAS-Pyd.RFP* lines were obtained, courtesy of R. Cagan. Pyd loss- and gain-of-function were assessed in a background with and without Kek5 misexpression (in a *ptc-GAL4* dependent manner) to determine whether the loss or gain of Pyd, or simultaneous modulation of Pyd and Kek5 levels could alter the elevated Arm phenotype. However, neither loss nor gain of Pyd were able to modulate Arm levels in the wing disc, whether alone or in the presence of misexpressed Kek5 (Fig 27), indicating that Kek5's effect on Arm levels is not mediated via an interaction with Pyd.

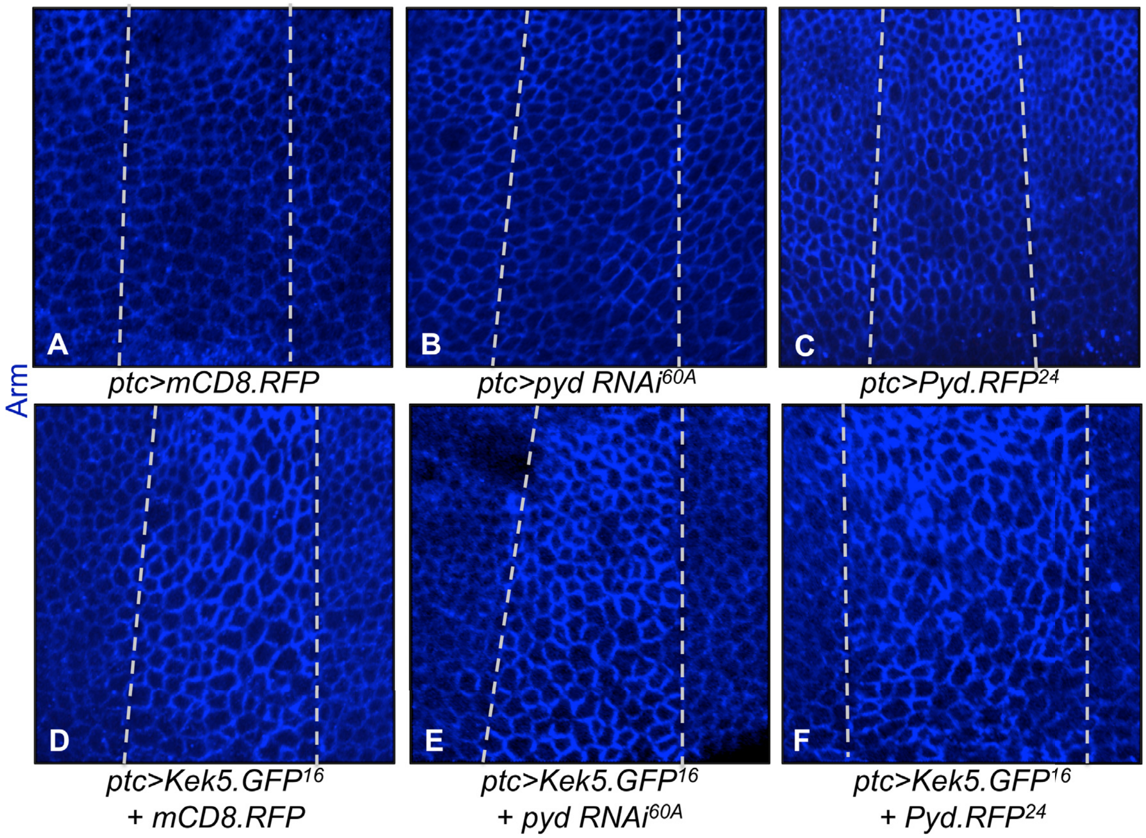


Figure 27: Kek5-mediated Arm upregulation is not affected by Pyd levels. Neither *pyd* RNAi nor Pyd.RFP alone affect Arm levels (blue; A-C). Similarly, neither is able to further modulate Kek5.GFP-induced Arm upregulation (D-F). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective.

As noted previously in regard to the Kek5 modifier screen, a hallmark Pyd mutant phenotype is thoracic/scutellar bristle duplication. This phenotype also occurs in *ptc-GAL4*-driven Kek5 adults and appears to be dominantly enhanced by loss of *pyd*. To confirm the separability of Kek5-induced Arm upregulation and the Pyd-related adult phenotypes, wild type Kek5, Kek5 ΔIg, and Kek5 ΔIC+PC (PDZ only) were assessed for adult bristle duplication (Fig 28). A wild type scutellum harbors 4 bristles, whereas misexpression of Kek5 results in an average of 14 bristles (Table 10). Misexpression of Kek5 ΔIg results in an effect similar to that of wild type Kek5, while Kek5 ΔIC+PC yields the normal number of bristles. Given that the Ig domain is essential for upregulation of Arm, but not for bristle duplications, this data supports the notion that Kek5's putative interaction with Pyd are separable from its ability to modulate Arm levels, and further argues that Pyd is unlikely to serve as a link between Kek5 and Arm upregulation.

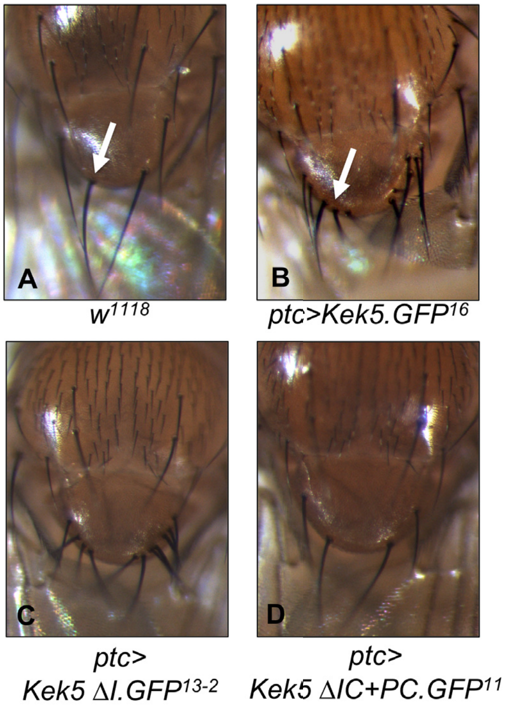


Figure 28: Kek5-induced Arm upregulation and Pyd-like effects are separable. Full-length *ptc-GAL4*-driven Kek5 causes significant duplication of scutellar bristles (A, B; compare white arrows). Kek5 ΔIg yields a comparable number (C), while Kek5 ΔIC+PC appears wild type (D). Bright field micrographs of adults acquired with 10X objective.

Table 10: Structure-function analysis of Kek5 with regard to Pyd-like adult phenotypes. Kek5 variants were driven with *ptc-GAL4* and assessed for scutellar bristle (SB) duplication, a phenotype potentially indicative of altered Pyd levels.

[*ptc-GAL4*] x	Avg. # SB	n (flies)
Kek5$^{16\ \mathrm{II}}$	14	52
Kek5 ΔI$^{13\text{-}2\ \mathrm{II}}$	10	47
Kek5 ΔIC+PC$^{11\ \mathrm{I}}$	4	108

Kek5 misexpression affects additional AJ components

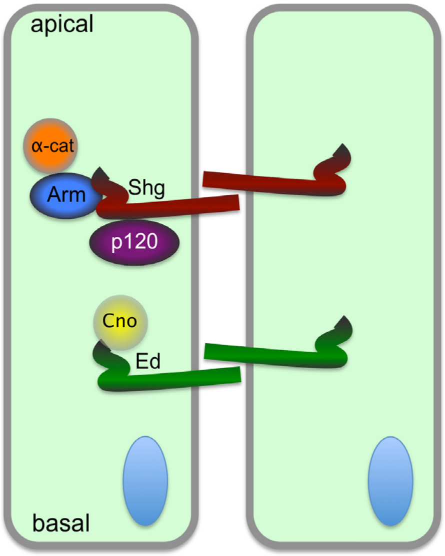


Figure 29: *Drosophila* Adherens Junction architecture. AJs contain two main complexes. The first is the cadherin-catenin complex, with transmembrane Shotgun and intracellular Armadillo, α-catenin, and p120-catenin. The second is the nectin-afadin complex, with transmembrane Echinoid and intracellular Canoe.

The dynamic, tightly regulated adherens junction architecture suggests that if Kek5 can modulate one AJ component, it may be able to affect others as well. If other components are indeed altered by Kek5 misexpression, we may be able to create a model that links an increase in Kek5 to elevated Arm levels. In terms of adherens junction architecture, there are two main complexes: the cadherin-catenin (Shg & Arm) complex and the nectin-afadin (Ed & Cno) complex (Fig 29). These are cooperative but separate entities at the AJ, and may be transiently linked by Bazooka, a PDZ domain-containing junction molecule that is able to associate with the PDZ domain-binding sites of both Arm and Ed.

Components of the cadherin-catenin complex were first examined with respect to Kek5 misexpression. As noted previously, Kek5-misexpressing cells do not show obvious modulation of Shg levels. It is possible that Shg is mildly upregulated, but this effect is not as consistent, and is certainly not as strong as the increase in Arm levels (Fig 30A-C'). When tested with the two most critical Kek5 variants, Kek5 ΔIg and Kek5 ΔIC+PC, it is interesting to note that deleting the Ig domain does not appear to modulate Shg levels, while deleting the IC domain, shows slight but definite upregulation (Table 11). Similarly, misexpression of Kek5 may increase p120-catenin levels, but again, such an effect, if any, is extremely mild (Fig 30D-F'). p120 levels were also assessed in response to misexpression of Kek5 ΔIg and Kek5 ΔIC+PC, both of which showed normal levels of p120 (Table 11). Assessment of α-catenin was also attempted, but the antibody (DSHB) proved to stain poorly in both imaginal disc tissue and egg chambers (data not shown and M. Arata, personal communication, respectively). Together, these results suggest that unlike Arm, currently assessed components of the cadherin-catenin complex do not display strong expression effects in response to Kek5. Thus, Kek5's effect on Arm is likely not mediated through the cadherin-catenin complex.

To examine other potential candidates, the nectin-afadin complex was next examined in Kek5-misexpressing cells. Interestingly, Cno (afadin) shows marked downregulation in the region of Kek5 misexpression (Fig 30G-I'). Preliminary assessments indicate that Kek5 may not interact with Ed (nectin). Although Pyd is not truly part of the nectin-afadin complex, it has been reported to bind Cno, and as noted previously, misexpression of Kek5 in adult tissues yields bristle duplication phenotypes similar to that of Pyd mutants. As with Cno, Pyd levels are lowered in Kek5-misexpressing cells, and a structure-function analysis again suggests that the Ig is not necessary for the effects on either Cno or Pyd, while the PDZ domain is necessary but not sufficient to induce such modulation (Fig 30J-L', Table 11). This is in stark contrast to a strict requirement for the Ig domain on the extracellular side and within the IC domain only the PDZ binding motif in Arm upregulation. It is also worth noting that in all cases, while levels of junction components may be affected by Kek5 misexpression, the junction location and cell polarity appear to be normal. To ensure that the adjacent SAC region is not affected by Kek5 misexpression, EGFR was examined under Kek5 misexpression (Fig 30M-O'). As expected no obvious defects were noted, again suggesting that Kek5 is acting primarily at the adherens junctions. *Together, these data suggest that the nectin-afadin/Pyd complex may play a role in Kek5's ability to induce morphological defects at the cellular level, but is not likely to mediate its effects on Arm levels.*

Table 11: Structure-function analysis of Kek5 with regard AJ architecture. Kek5 variants were driven with *ptc-GAL4* and assessed for modulation of a number of AJ components. Variants were ranked upregulated (+), downregulated (-), or no effect (NE). NT indicates a variant not tested. Parentheses indicate a subtle effect/uncertainty.

[*ptc-GAL4*] x	**Arm**	**Shg**	**p120**	**Cno**	**Pyd**	**EGFR**
Kek5$^{16\ \mathrm{II}}$	+	(+)	(+)	-	-	NE
Kek5 Δ$\mathrm{I}^{13\text{-}2\ \mathrm{II}}$	NE	NE	NE	-	-	NT
Kek5 ΔIC+$\mathrm{PC}^{11\ \mathrm{I}}$	+	+	NE	NE	NE	NT

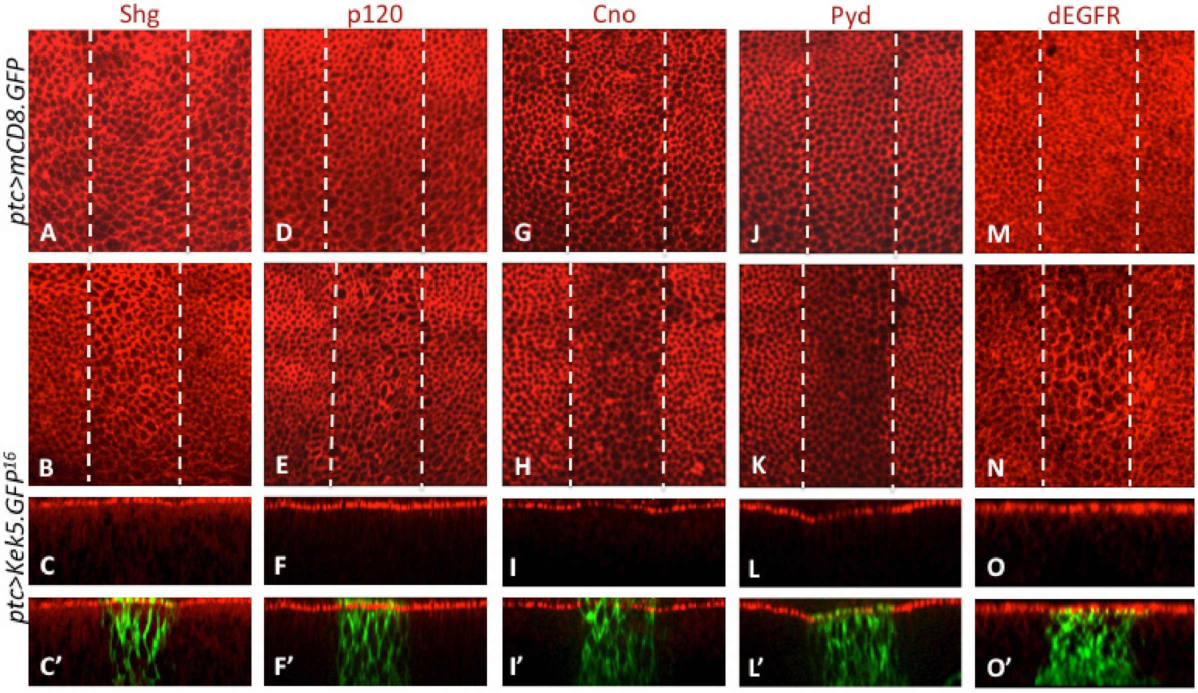


Figure 30: Kek5 misexpression affects additional architectural components. *ptc-GAL4*-driven Kek5.GFP does not induce obvious changes in Shg (red; A-C') or p120-catenin (red; D-F') levels, although there may be minor upregulation of each. In contrast, Cno (red; G-I') and Pyd (red; J-L') appear downregulated upon Kek5 misexpression. dEGFR, a subapical component (red; M-O') is not affected by Kek5 misexpression, as expected. In all cases, localization and cell polarity appear normal (C, F, I, L, and O are cross-sections with the indicated staining only; C', F', I', L', and O' show the same cross-section and staining with a GFP overlay). Fluorescent micrographs of 3rd instar wing imaginal discs acquired with 40X oil objective. Cross-sections processed with Axiovision.

Transfection of Kek5 does not recapitulate Arm upregulation in cell culture

Although a number of AJ components are affected by Kek5 *in vivo*, none of these molecules appears to be the basis for Kek5's ability to upregulate Arm. If these effects could be recapitulated in cell culture, it would then be possible to do a genome-wide screen for molecules essential for Kek5's ability to upregulate Arm. This could be accomplished through simultaneous transfection of Kek5.GFP and gene specific RNAi-induced knockdown to determine whether misexpressed Kek5 remains able to induce Arm upregulation in the absence of the targeted gene.

Schneider's embryonic S3 (Fig 31) cells were cultured under normal conditions and transfected with Kek5.GFP, as well as Kek1.GFP as a negative control. The cells were resuspended and fixed to poly-L-lysine-treated multi-well slides and stained for Arm, Shg, and p120 (Fig 32). In wild type S3 cells (no Kek5 misexpression), each junction component was not recruited to the membrane unless in contact with another cell, and although in many cases where there was cell-cell contact, membrane localization was absent. Kek5.GFP-expressing cells were unable to induce the same upregulation of Arm observed in larval tissue. With the knowledge that Arm upregulation in wild type S3 cells appears to be triggered by adhesion with adjacent cells, wild type and Kek5-GFP-transfected cells were also grown on a coverslip to avoid the need to resuspend the cells before fixing and staining. However, even under these conditions, Kek5.GFP misexpressing cells were still unable to recapitulate an increase in membrane-bound Arm (Fig 33). With the thinking that perhaps the proper molecular machinery required for this effect is not present in S3 cells, a third instar wing imaginal disc line (Clone8, or CL.8) was obtained and similarly tested (Fig 31). Given that Kek5 affects Arm in wing tissue, it would be expected that these cells have the appropriate molecular components to confer this effect. Unfortunately, as with S3 cells, misexpression of Kek5 did not induce recruitment of Arm to the membrane (Fig 34).

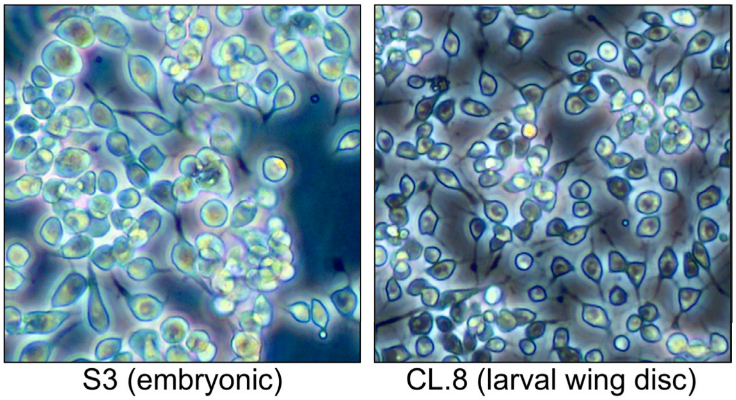


Figure 31: Morphology of S3 and Clone8 cell lines. Both *Drosophila* cell lines are semi-adherent and exhibit varied morphologies ranging from circular to neuron-like. Phase contrast micrographs acquired with inverted 20X objective plus 3X optical camera zoom.

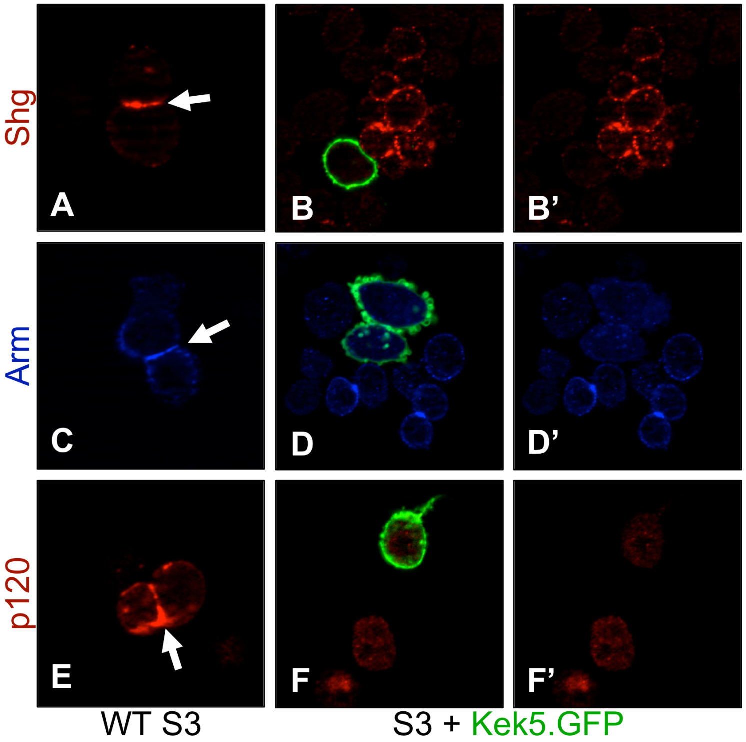


Figure 32: Transfection of Kek5 does not affect AJ components in resuspended S3 cells. In wild type S3 cells, Shg (red; A), Arm (blue; C), and p120 (red; E) occasionally exhibit membrane enrichment upon cell-cell contact (white arrows). Transfection of Kek5.GFP does not induce an observable change in levels of Shg (B, B'), Arm (D, D'), or p120 (F, F'). Fluorescent micrographs of resuspended S3 cells acquired with 40X oil objective.

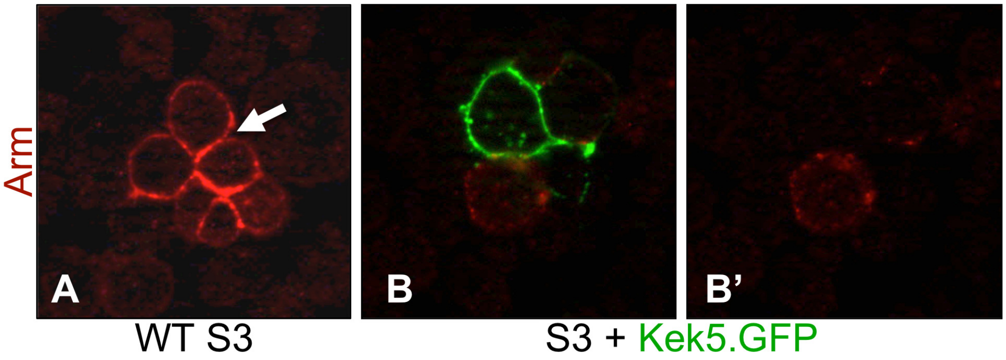


Figure 33: Grown S3 cells do not exhibit Arm upregulation upon Kek5 transfection. Wild type S3 cells still only exhibit occasional membrane Arm enrichment (red) when grown on a coverslip (A; white arrow). Similarly, transfection of Kek5.GFP still does not cause Arm upregulation (B, B'). Fluorescent micrographs of S3 cells grown on a coverslip acquired with 40X oil objective.

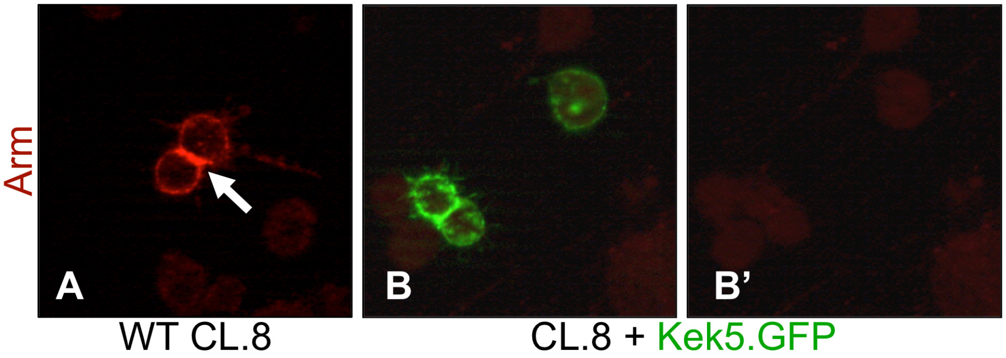


Figure 34: Transfection of Kek5 does not induce Arm upregulation in Clone8 cells. Wild type wing disc-derived Clone8 cells exhibit occasional membrane Arm enrichment (red) after resuspension (A; white arrow). Similar to S3 cells, transfection of Kek5.GFP does not induce membrane Arm enrichment (B, B'). Fluorescent micrographs of resuspended Clone8 cells acquired with 40X oil objective.

This assay may still be possible with another tissue culture cell type and conditions, and perhaps primary culture of Kek5-misexpressing wing disc tissue would be the key to recapitulating effects observed *in vivo* (Bai et al., 2008). In preparation for a possible primary cell based screen all *Drosophila* PDZ domain-containing molecules were identified from the *Drosophila* database, given the importance of the PDZ domain-binding site in Arm upregulation (Fig 35). With this knowledge, it may be possible to use a combined cell culture and RNAi-based approach to identify putative Kek5 binding partners mediating its effect on Arm. As such, a reliable assay that recapitulates Kek5's ability to upregulate Arm must be developed first, whether it be locating an appropriate cell line or primary culture of Kek5-misexpressing disc tissue.

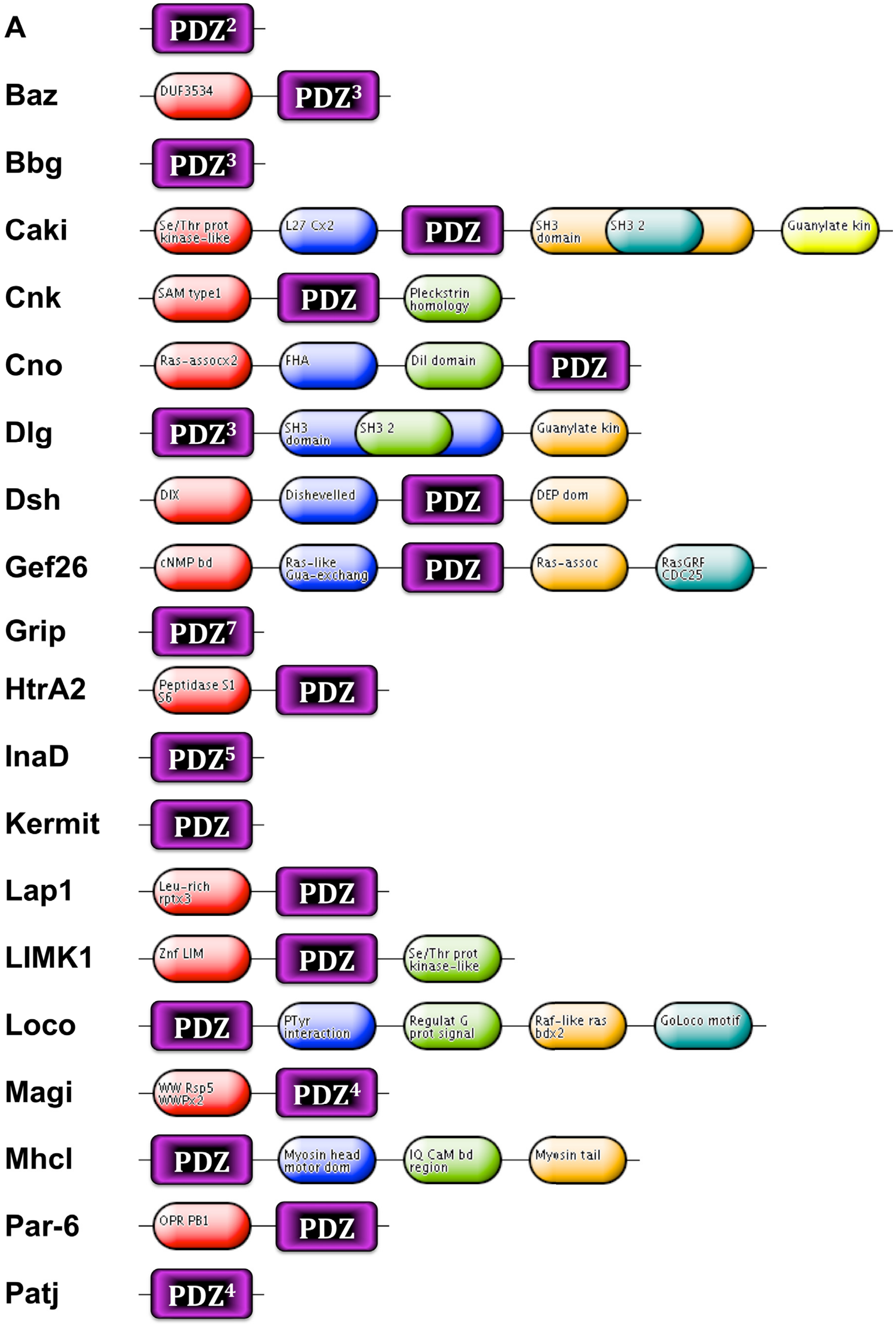


Figure 35: *Drosophila* PDZ domain-containing proteins. All known *Drosophila* PDZ domain-containing are shown with their respective PDZ domains, as well as any additional identifiable sequence elements.

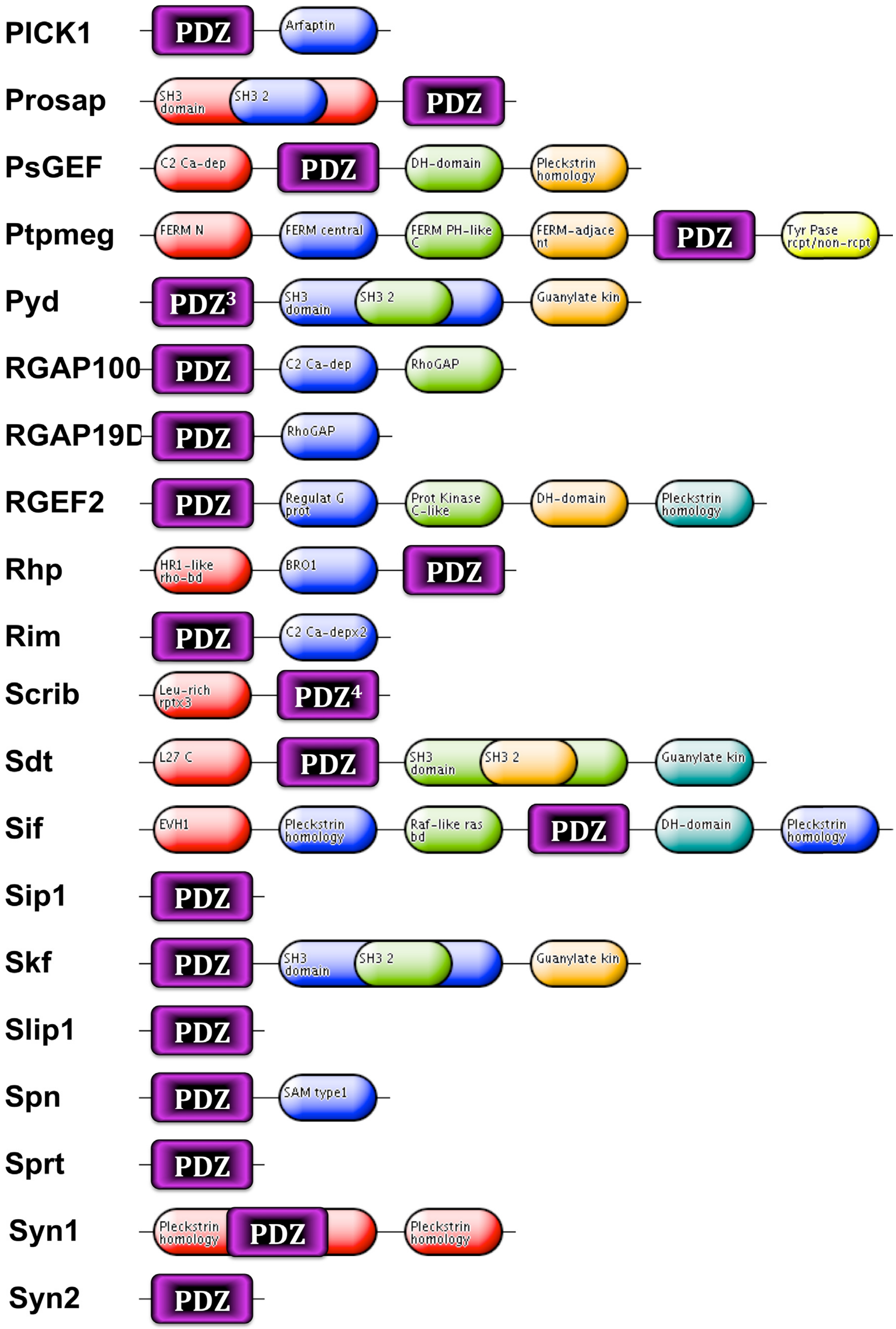


Figure 35 (continued)

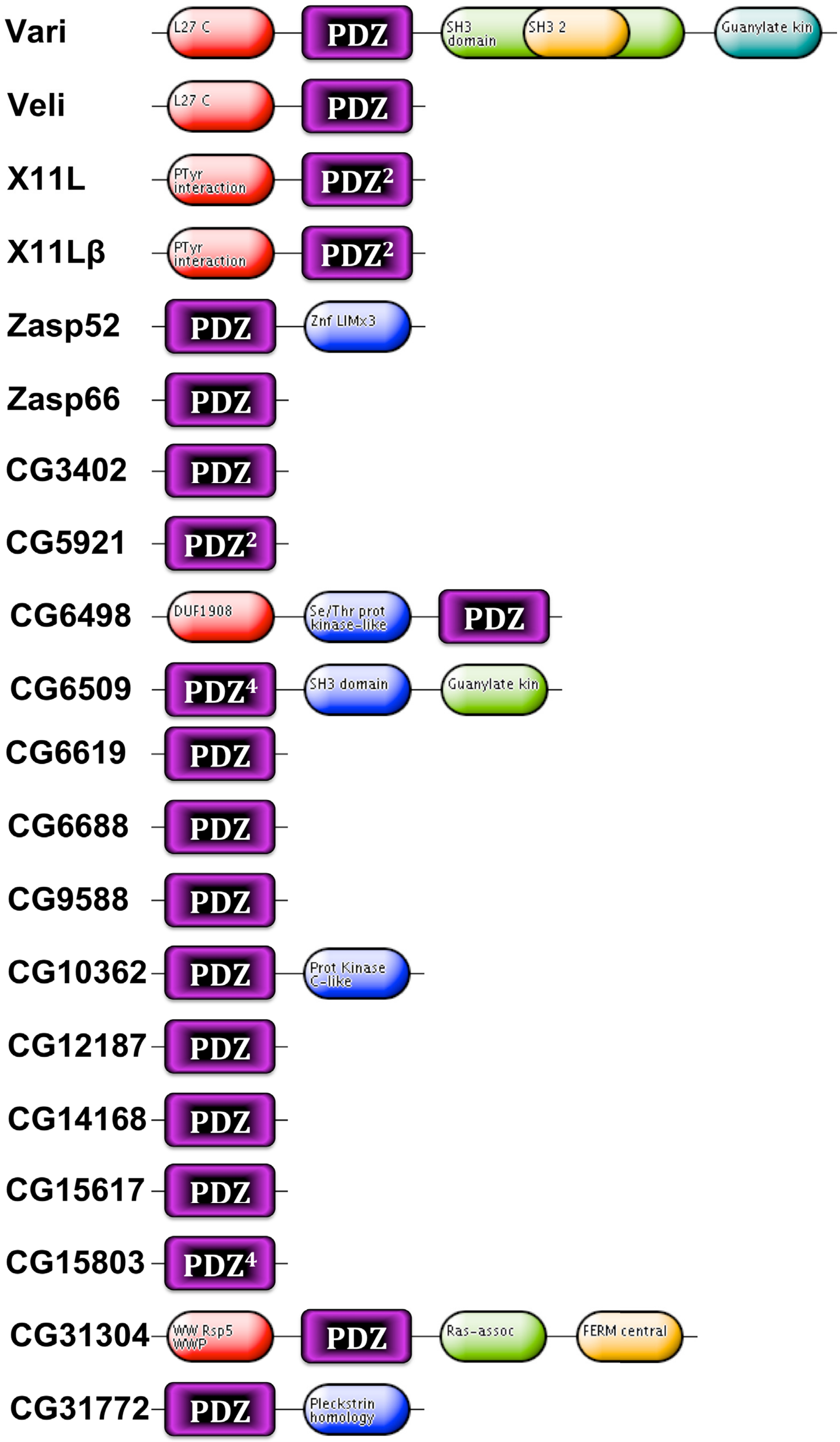


Figure 35 (continued)

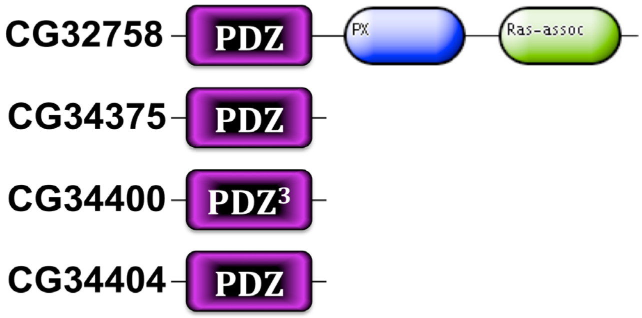


Figure 35 (continued)

Development of an *in vitro* high-throughput assay to identify LIG binding partners

With the importance of a select few LIGs in cell-cell communication, including Kek1 in EGFR signaling, Kek5 in BMP signaling and Arm upregulation, and Lingo1 in Nogo signaling, having been demonstrated, the next aim was to create an approach to more broadly assess the role(s) of the LIG family. Such an approach could also be used to identify molecules that interact with the Ig domain of Kek5 and mediate its ability to upregulate Arm. Such an effort would require the development of more rapid high-throughput assays to assess putative binding partners on a larger scale. Although tenuous, claims that EGFR (ErbB) family members interact with various LIG members, and our work with Kek1, make the EGFR family an appropriate starting point for developing such an assay. These assays can take one of two forms: (1) smaller scale co-immunoprecipitations with full-length transmembrane molecules or (2) larger scale ELISA-based assays with secreted versions of the proteins, as has been utilized in assessing Dscam isoform specificity (Wojtowicz et al., 2007).

Both assays make use of the GAL4-UAS system to express the molecules of interest in cell culture. As summarized in Fig 36, *metallothionine-GAL4* (*mt-GAL4*) and the construct(s) of interest are transiently transfected into S3 cells. Upon addition of $CuSO_4$, *mt-GAL4* is activated, which subsequently induces expression of the GFP- or V5/6XHis-tagged molecules of interest.

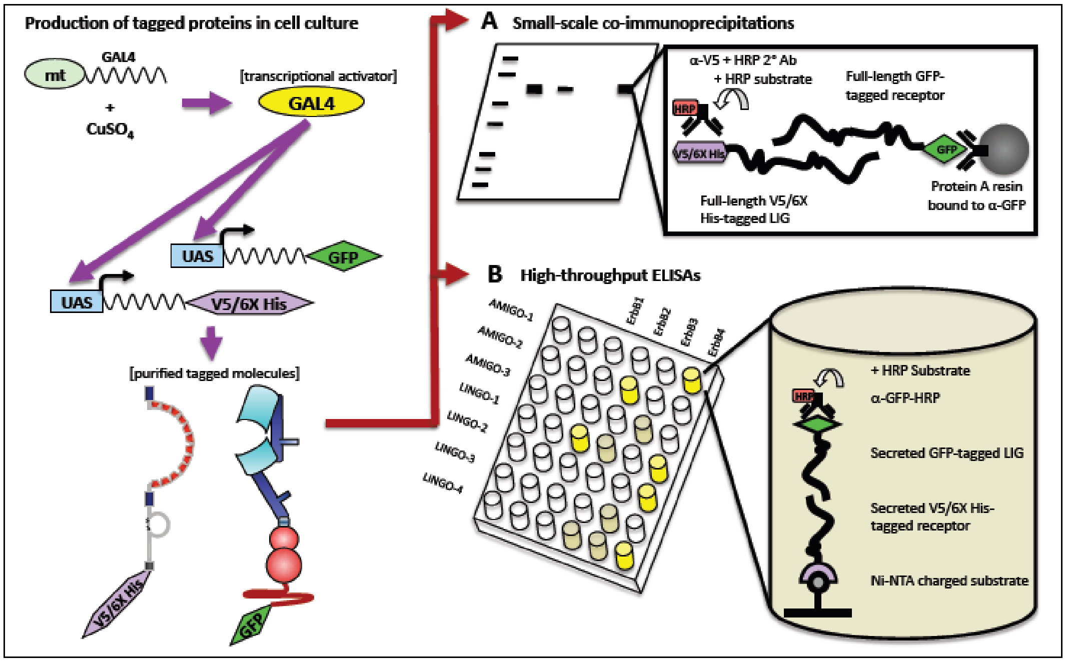


Figure 36: Development of *in vitro* assays for LIG binding partners. Proteins of interest are produced in cell culture under the expression of copper-inducible *mt-GAL4*. Full-length molecules will be used for co-immunoprecipitations (A), while secreted versions will be used in an ELISA-based assay (B).

To start the smaller-scale co-immunoprecipitations, a small number of human LIGs were chosen for binding analysis with the human ErbB family. GFP-tagged constructs of all four members of the ErbB family were previously generated in the Duffy lab. A list of currently available LIG cDNAs is included in Appendix A, as obtained from the NIH Mammalian Gene Collection and OpenBiosystems. Full-length versions of AMIGOs 1-3 and LINGO1 were cloned into *pUAST-a-V5/6XHis* using the Gateway recombination-based cloning system. Note that the attB recombination sites were modified in accordance with improved consensus sequences to increase cloning efficiency (Rual et al., 2004). A complete schematic of the altered attB sites can be found in Appendix B. Two of these constructs, LINGO1 and AMIGO3 have been successfully transfected into S3 cells and detected in cell lysates on a Western blot with monoclonal anti-V5. Once all constructs have been transfected and detected in sufficient quantities, co-immunoprecipitations may be completed. For this assay, GFP-tagged EGFRs will be pulled down with polyclonal anti-GFP and assessed for the presence of V5-tagged LIGs.

In preparation of a larger scale ELISA-based screen, several secreted control constructs have been generated. Previous data using surface Plasmon resonance has demonstrated high affinity *in vitro* binding of secreted dEGFR with Kek1, but not Kek2, serving as appropriate positive and negative controls for assay development (D. Alvarado, personal communication). Again using the Gateway recombination cloning system, secreted Kek1, Kek2, and dEGFR were cloned into both *pUAST-a-GFP* and *pUAST-a-V5/6XHis*. Secreted Kek1.GFP has since been successfully transfected into S3 cells, and high expression levels have been detected in the supernatant on a Western blot with monoclonal anti-GFP (D. Afezolli, personal communication). Once all constructs have been transfected and harvested in the supernatant, ELISAs can be run. His-tagged dEGFR would be secured to a nickel-embedded 96-well plate, and assessed for Kek1.GFP and Kek2.GFP binding. Interactions may be quantified by fused GFP fluorescence, or by using HRP-conjugated antibodies and detecting substrate-induced chemiluminescence.

As it stands, most of the appropriate controls for both co-immunoprecipitations and ELISAs have been generated and successfully expressed; the assays can now be optimized and can then be utilized to start screening the LIG family more broadly.

Discussion

The LIG family is a novel class of transmembrane proteins that already shows intriguing promise as potential therapeutic targets for cancers and neurodegenerative diseases (Chen et al., 2006; MacLaren et al., 2004). Much has yet to be learned in regard to the overall role(s) of the family, but a number of human family members have been implicated in EGFR signaling, as well as neuronal outgrowth, synaptic plasticity, and oligodendrocyte differentiation. The common core of LRRs and Ig domains points to a number of possible cellular functions, such as adhesion, growth, and proliferation, likely mediated through homo- and heterophilic protein-protein interactions. The study of both human and *Drosophila* LIGs will lead to a better understanding of these processes, as well as the individual roles of its members.

Previous studies of the *Drosophila* LIG Kekkon5 have been suggestive of a number of functions, with special regard to BMP signaling and cellular adhesion (Evans et al., 2009). In this work, the role of Kek5 in epithelial junction architecture was examined in an effort to elucidate its function in cellular adhesion. At the cellular level, misexpression of Kek5 induces morphological defects such as epithelial extrusion and apparent cell enlargement. It was determined that the LRRs and PDZ domain-binding site are crucial for protein stability and membrane localization, respectively, while additional conserved intracellular motifs are required for 'normal' gain-of-function activity, both in adult and larval wing tissues.

Upon examination of intracellular adherens junction component Armadillo, part of a critical transmembrane complex linking extracellular cell-cell contacts with the cytoskeleton, as well as a key player in Wnt signaling, it was determined that

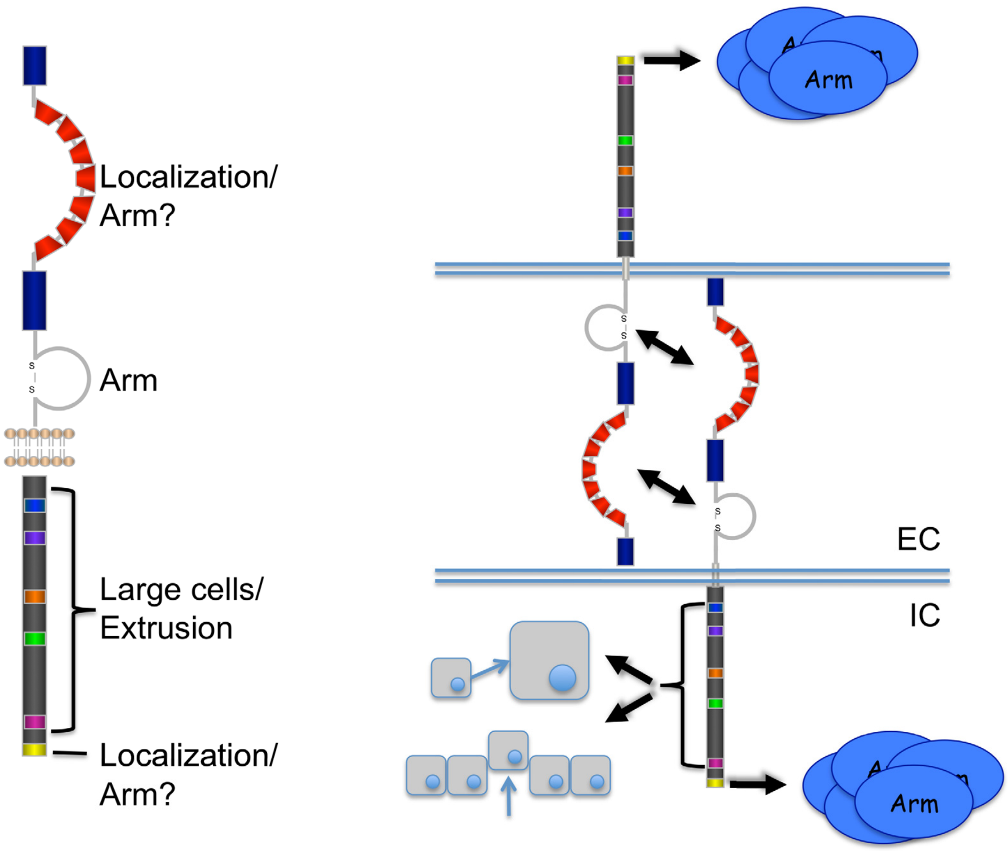


Figure 37: Kek5-mediated Arm upregulation is separable from its ability to confer cellular defects. Full-length misexpressed Kek5 induces both an increase in membrane-associated Arm, as well as cell enlargement and basal extrusion. Deletion of the Ig domain abrogates its ability to upregulate Arm, but still allows for it to confer other cellular effects. Conversely, the PDZ domain-binding site allows for recruitment of Arm to the membrane, but eliminates Kek5's activity in terms of altered cellular morphology.

misexpression of Kek5 dramatically increases its levels at the apical membrane. Given that Arm and misexpressed Kek5 do not colocalize, it seems likely that Kek5 mediates this effect indirectly. In contrast to other Kek5-induced cellular effects, the Ig domain is critical for such upregulation to occur, and in regard to the intracellular domain, the PDZ domain-binding site alone is sufficient to confer the effect (Fig 37). Particularly striking is the demonstration that the Ig domain is required for Arm upregulation as it provides for the first time a demonstrable function for an Ig domain among Kek family members. Interestingly, misexpressed Kek5 is also required in two adjoining cells in order for membrane-associated Arm levels to increase. This is consistent with published results demonstrating homo- and heterophilic interactions among Kek family members and supports a model for Kek5 function involving homophilic interactions (MacLaren et al., 2004).

A number of models for Arm upregulation are possible (Fig 38). One is that homophilic Ig-mediated binding occurs between Kek5-expressing cells, somehow directly enabling intracellular recruitment of Arm through an intermediate PDZ domain-containing molecule. Another possibility is that the PDZ domain-binding site is utilized solely for Kek5 localization, while extracellular Ig-mediated binding triggers an entirely separate cascade back into the cytoplasm, resulting in recruitment of Arm to the membrane. To begin to determine which, if either, model is correct, PDZ domain-binding site swaps with other Kek family members could be generated and tested for Arm upregulation. If Kek5's PDZ domain-binding site is critical in mediating Arm recruitment, it would be expected that Kek5 with other PDZ domain-binding sites would be unable to induce such an effect. Regardless, either model must also take into consideration that other Kek5-induced cellular effects do not require the presence of the Ig domain, but do function through some combination of intracellular motifs. Currently, these two mechanisms appear to be entirely separable.

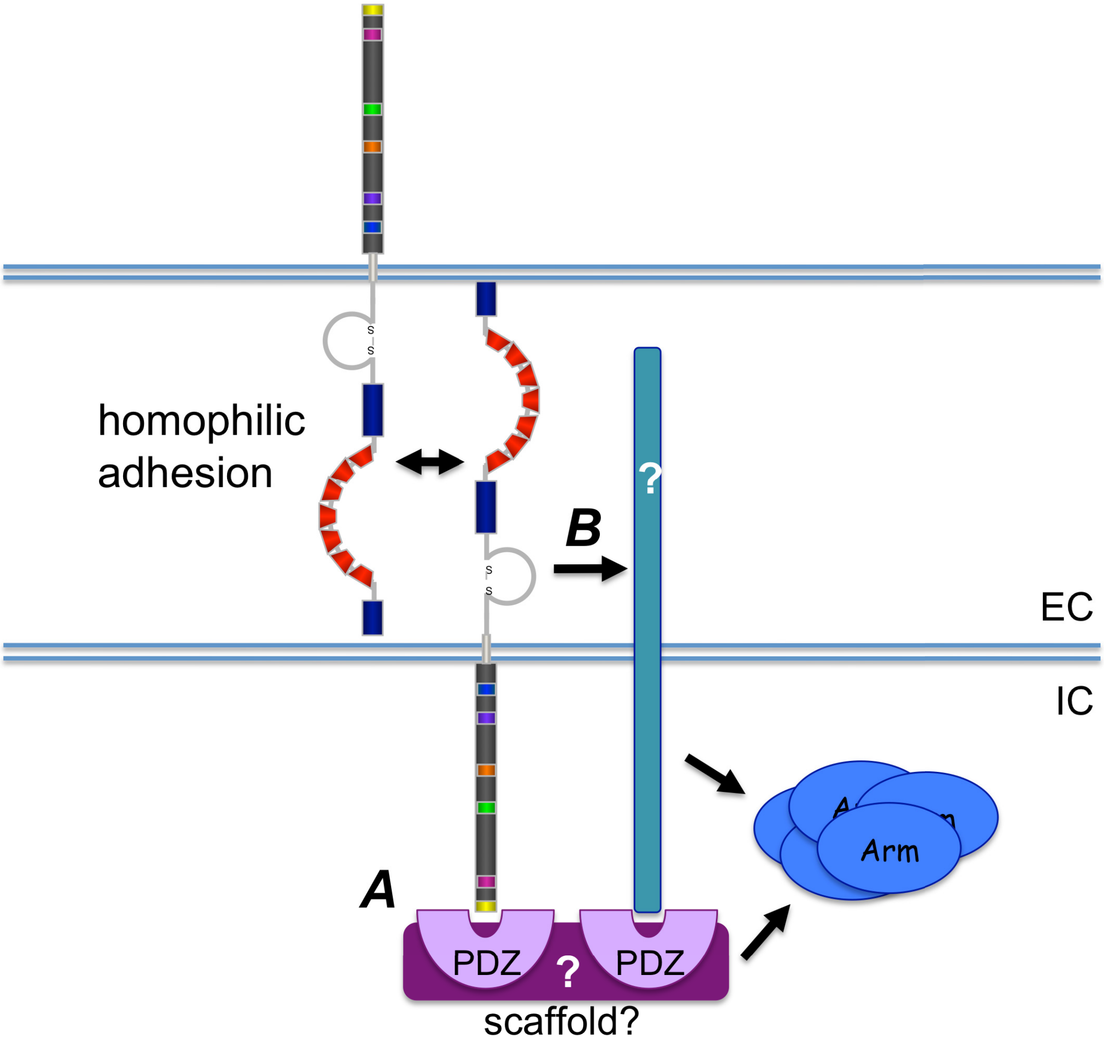


Figure 38: Putative models for Kek5-mediated Arm upregulation. Kek5 first associates with itself, triggering binding with an intracellular PDZ-containing molecule, causing Arm recruitment (A). Alternatively, homophilic adhesion may activate its ability to interact with an unknown transmembrane molecule, which in turn recruits Arm (B). The PDZ-containing component may also serve as a scaffold for an entire membrane-associated complex that eventually causes Arm to accumulate at the membrane.

Given that homophilic association is likely to play a role in at least one of Kek5's roles, it becomes of interest to determine exactly *how* Kek5 would interact with itself, as there is precedence for LRR-LRR, Ig-Ig, and LRR-Ig associations. Based on the crystal structure of Kek1 (D. Alvarado, unpublished), one of three models may be possible (Fig 39). The first possibility is that only the LRRs bind to each other. One example of LRR-mediated self-association is LINGO1, which forms a homo-tetramer (Fig 40)(Mosyak et al., 2006). However, if this were the case, the first model for Arm upregulation would be unlikely, as homophilic association would still occur in the absence of Ig domains. The second model is that the opposing Kek5 molecules form two reciprocal LRR-Ig pairs. In this case, deletion of the Ig domain would prevent homophilic association. There is precedence for such an

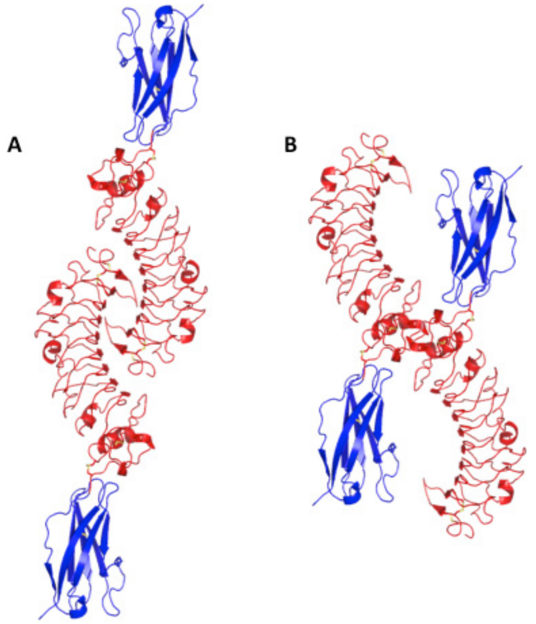


Figure 39: Two models for extracellular Kek5 homophilic association. Kek5 may self-associate via the LRRs alone (red-red; A), or may form two reciprocal LRR-Ig pairs (red-blue; B). (D. Alvarado, unpublished)

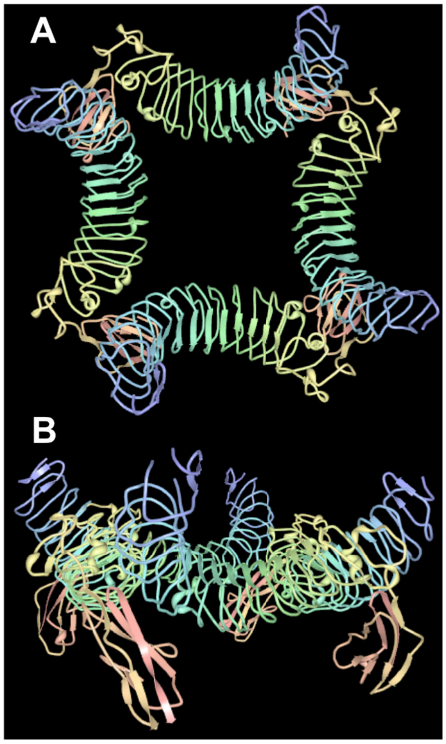


Figure 40: LRRs self-associate in LINGO1 tetramers. Top-down (A) and side (B) views of LINGO1 ectodomain homo-tetramer. In A, the cell surface would be behind the complex, while in B, the cell surface would be below the complex. (PBD ID: 2ID5)

interaction in Slit-Robo signaling, as the crystal structure has been elucidated for the association of LRR domain 2 of Slit and Ig domain 1 of Robo (Fig 41)(Morlot et al., 2007). In the third model, opposing Kek5 molecules associate via an Ig-Ig interaction. As in the second model, deletion of the Ig domain would prevent homophilic association and abrogate Arm upregulation. Assuming the PDZ domain-binding site is not crucial, but rather simply ensures membrane localization, either the second or third mechanism of Kek5 self-association could be possible.

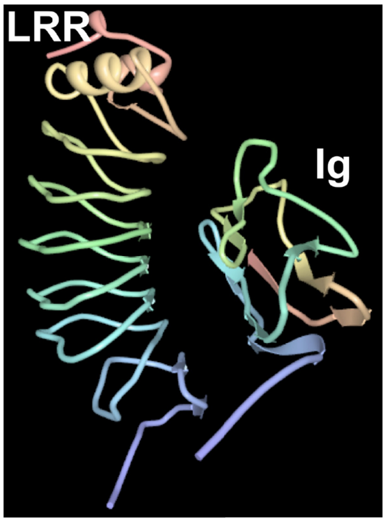


Figure 41: LRRs bind Ig domains. Interface between LRR D2 of Slit (left) and Ig 1 of Robo (right). (PDB ID: 2V9T)

At this point, the link between Kek5 and the elevation of membrane-associated Arm remains unknown. From further junctional analyses, it appears that Kek5 is able to affect one or more components of the nectin-afadin complex (including Polychaetoid). However, based on structure-function correlation, it is more likely that Kek5's ability to modulate these components is related to observed morphological defects. A detailed analysis of the cadherin-catenin will be necessary to determine whether additional components, such as α-catenin and p120-catenin, play a role in mediating Arm upregulation. It is also possible that non-AJ components may be involved in such mechanisms. One such candidate is Bazooka, an intracellular PDZ domain-containing molecule that resides in a tripartite complex in sub-apical region, and has been reported to interact with both Arm and Echinoid (Wei et al., 2005). A number of other binding partners may be possible, given the high density of protein complexes localized to apical epithelial junctions.

A more efficient way to screen for putative Kek5 binding partners would be to recapitulate Kek5-induced Arm upregulation in cell culture. Once this can be done, a large number of candidates can be assessed efficiently through the use of RNAi in Kek5-transfected cells. This was attempted in embryonic S3 and larval wing disc Clone8 cell lines, but neither exhibited a response to transfected Kek5. It is unknown whether either of these lines harbors the appropriate cellular machinery, so the most promising approach may be primary cell culture in which cells from Kek5-misexpressing wing imaginal discs are dissociated and cultured.

A larger scale analysis of the vertebrate LIG family will also likely shed light on the roles of these molecules and any common mechanisms by which they function. Currently, two assays are being developed. One is a smaller scale approach, making use of co-immunoprecipitations to reassess reported interactions between full-length LIGS (AMIGOs, LINGOs, and LRIGs in particular) and the EGFR family. A second high-throughput method will consist of an ELISA-based assay in which secreted versions of the human LIGs will be assessed for binding with the EGFR family and other RTKs. Once these assays are established with appropriate controls, a broader, more informative view of the LIG family may begin to emerge.

References

Alvarado, D., Rice, A. H. and Duffy, J. B. (2004a). Bipartite inhibition of Drosophila epidermal growth factor receptor by the extracellular and transmembrane domains of Kekkon1. *Genetics* **167**, 187-202.

Alvarado, D., Rice, A. H. and Duffy, J. B. (2004b). Knockouts of Kekkon1 define sequence elements essential for Drosophila epidermal growth factor receptor inhibition. *Genetics* **166**, 201-11.

Arrate, M. P., Rodriguez, J. M., Tran, T. M., Brock, T. A. and Cunningham, S. A. (2001). Cloning of human junctional adhesion molecule 3 (JAM3) and its identification as the JAM2 counter-receptor. *J Biol Chem* **276**, 45826-32.

Bai, J., Binari, R., Ni, J. Q., Vijayakanthan, M., Li, H. S. and Perrimon, N. (2008). RNA interference screening in Drosophila primary cells for genes involved in muscle assembly and maintenance. *Development* **135**, 1439-49.

Bork, P., Holm, L. and Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* **242**, 309-20.

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-15.

Brembeck, F. H., Rosario, M. and Birchmeier, W. (2006). Balancing cell adhesion and Wnt signaling, the key role of beta-catenin. *Curr Opin Genet Dev* **16**, 51-9.

Brower, D. L., Brabant, M. C. and Bunch, T. A. (1995). Role of the PS integrins in Drosophila development. *Immunol Cell Biol* **73**, 558-64.

Carmena, A., Speicher, S. and Baylies, M. (2006). The PDZ protein Canoe/AF-6 links Ras-MAPK, Notch and Wingless/Wnt signaling pathways by directly interacting with Ras, Notch and Dishevelled. *PLoS One* **1**, e66.

Cavey, M. and Lecuit, T. (2009). Molecular bases of cell-cell junctions stability and dynamics. *Cold Spring Harbor Perspect Biol* **1**, a002998.

Chen, Y., Aulia, S., Li, L. and Tang, B. L. (2006). AMIGO and friends: an emerging family of brain-enriched, neuronal growth modulating, type I transmembrane proteins with leucine-rich repeats (LRR) and cell adhesion molecule motifs. *Brain Res Rev* **51**, 265-74.

Coates, J. C. (2003). Armadillo repeat proteins: beyond the animal kingdom. *Trends Cell Biol* **13**, 463-71.

Cohen, S. M. (1993). Imaginal Disc Development. In *The Development of Drosophila melanogaster*, vol. II (ed. M. Bate and A. Martinez Arias). Plainview, NY: Cold Spring Harbor Laboratory Press.

Crossin, K. L. and Krushel, L. A. (2000). Cellular signaling by neural cell adhesion molecules of the immunoglobulin superfamily. *Dev Dyn* **218**, 260-79.

Duffy, J. B. (2002). GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis* **34**, 1-15.

Ebnet, K. (2008). Organization of multiprotein complexes at cell-cell junctions. *Histochem Cell Biol* **130**, 1-20.

Ernst, C. L. (2008). Structural and functional dissection of kekkon5. Senior Undergraduate MQP. Worcester, MA: Worcester Polytechnic Institute.

Evans, T. (2006). Characterization of Kekkon5, a Drosophila LIG Protein that Modulates BMP and Integrin Function. PhD Thesis. Indiana University.

Evans, T. A., Haridas, H. and Duffy, J. B. (2009). Kekkon5 is an extracellular regulator of BMP signaling. *Dev Biol* **326**, 36-46.

Franke, W. W. (2009). Discovering the molecular components of intercellular junctions-a historical view. *Cold Spring Harbor Perspect Biol* **1**, a003061.

Fristrom, D. and Fristrom, J. W. (1993). The Metamorphic Development of the Adult Epidermis. In *The Development of Drosophila melanogaster*, vol. II (ed. M. Bate and A. Martinez Arias). Plainview, NY: Cold Spring Harbor Laboratory Press.

Fristrom, D., Wilcox, M. and Fristrom, J. (1993). The distribution of PS integrins, laminin A and F-actin during key stages in Drosophila wing development. *Development* **117**, 509-23.

Furuse, M. and Tsukita, S. (2006). Claudins in occluding junctions of humans and flies. *Trends Cell Biol* **16**, 181-8.

Ghiglione, C., Amundadottir, L., Andresdottir, M., Bilder, D., Diamonti, J. A., Noselli, S., Perrimon, N. and Carraway, I. K. (2003). Mechanism of inhibition of the Drosophila and mammalian EGF receptors by the transmembrane protein Kekkon 1. *Development* **130**, 4483-93.

Ghiglione, C., Carraway, K. L., 3rd, Amundadottir, L. T., Boswell, R. E., Perrimon, N. and Duffy, J. B. (1999). The transmembrane molecule kekkon 1 acts in a feedback loop to negatively regulate the activity of the Drosophila EGF receptor during oogenesis. *Cell* **96**, 847-56.

Gooding, J. M., Yap, K. L. and Ikura, M. (2004). The cadherin-catenin complex as a focal point of cell adhesion and signalling: new insights from three-dimensional structures. *Bioessays* **26**, 497-511.

Hatzfeld, M. (2005). The p120 family of cell adhesion molecules. *Eur J Cell Biol* **84**, 205-14.

Hedman, H. and Henriksson, R. (2007). LRIG inhibitors of growth factor signalling - double-edged swords in human cancer? *Eur J Cancer* **43**, 676-82.

Ho, Y. H., Lien, M. T., Lin, C. M., Wei, S. Y., Chang, L. H. and Hsu, J. C. (2010). Echinoid regulates Flamingo endocytosis to control ommatidial rotation in the Drosophila eye. *Development* **137**, 745-54.

Homma, S., Shimada, T., Hikake, T. and Yaginuma, H. (2009). Expression pattern of LRR and Ig domain-containing protein (LRRIG protein) in the early mouse embryo. *Gene Expr Patterns* **9**, 1-26.

Huang, E. J. and Reichardt, L. F. (2003). Trk receptors: roles in neuronal signal transduction. *Annu Rev Biochem* **72**, 609-42.

Hung, A. Y. and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. *J Biol Chem* **277**, 5699-702.

Inoue, H., Lin, L., Lee, X., Shao, Z., Mendes, S., Snodgrass-Belt, P., Sweigard, H., Engber, T., Pepinsky, B., Yang, L. et al. (2007). Inhibition of the leucine-rich repeat protein LINGO-1 enhances survival, structure, and function of dopaminergic neurons in Parkinson's disease models. *Proc Natl Acad Sci U S A* **104**, 14430-5.

Islam, R., Wei, S. Y., Chiu, W. H., Hortsch, M. and Hsu, J. C. (2003). Neuroglian activates Echinoid to antagonize the Drosophila EGF receptor signaling pathway. *Development* **130**, 2051-9.

Itoh, M., Furuse, M., Morita, K., Kubota, K., Saitou, M. and Tsukita, S. (1999). Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and ZO-3, with the COOH termini of claudins. *J Cell Biol* **147**, 1351-63.

Janssens, B. and Chavrier, P. (2004). Mediterranean views on epithelial polarity. *Nat Cell Biol* **6**, 493-6.

Jung, A. C., Ribeiro, C., Michaut, L., Certa, U. and Affolter, M. (2006). Polychaetoid/ZO-1 is required for cell specification and rearrangement during Drosophila tracheal morphogenesis. *Curr Biol* **16**, 1224-31.

Karaulanov, E. E., Bottcher, R. T. and Niehrs, C. (2006). A role for fibronectin-leucine-rich transmembrane cell-surface proteins in homotypic cell adhesion. *EMBO Rep* **7**, 283-90.

Kedzierski, L., Montgomery, J., Curtis, J. and Handman, E. (2004). Leucine-rich repeats in host-pathogen interactions. *Arch Immunol Ther Exp (Warsz)* **52**, 104-12.

Ko, J. and Kim, E. (2007). Leucine-rich repeat proteins of synapses. *J Neurosci Res* **85**, 2824-32.

Kobe, B. and Kajava, A. V. (2001). The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* **11**, 725-32.

Kostrewa, D., Brockhaus, M., D'Arcy, A., Dale, G. E., Nelboeck, P., Schmid, G., Mueller, F., Bazzoni, G., Dejana, E., Bartfai, T. et al. (2001). X-ray structure of junctional adhesion molecule: structural basis for homophilic adhesion via a novel dimerization motif. *EMBO J* **20**, 4391-8.

Lal-Nag, M. and Morin, P. J. (2009). The claudins. *Genome Biol* **10**, 235.

MacLaren, C. M., Evans, T. A., Alvarado, D. and Duffy, J. B. (2004). Comparative analysis of the Kekkon molecules, related members of the LIG superfamily. *Dev Genes Evol* **214**, 360-6.

Mandell, K. J. and Parkos, C. A. (2005). The JAM family of proteins. *Adv Drug Deliv Rev* **57**, 857-67.

Marygold, S. J. and Vincent, J. P. (2003). Armadillo levels are reduced during mitosis in Drosophila. *Mech Dev* **120**, 157-65.

Matsuo, T., Takahashi, K., Suzuki, E. and Yamamoto, D. (1999). The Canoe protein is necessary in adherens junctions for development of ommatidial architecture in the Drosophila compound eye. *Cell Tissue Res* **298**, 397-404.

Matsushima, N., Tachi, N., Kuroki, Y., Enkhbayar, P., Osaki, M., Kamiya, M. and Kretsinger, R. H. (2005). Structural analysis of leucine-rich-repeat variants in proteins associated with human diseases. *Cell Mol Life Sci* **62**, 2771-91.

McNeill, H. (2000). Sticking together and sorting things out: adhesion as a force in development. *Nat Rev Genet* **1**, 100-8.

Mi, S., Hu, B., Hahm, K., Luo, Y., Kam Hui, E. S., Yuan, Q., Wong, W. M., Wang, L., Su, H., Chu, T. H. et al. (2007). LINGO-1 antagonist promotes spinal cord remyelination and axonal integrity in MOG-induced experimental autoimmune encephalomyelitis. *Nat Med* **13**, 1228-33.

Mi, S., Lee, X., Shao, Z., Thill, G., Ji, B., Relton, J., Levesque, M., Allaire, N., Perrin, S., Sands, B. et al. (2004). LINGO-1 is a component of the Nogo-66 receptor/p75 signaling complex. *Nat Neurosci* **7**, 221-8.

Mi, S., Miller, R. H., Lee, X., Scott, M. L., Shulag-Morskaya, S., Shao, Z., Chang, J., Thill, G., Levesque, M., Zhang, M. et al. (2005). LINGO-1 negatively regulates myelination by oligodendrocytes. *Nat Neurosci* **8**, 745-51.

Miyamoto, H., Nihonmatsu, I., Kondo, S., Ueda, R., Togashi, S., Hirata, K., Ikegami, Y. and Yamamoto, D. (1995). canoe encodes a novel protein containing a GLGF/DHR motif and functions with Notch and scabrous in common developmental pathways in Drosophila. *Genes Dev* **9**, 612-25.

Morlot, C., Thielens, N. M., Ravelli, R. B., Hemrika, W., Romijn, R. A., Gros, P., Cusack, S. and McCarthy, A. A. (2007). Structural insights into the Slit-Robo complex. *Proc Natl Acad Sci U S A* **104**, 14923-8.

Mosyak, L., Wood, A., Dwyer, B., Buddha, M., Johnson, M., Aulabaugh, A., Zhong, X., Presman, E., Benard, S., Kelleher, K. et al. (2006). The structure of the Lingo-1 ectodomain, a module implicated in central nervous system repair inhibition. *J Biol Chem* **281**, 36378-90.

Nakagawara, A. (2001). Trk receptor tyrosine kinases: a bridge between cancer and neural development. *Cancer Lett* **169**, 107-14.

Nelson, W. J. (2008). Regulation of cell-cell adhesion by the cadherin-catenin complex. *Biochem Soc Trans* **36**, 149-55.

Nelson, W. J. and Nusse, R. (2004). Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* **303**, 1483-7.

Padmanabhan, M., Cournoyer, P. and Dinesh-Kumar, S. P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cell Microbiol* **11**, 191-8.

Palsson-McDermott, E. M. and O'Neill, L. A. (2007). Building an immune system from nine domains. *Biochem Soc Trans* **35**, 1437-44.

Pignoni, F. and Zipursky, S. L. (1997). Induction of Drosophila eye development by decapentaplegic. *Development* **124**, 271-8.

Rogulja, D. and Irvine, K. D. (2005). Regulation of cell proliferation by a morphogen gradient. *Cell* **123**, 449-61.

Rogulja, D., Rauskolb, C. and Irvine, K. D. (2008). Morphogen control of wing growth through the Fat signaling pathway. *Dev Cell* **15**, 309-21.

Rual, J. F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P. O. et al. (2004). Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res* **14**, 2128-35.

Rubin, C., Gur, G. and Yarden, Y. (2005). Negative regulation of receptor tyrosine kinases: unexpected links to c-Cbl and receptor ubiquitylation. *Cell Res* **15**, 66-71.

Satir, P. and Gilula, N. B. (1973). The fine structure of membranes and intercellular communication in insects. *Annu Rev Entomol* **18**, 143-66.

Sawyer, J. K., Harris, N. J., Slep, K. C., Gaul, U. and Peifer, M. (2009). The Drosophila afadin homologue Canoe regulates linkage of the actin cytoskeleton to adherens junctions during apical constriction. *J Cell Biol* **186**, 57-73.

Seppa, M. J., Johnson, R. I., Bao, S. and Cagan, R. L. (2008). Polychaetoid controls patterning by modulating adhesion in the Drosophila pupal retina. *Dev Biol* **318**, 1-16.

Speicher, S., Fischer, A., Knoblich, J. and Carmena, A. (2008). The PDZ protein Canoe regulates the asymmetric division of Drosophila neuroblasts and muscle progenitors. *Curr Biol* **18**, 831-7.

Takahashi, K., Hamada, N. and Yamamoto, D. (2002). Ras target protein canoe is a substrate for Cdc2 and Cdk5 kinases. *Arch Insect Biochem Physiol* **49**, 102-7.

Wei, S. Y., Escudero, L. M., Yu, F., Chang, L. H., Chen, L. Y., Ho, Y. H., Lin, C. M., Chou, C. S., Chia, W., Modolell, J. et al. (2005). Echinoid is a component of adherens junctions that cooperates with DE-Cadherin to mediate cell adhesion. *Dev Cell* **8**, 493-504.

Wei, T., Gong, J., Jamitzky, F., Heckl, W. M., Stark, R. W. and Rossle, S. C. (2008). LRRML: a conformational database and an XML description of leucine-rich repeats (LRRs). *BMC Struct Biol* **8**, 47.

Wei, X. and Ellis, H. M. (2001). Localization of the Drosophila MAGUK protein Polychaetoid is controlled by alternative splicing. *Mech Dev* **100**, 217-31.

Wirtz-Peitz, F. and Zallen, J. A. (2009). Junctional trafficking and epithelial morphogenesis. *Curr Opin Genet Dev* **19**, 350-6.

Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D. and Zipursky, S. L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* **130**, 1134-45.

Xu, W. and Kimelman, D. (2007). Mechanistic insights from structural studies of beta-catenin and its binding partners. *J Cell Sci* **120**, 3337-44.

Appendix A:

Human LIG cDNA Availability

Gene Name	IMAGE ID	Accession	Library	Vector	OpenBiosystems #	Alternatives
AMIGO1	5725643	BC040879	NIH MGC 125	pCMV-Sport6	**5725643**	
AMIGO2	4828334	BC047595	NIH MGC 97	pCMV-Sport6.1	**30387316**	4828334; pBluescriptR
AMIGO3	5211193	BC110418	NIH MGC 118	pCMV-Sport6	**5211193**	40025738; pCR-BluntII-TOPO
GPR124	9007223	BC146774	Kazusa-KIAA-pBluescriptIISK+	N/A	**EST cDNAs only**	
GPR125	N/A	BC026009	N/A	pDNR-LIB	**4714761**	
IGSF10	N/A	N/A	N/A	N/A	**EST cDNAs only**	
ISLR	5763157	BC111013	NIH_MGC_114	pCMV-SPORT6	**5763157**	4794738; pBluescriptR
ISLR2	8862545	BC152429	Kazusa-KIAA-pBluescriptIISK+	N/A	**EST cDNAs only**	
LINGO1	4797856	BC068558	NIH_MGC_96	pCMV-Sport6	**4214343**	4797856; pBluescriptR
LINGO2	9021514	BC137514	NIH_MGC_310	pCR4-TOPO	**9021514**	
LINGO3	N/A	N/A	N/A	N/A	N/A	
LINGO4	9021217	BC137220	NIH_MGC_362	pCR4-TOPO	**9021217**	902121;, same vector
LRFN1	N/A	N/A	N/A	N/A	N/A	
LRFN2	40147341	BC142616	NIH_MGC_417	pCMV-SPORT6	**40147341**	
LRFN3	3537259	BC003578	NIH_MGC_7	pOTB7	**3537259**	
LRFN4	4637053	BC015581	NIH_MGC_18	pOTB7	**4637053**	
LRFN5	5286818	BC043165	NIH_MGC_96	pBluescriptR	**5286818**	
LRIG1	30340674	BC071561	NIH_MGC_147	pBluescriptR	**30340674**	
LRIG2	40125921	BC117370	NIH_MGC_311	pCR4-TOPO	**40125921**	
LRIG3	8991885	BC126169	NIH_MGC_310	pCR4-TOPO	**8991885**	
LRIT1	N/A	N/A	N/A	N/A	N/A	
LRIT2	9053002	BC144475	NIH_MGC_362	N/A	**EST cDNAs only**	
LRIT3	40026462	BC104037	NIH_MGC_283	pCR-BluntII-TOPO	**40026462**	
LRRC24	5767723	BC111067	NIH_MGC_121	pCMV-SPORT6	**5767723**	
LRRC4/NAG14	40068142	BC111561	NIH_MGC_282	pCR-BluntII-TOPO	**40068142**	
LRRC4B/LRIG4	N/A	N/A	N/A	N/A	N/A	
LRRC4C/NGL1	5274422	BC041374	NIH_MGC_95	pBluescriptR	**5274422**	
LRRN1	4826337	BC034947	NIH_MGC_97	pBluescriptR	**4826337**	
LRRN2	5301942	BC068541	NIH_MGC_96	pBluescriptR	**5301942**	4798822; same vector
LRRN3	5263900	BC035133	NIH_MGC_95	pBluescriptR	**5263900**	
MXRA5	N/A	N/A	N/A	N/A	N/A	
NTRK1	9052759	BC144239	NIH_MGC_361	N/A	**EST cDNAs only**	
NTRK2	4941763	BC031835	NCI_CGAP_Brn67	pCMV-SPORT6	**4941763**	
NTRK3	4181574	BC013693	NCI_CGAP_Brn67	pCMV-SPORT6	**4181574**	
PXDN	N/A	N/A	N/A	N/A	**EST cDNAs only**	
PXDNL	N/A	BC132813	N/A	pCR4-TOPO	**40146835**	

Appendix B:

Modified Gateway Cloning

GATEWAY CLONING WITH MODIFIED ATT SITES (B1.1, 2.1)

(Rual et al. Genome Res. 2004 14: 2128-2135)

attB sites in PCR product

attB1: 5′ ACAAGTTT**GTACAAAAAAGCAGGCT** 3′ **attB1.1:** 5′ ACAACTTT**GTACAAAAAAGTTG** 3′
TGTTCAAACATGTTT**TTTGCACCGA** TGTTGAAACATGTTT**TTTCAAC**

attB2: 5′ **ACCCAGCTTT**CTTGTACTTTGTGGT 3′ **attB2.1:** 5′ **CAACTTT**CTTGTACAAAGTTGT 3′
TGGGTCGAAAGAACATGTTTCACCA **GTTGAAAGAACATG**TTTCAACA

Notes:
Bases are kept throughout recombination steps.
Bases have changed between versions.
Bases participate in recombination; these bases are unchanged, making the new sites compatible with original P sites.

```
        [CAGGCT]             [GACCCAG]                        --|
GTACAAAAAAGTTG--------PCR--------GCAACTTT           --| B1.1, 2.1 | B1, 2
      TTTCAAC--------PCR--------CGTTGAAAGAACATG     --|           |
        [GTCCGA]             [CTGGGTC]                        --|
```

(C or G – *only* added to put in frame with pUAST fusion expression vectors)

attL sites in resulting pENTR vector

```
5′ N75-CCAACTTT-----------insert-goes-here-----CTTGTACAAAGTTGG-N75 3′
   N75-GGTTGAAACATGTTT----insert-goes-here------------TTTCAACC-N75

                              |||
                              VVV

5′ N75-CCAACTTTGTACAAAAAAGTTG----PCR----CAACTTTCTTGTACAAAGTTGG-N75 3′
   N75-GGTTGAAACATGTTTTTTCAAC----PCR----GTTGAAAGAACATGTTTCAACC-N75
```

Note: the regions outside of the recombination sites (non-underlined bases in figure above) will *not* be affected by new attB sites.

attB sites in resulting expression construct (pUAST, etc)
(this will be a *hybrid* of the attB1/2 and attB1.1/2.1 sites)

```
5′ ACAAGTTTGTACAAAAAAGTTG----PCR----CAACTTTCTTGTACTTTGTGGT 3′
   TGTTCAAACATGTTTTTTCAAC----PCR----GTTGAAAGAACATGTTTCACCA

   |--B1--|     |-B1.1-|          |-B2.1-|     |--B2--|
```

Appendix C:

Molecular Cloning: Primers

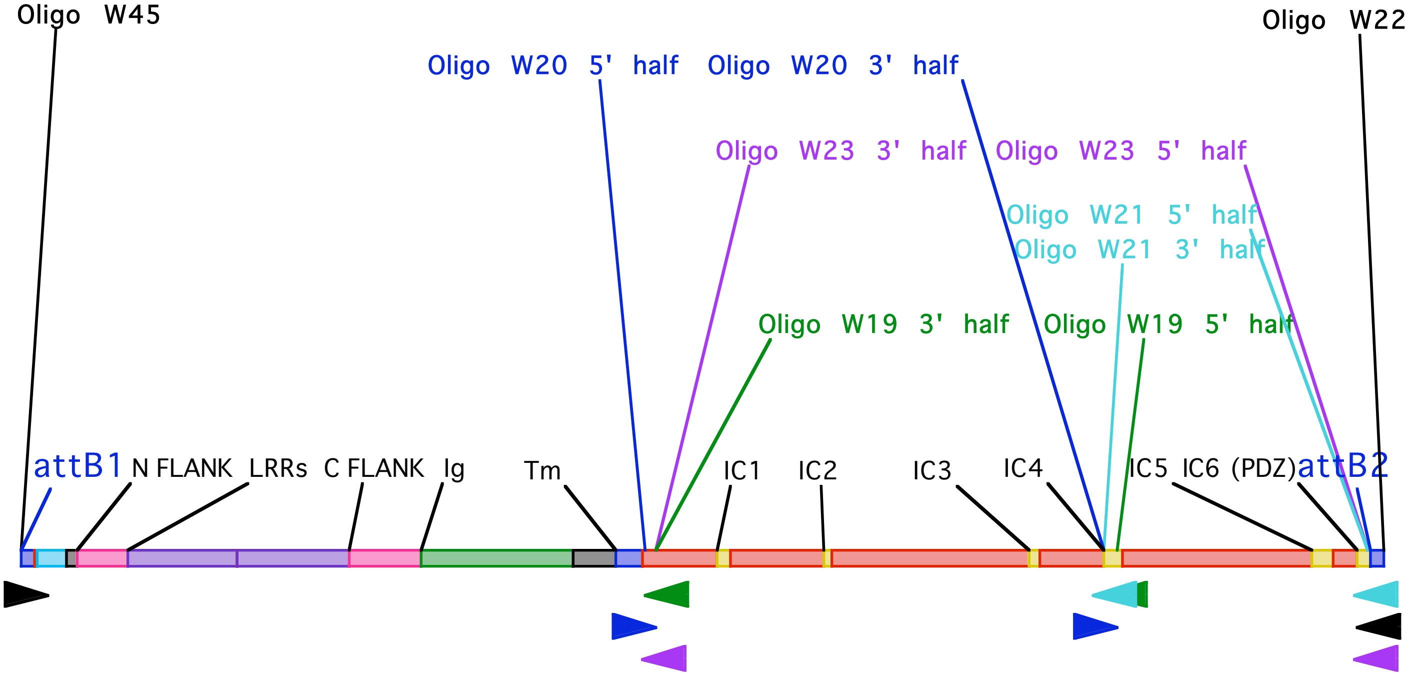
W104 (AMIGO1 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGCACCCCCACCGTGACCCG
W105 (AMIGO1 3′)
GGGGACAACTTTGTACAAGAAAGTTGGCACCACAATGGGCGTATCAGAGAAG
W106 (AMIGO2 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGTCGTTACGTGTACACACTCTGC
W107 (AMIGO2 3′)
GGGGACAACTTTGTACAAGAAAGTTGGAGTGGACGCGACAAAAGGTGTGTC
W108 (AMIGO3 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGACGTGGTTGGTGCTGCTGGG
W109 (AMIGO3 3′)
GGGGACAACTTTGTACAAGAAAGTTGGGGTTGTCATGGGACCCTCGGAG
W110 (LINGO1 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGCAGGTGAGTAAGAGGATGCTG
W111 (LINGO1 3′)
GGGGACAACTTTGTACAAGAAAGTTGGTATCATCTTCATGTTGAACTTGCG

W141 (sDER 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGCTGCTACGACGGCGCAAC
W142 (sDER 3′)
GGGGACAACTTTGTACAAGAAAGTTGGCTTGCTGCTCCTCGGCGGAC
W143 (sKEK1 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGCATATCAGGGAAGCAGTTTTC
W144 (sKEK1 3′)
GGGGACAACTTTGTACAAGAAAGTTGGTTGGCCACTACCCAGGGAGGC
W145 (sKEK2 5′)
GGGGACAACTTTGTACAAAAAAGTTGGAAAATGAGTGGTCTGCCAATCTGG
W146 (sKEK2 3′)
GGGGACAACTTTGTACAAGAAAGTTGGGTTCATGTAATCCCTGGGGAAG

ATTB 1.1/2.1 SITES
CODING REGION

Kek5

IC Deletion Primer Locations


Oligo W45
Oligo W22
Oligo W20 5' half
Oligo W20 3' half
Oligo W23 3' half
Oligo W23 5' half
Oligo W21 5' half
Oligo W21 3' half
Oligo W19 3' half
Oligo W19 5' half
attB1
N FLANK
LRRs
C FLANK
Ig
Tm
IC1
IC2
IC3
IC4
IC5
IC6 (PDZ)
attB2

Appendix D:

Molecular Cloning: Construct Sequences

pUAST-Kek5Δ123-GFP

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGAC

  77 GGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCT
     CCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGA

 153 CGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAA
     GCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTT

 229 GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCG
     CCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGC

 305 ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTA
     TGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAAT

 381 CCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTC
     GGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAG

 457 GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
     CCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCA

 533 AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
     TTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGT

 609 GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
     CTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAA

 685 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
     ACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG

 761 GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
     CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACT

 837 TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
     AGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTC

 913 GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
     CTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGA

 989 GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
     CTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTG

1065 TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
     AGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGG

1141 ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
     TGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGA

1217 TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA
     AATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGT

1293 ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
     TGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGT

1369 ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
     TGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAG

1445 AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
     TCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGC

1521 TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
     ATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAG

1597 TTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
     AACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGA

1673 TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
     AGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTA

1749 CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
     GAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTA
```

```
1825 AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
     TTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACA
-----------------------------------------------------------------------------------
1901 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
     GAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTC
-----------------------------------------------------------------------------------
1977 TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
     ACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGC
-----------------------------------------------------------------------------------
2053 TCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA
     AGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATT
-----------------------------------------------------------------------------------
2129 GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG
     CGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATAC
-----------------------------------------------------------------------------------
2205 CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACG
     GCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGC
-----------------------------------------------------------------------------------
2281 ACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTA
     TGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGAT
-----------------------------------------------------------------------------------
2357 CGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCC
     GCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGG
-----------------------------------------------------------------------------------
2433 GGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAAT
     CCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTA
-----------------------------------------------------------------------------------
2509 CCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACA
     GGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGT
-----------------------------------------------------------------------------------
2585 AATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAAACAAG
     TTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTTGTTC
-----------------------------------------------------------------------------------
2661 CAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGA
     GTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCT
-----------------------------------------------------------------------------------
                             3'P
2737 AGCTCACGATGAGAATGGCCAGACCATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACG
     TCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGC
                             ____________________________________________________
-----------------------------------------------------------------------------------
2813 TATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGA
     ATACGAACGTTATTCACGCTCACTTTCCTTATCATAAGACTCACAGCATAACTCAGACTCACTCTGTCGCTATACT
     ____________________________________________________________________________
-----------------------------------------------------------------------------------
2889 TTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAG
     AACAACTAATTGGGAATCGTACAGGCACCCCAAACTTAATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTC
     ____________________________________________________________________________
-----------------------------------------------------------------------------------
2965 TTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGG
     AAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCC
     ____________________________>      ________________________>   _______________
-----------------------------------------------------------------------------------
3041 AGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGG
     TCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCC
     ____________________________________________________________________________
-----------------------------------------------------------------------------------
                               UAS sites
3117 AGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATT
     TCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAA
     __________________________>   ______________________________________________
-----------------------------------------------------------------------------------
3193 CAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAA
     GTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTT
     ____________________________________________________________________________
-----------------------------------------------------------------------------------
3269 CAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACT
     GTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGA
     ____________________________________________________________________________
-----------------------------------------------------------------------------------
```

```
3345 GAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACA GAT
     CTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGT CTA
-------------------------------------------------------------------------------
         attB1
3420 CTG ACA AGT TTG TAC AAA AAA GCA GGC TCCAGGAAA ATG ATC CTT CTG CTG CTG GGT
     GAC TGT TCA AAC ATG TTT TTT CGT CCG AGGTCCTTT TAC TAG GAA GAC GAC GAC CCA
                                                1▸ M   I   L   L   L   L   G
-------------------------------------------------------------------------------
3477 GTG CTA GTG GTT CTA ATG GCC CTA CCG CCG CCC ACC GCA GGC ACC ACC GAT TGG ATG
     CAC GAT CAC CAA GAT TAC CGG GAT GGC GGC GGG TGG CGT CCG TGG TGG CTA ACC TAC
   8▸ V   L   V   V   L   M   A   L   P   P   P   T   A   G   T   T   D   W   M
-------------------------------------------------------------------------------
3534 CAG AGC TGC GGT ACA TGC CAC TGT CAG TGG AAT TCG GGC AAG AAG AGC GCC GAC TGC
     GTC TCG ACG CCA TGT ACG GTG ACA GTC ACC TTA AGC CCG TTC TTC TCG CGG CTG ACG
  27▸ Q   S   C   G   T   C   H   C   Q   W   N   S   G   K   K   S   A   D   C
-------------------------------------------------------------------------------
3591 AAG AAC AAG GCG CTA ACC AAA ATT CCG CAG GAC ATG AGC AAC GAG ATG CAG GTG CTG
     TTC TTG TTC CGC GAT TGG TTT TAA GGC GTC CTG TAC TCG TTG CTC TAC GTC CAC GAC
  46▸ K   N   K   A   L   T   K   I   P   Q   D   M   S   N   E   M   Q   V   L
-------------------------------------------------------------------------------
3648 GAC TTT GCC CAC AAT CAA ATA CCC GAG CTG CGG CGC GAA GAG TTC CTA CTG GCC GGT
     CTG AAA CGG GTG TTA GTT TAT GGG CTC GAC GCC GCG CTT CTC AAG GAT GAC CGG CCA
  65▸ D   F   A   H   N   Q   I   P   E   L   R   R   E   E   F   L   L   A   G
-------------------------------------------------------------------------------
3705 CTG CCC AAT GTG CAC AAG ATC TTT TTG CGC AAC TGC ACC ATC CAG GAG GTG CAT CGC
     GAC GGG TTA CAC GTG TTC TAG AAA AAC GCG TTG ACG TGG TAG GTC CTC CAC GTA GCG
  84▸ L   P   N   V   H   K   I   F   L   R   N   C   T   I   Q   E   V   H   R
-------------------------------------------------------------------------------
3762 GAG GCC TTC AAG GGT CTG CAT ATC CTA ATC GAG CTG GAC CTG TCG GGC AAT CGG ATA
     CTC CGG AAG TTC CCA GAC GTA TAG GAT TAG CTC GAC CTG GAC AGC CCG TTA GCC TAT
 103▸ E   A   F   K   G   L   H   I   L   I   E   L   D   L   S   G   N   R   I
-------------------------------------------------------------------------------
3819 CGG GAA CTG CAT CCG GGC ACT TTC GCC GGC CTG GAG AAG CTG CGC AA T GTG ATC ATC
     GCC CTT GAC GTA GGC CCG TGA AAG CGG CCG GAC CTC TTC GAC GCG TT A CAC TAG TAG
 122▸ R   E   L   H   P   G   T   F   A   G   L   E   K   L   R   N    V   I   I
-------------------------------------------------------------------------------
3876 AAC AAC AAC GAG ATC GAG GTG CTG CCC AAC CAT CTG TTC GTC AAC CTG AGC TTC CTG
     TTG TTG TTG CTC TAG CTC CAC GAC GGG TTG GTA GAC AAG CAG TTG GAC TCG AAG GAC
 141▸ N   N   N   E   I   E   V   L   P   N   H   L   F   V   N   L   S   F   L
-------------------------------------------------------------------------------
3933 TCG CGC ATC GAG TTC CGG AAC AAT CGA TTG CGC CAG GTG CAG CTG CAC GTC TTC GCT
     AGC GCG TAG CTC AAG GCC TTG TTA GCT AAC GCG GTC CAC GTC GAC GTG CAG AAG CGA
 160▸ S   R   I   E   F   R   N   N   R   L   R   Q   V   Q   L   H   V   F   A
-------------------------------------------------------------------------------
3990 GGC ACA ATG GCG CTG AGC GCC ATT TCG CTG GAA CAG AAC CGC CTC TCA CAT CTG CAC
     CCG TGT TAC CGC GAC TCG CGG TAA AGC GAC CTT GTC TTG GCG GAG AGT GTA GAC GTG
 179▸ G   T   M   A   L   S   A   I   S   L   E   Q   N   R   L   S   H   L   H
-------------------------------------------------------------------------------
4047 AAG GAG ACA TTC AAG GAT CTG CAG AAG CTG ATG CAT CTA TCG CTG CAG GGT AAC GCA
     TTC CTC TGT AAG TTC CTA GAC GTC TTC GAC TAC GTA GAT AGC GAC GTC CCA TTG CGT
 198▸ K   E   T   F   K   D   L   Q   K   L   M   H   L   S   L   Q   G   N   A
-------------------------------------------------------------------------------
4104 TGG AAC TGC AGC TGC GAG CTG CAG GAC TTT CGC GAC TTT GCG ATC AGC AAA CGG CTC
     ACC TTG ACG TCG ACG CTC GAC GTC CTG AAA GCG CTG AAA CGC TAG TCG TTT GCC GAG
 217▸ W   N   C   S   C   E   L   Q   D   F   R   D   F   A   I   S   K   R   L
-------------------------------------------------------------------------------
4161 TAC ACA CCG CCC ACC GAT TGC CAG GAG CCG CCA CAG CTG CGC GGC AAG CTG TGG AGC
     ATG TGT GGC GGG TGG CTA ACG GTC CTC GGC GGT GTC GAC GCG CCG TTC GAC ACC TCG
 236▸ Y   T   P   P   T   D   C   Q   E   P   P   Q   L   R   G   K   L   W   S
-------------------------------------------------------------------------------
4218 GAG GTG CCA TCG GAG AAC TTC GCC TGC CGG CCG CGC ATT TTG GGT TCC GTG CGC TCC
     CTC CAC GGT AGC CTC TTG AAG CGG ACG GCC GGC GCG TAA AAC CCA AGG CAC GCG AGG
 255▸ E   V   P   S   E   N   F   A   C   R   P   R   I   L   G   S   V   R   S
-------------------------------------------------------------------------------
4275 TTC ATC GAG GCC AAT CAC GAC AAT ATC TCG CTA CCC TGC CGC ATT GTC GGC AGT CCG
     AAG TAG CTC CGG TTA GTG CTG TTA TAG AGC GAT GGG ACG GCG TAA CAG CCG TCA GGC
 274▸ F   I   E   A   N   H   D   N   I   S   L   P   C   R   I   V   G   S   P
-------------------------------------------------------------------------------
```

```
4332 CGT CCC AAT GTC ACC TGG GTG TAC AAC AAG CGG CCA TTG CAG CAG TAC GAC CCG CGT
     GCA GGG TTA CAG TGG ACC CAC ATG TTG TTC GCC GGT AAC GTC GTC ATG CTG GGC GCA
  293▸ R   P   N   V   T   W   V   Y   N   K   R   P   L   Q   Q   Y   D   P   R
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4389 GTG CGT GTC CTC ACC TCC GTG GAA CAG ATG CCG GAG CAG CCC TCC CAG GTG CTC ACC
     CAC GCA CAG GAG TGG AGG CAC CTT GTC TAC GGC CTC GTC GGG AGG GTC CAC GAG TGG
  312▸ V   R   V   L   T   S   V   E   Q   M   P   E   Q   P   S   Q   V   L   T
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4446 TCG GAG CTG CGC ATC GTG GGC GTA CGG GCC TCC GAC AAG GGT GCC TAC ACC TGT GTG
     AGC CTC GAC GCG TAG CAC CCG CAT GCC CGG AGG CTG TTC CCA CGG ATG TGG ACA CAC
  331▸ S   E   L   R   I   V   G   V   R   A   S   D   K   G   A   Y   T   C   V
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4503 GCG GAT AAC CGG GGC GGA CGG GCG GAG GCC GAG TTC CAG CTG CTC GTG AGC GGT GAC
     CGC CTA TTG GCC CCG CCT GCC CGC CTC CGG CTC AAG GTC GAC GAG CAC TCG CCA CTG
  350▸ A   D   N   R   G   G   R   A   E   A   E   F   Q   L   L   V   S   G   D
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4560 TAT GCC GGC GCG GTA TCC GCC TCC GAT GGC ATG GGC ATG GGC GCC ATT GGG GCA CCA
     ATA CGG CCG CGC CAT AGG CGG AGG CTA CCG TAC CCG TAC CCG CGG TAA CCC CGT GGT
  369▸ Y   A   G   A   V   S   A   S   D   G   M   G   M   G   A   I   G   A   P
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

                                              TM
4617 ACC ATT GAT CCG CAA ACG AAC ATG TTT CTC ATC ATC TGT CTA ATC ATT ACG ACG CTG
     TGG TAA CTA GGC GTT TGC TTG TAC AAA GAG TAG TAG ACA GAT TAG TAA TGC TGC GAC
  388▸ T   I   D   P   Q   T   N   M   F   L   I   I   C   L   I   I   T   T   L
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4674 CTG CTC CTG CTG CTC GTG GCG GTG CTG ACG CTC TTC TGG TAC TGC CGa CGC ATC AAG
     GAC GAG GAC GAC GAG CAC CGC CAC GAC TGC GAG AAG ACC ATG ACG GCt GCG TAG TTC
  407▸ L   L   L   L   L   V   A   V   L   T   L   F   W   Y   C   R   R   I   K
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

         TM-IC4 merge
4731 ACC GAG CCA GTG TAC GAC AAC TTG GGA TTG CGA ACC ACT GCC GGC GGC AAC TCC ACC
     TGG CTC GGT CAC ATG CTG TTG AAC CCT AAC GCT TGG TGA CGG CCG CCG TTG AGG TGG
  426▸ T   E   P   V   Y   D   N   L   G   L   R   T   T   A   G   G   N   S   T
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4788 CTC AAT CTG ACC AAG ATC GCC GGC TCA CAG GGG GGC GCT GGT CAG CAG TAC TCG ATG
     GAG TTA GAC TGG TTC TAG CGG CCG AGT GTC CCC CCG CGA CCA GTC GTC ATG AGC TAC
  445▸ L   N   L   T   K   I   A   G   S   Q   G   G   A   G   Q   Q   Y   S   M
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4845 CGG GAC CGA CCA CTT CCG GCC ACG CCC AGC CTG ACA TCG GTG TCC TCG GCG ACC AAT
     GCC CTG GCT GGT GAA GGC CGG TGC GGG TCG GAC TGT AGC CAC AGG AGC CGC TGG TTA
  464▸ R   D   R   P   L   P   A   T   P   S   L   T   S   V   S   S   A   T   N
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4902 GCC AGT AAG ATT TAC GAG CCC ATA CAC GAG CTG ATT CAG CAG CAA CAG CAG TTG CAA
     CGG TCA TTC TAA ATG CTC GGG TAT GTG CTC GAC TAA GTC GTC GTT GTC GTC AAC GTT
  483▸ A   S   K   I   Y   E   P   I   H   E   L   I   Q   Q   Q   Q   Q   L   Q
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4959 CAA CAA CAA CAG CAG CAG CAG CAG CGA CTG GGC TCC ATG GAC ACG GAA CCC CTG TAC
     GTT GTT GTT GTC GTC GTC GTC GTC GCT GAC CCG AGG TAC CTG TGC CTT GGG GAC ATG
  502▸ Q   Q   Q   Q   Q   Q   Q   Q   R   L   G   S   M   D   T   E   P   L   Y
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5016 GGA GTT CGG CAA CAG GGG ATC ACG ATA CTG CCC GGC TCG AGC ATT AGC GGT GCC GGA
     CCT CAA GCC GTT GTC CCC TAG TGC TAT GAC GGG CCG AGC TCG TAA TCG CCA CGG CCT
  521▸ G   V   R   Q   Q   G   I   T   I   L   P   G   S   S   I   S   G   A   G
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5073 CTG GGC CAC GCC GCC TAC CTT TCA CCC GGC TCG GGT GCC GCC GTA TCG CCA AGC CAC
     GAC CCG GTG CGG CGG ATG GAA AGT GGG CCG AGC CCA CGG CGG CAT AGC GGT TCG GTG
  540▸ L   G   H   A   A   Y   L   S   P   G   S   G   A   A   V   S   P   S   H
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5130 GCC AGC AGC AGC GGT GAC TCT CCG AAG GCC GCC AAG ATC CCA CCA CGC CCA CCA CCG
     CGG TCG TCG TCG CCA CTG AGA GGC TTC CGG CGG TTC TAG GGT GGT GCG GGT GGT GGC
  559▸ A   S   S   S   G   D   S   P   K   A   A   K   I   P   P   R   P   P   P
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5187 AAG CCC AAG AAG AAG ATG TCC GTG ACG ACG ACG CGC AGC GGC CAG GGC AGC ACC AGC
     TTC GGG TTC TTC TTC TAC AGG CAC TGC TGC TGC GCG TCG CCG GTC CCG TCG TGG TCG
  578▸ K   P   K   K   K   M   S   V   T   T   T   R   S   G   Q   G   S   T   S
```

```
                                                                         attB2
5244 CAG CTC TTC GAC GAC GAG GGC GAG GAT GGC ACC GAG GTC GAC CCA GCT TTC TTG
     GTC GAG AAG CTG CTG CTC CCG CTC CTA CCG TGG CTC CAG CTG GGT CGA AAG AAC
 597▸ Q   L   F   D   D   E   G   E   D   G   T   E   V   D   P   A   F   L
-------------------------------------------------------------------------------
                                                                     EGFPN1
5298 TAC AAA GTG GTG GTA CCG CGG GCC CGG GAT CCA CCG GTC GCC ACC ATG GTG AGC
     ATG TTT CAC CAC CAT GGC GCC CGG GCC CTA GGT GGC CAG CGG TGG TAC CAC TCG
 615▸ Y   K   V   V   V   P   R   A   R   D   P   P   V   A   T   M   V   S
-------------------------------------------------------------------------------
5352 AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC
     TTC CCG CTC CTC GAC AAG TGG CCC CAC CAC GGG TAG GAC CAG CTC GAC CTG CCG CTG
 633▸ K   G   E   E   L   F   T   G   V   V   P   I   L   V   E   L   D   G   D
-------------------------------------------------------------------------------
5409 GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC
     CAT TTG CCG GTG TTC AAG TCG CAC AGG CCG CTC CCG CTC CCG CTA CGG TGG ATG CCG
 652▸ V   N   G   H   K   F   S   V   S   G   E   G   E   G   D   A   T   Y   G
-------------------------------------------------------------------------------
5466 AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC
     TTC GAC TGG GAC TTC AAG TAG ACG TGG TGG CCG TTC GAC GGG CAC GGG ACC GGG TGG
 671▸ K   L   T   L   K   F   I   C   T   T   G   K   L   P   V   P   W   P   T
-------------------------------------------------------------------------------
5523 CTC GTG ACC ACC CTG ACC TAC GGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG
     GAG CAC TGG TGG GAC TGG ATG CCG CAC GTC ACG AAG TCG GCG ATG GGG CTG GTG TAC
 690▸ L   V   T   T   L   T   Y   G   V   Q   C   F   S   R   Y   P   D   H   M
-------------------------------------------------------------------------------
5580 AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC
     TTC GTC GTG CTG AAG AAG TTC AGG CGG TAC GGG CTT CCG ATG CAG GTC CTC GCG TGG
 709▸ K   Q   H   D   F   F   K   S   A   M   P   E   G   Y   V   Q   E   R   T
-------------------------------------------------------------------------------
5637 ATC TTC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC
     TAG AAG AAG TTC CTG CTG CCG TTG ATG TTC TGG GCG CGG CTC CAC TTC AAG CTC CCG
 728▸ I   F   F   K   D   D   G   N   Y   K   T   R   A   E   V   K   F   E   G
-------------------------------------------------------------------------------
5694 GAC ACC CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC
     CTG TGG GAC CAC TTG GCG TAG CTC GAC TTC CCG TAG CTG AAG TTC CTC CTG CCG TTG
 747▸ D   T   L   V   N   R   I   E   L   K   G   I   D   F   K   E   D   G   N
-------------------------------------------------------------------------------
5751 ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC ATG GCC
     TAG GAC CCC GTG TTC GAC CTC ATG TTG ATG TTG TCG GTG TTG CAG ATA TAG TAC CGG
 766▸ I   L   G   H   K   L   E   Y   N   Y   N   S   H   N   V   Y   I   M   A
-------------------------------------------------------------------------------
5808 GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC
     CTG TTC GTC TTC TTG CCG TAG TTC CAC TTG AAG TTC TAG GCG GTG TTG TAG CTC CTG
 785▸ D   K   Q   K   N   G   I   K   V   N   F   K   I   R   H   N   I   E   D
-------------------------------------------------------------------------------
5865 GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC
     CCG TCG CAC GTC GAG CGG CTG GTG ATG GTC GTC TTG TGG GGG TAG CCG CTG CCG GGG
 804▸ G   S   V   Q   L   A   D   H   Y   Q   Q   N   T   P   I   G   D   G   P
-------------------------------------------------------------------------------
5922 GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC CAG TCC GCC CTG AGC AAA GAC CCC
     CAC GAC GAC GGG CTG TTG GTG ATG GAC TCG TGG GTC AGG CGG GAC TCG TTT CTG GGG
 823▸ V   L   L   P   D   N   H   Y   L   S   T   Q   S   A   L   S   K   D   P
-------------------------------------------------------------------------------
5979 AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT
     TTG CTC TTC GCG CTA GTG TAC CAG GAC GAC CTC AAG CAC TGG CGG CGG CCC TAG TGA
 842▸ N   E   K   R   D   H   M   V   L   L   E   F   V   T   A   A   G   I   T
-------------------------------------------------------------------------------
                                                                 SV40 Poly A
6036 CTC GGC ATG GAC GAG CTG TAC AAG TAA AGCGGCCGCGACTCTAGAGGATCTTTGTGAAGGAACCTTA
     GAG CCG TAC CTG CTC GAC ATG TTC ATT TCGCCGGCGCTGAGATCTCCTAGAAACACTTCCTTGGAAT
 861▸ L   G   M   D   E   L   Y   K   •
-------------------------------------------------------------------------------
6103 CTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGT
     GAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCA
-------------------------------------------------------------------------------
```

```
6179 GTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGC
     CATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCG
---------------------------------------------------------------------------------
6255 AGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGC
     TCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACG
---------------------------------------------------------------------------------
6331 TGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTA
     ACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGAT
---------------------------------------------------------------------------------
6407 AGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTG
     TCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGAC
---------------------------------------------------------------------------------
6483 CACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACAT
     GTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTA
---------------------------------------------------------------------------------
6559 ACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGC
     TGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACATGGAAATCG
---------------------------------------------------------------------------------
6635 TTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACC
     AAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGG
---------------------------------------------------------------------------------
6711 ACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAA
     TGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTT
---------------------------------------------------------------------------------
6787 TTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
     AACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATT
---------------------------------------------------------------------------------
                                                                                white
6863 AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGGATCC
     TCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCCTAGG
---------------------------------------------------------------------------------
6939 ACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATTTTT
     TGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAAAAA
---------------------------------------------------------------------------------
7015 TTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTATTT
     AAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATAAA
---------------------------------------------------------------------------------
7091 GAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTCT
     CTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAGA
---------------------------------------------------------------------------------
7167 CTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTA
     GACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGAT
---------------------------------------------------------------------------------
7243 GCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTA
     CGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAAT
---------------------------------------------------------------------------------
7319 AACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGG
     TTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACC
---------------------------------------------------------------------------------
7395 TTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCG
     AAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGC
---------------------------------------------------------------------------------
```

```
7471 GAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGA
     CTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACT

7547 TCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAA
     AGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTT

7623 TCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCA
     AGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGT

7699 TATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAGATATT
     ATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAA

7775 TTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGG
     AATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCC

7851 TTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAG
     AATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTC

7927 AGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCG
     TCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGC

8003 GTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCA
     CAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGT

8079 GCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCC
     CGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGG

8155 AGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCA
     TCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGT

8231 ATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCC
     TAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGG

8307 GGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGAC
     CCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTG

8383 TGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTAT
     ACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATA

8459 CGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAG
     GCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTC

8535 CGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCG
     GCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGC

8611 GCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCT
     CGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGA

8687 TCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTG
     AGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGAC
```

```
8763 TCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCC
     AGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGG

8839 TTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCG
     AAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGC

8915 ATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAA
     TACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTT

8991 CTACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATC
     GATGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAG

9067 GCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGG
     CGGTTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACC

9143 AGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTG
     TCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACAC

9219 GCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCC
     CGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGG

9295 AGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCC
     TCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGG

9371 ACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCG
     TGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGC

9447 CGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGG
     GCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACC

9523 CCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACC
     GGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGG

9599 AACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTA
     TTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAAT

9675 CTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTC
     GATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAG

9751 GACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGT
     CTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCA

9827 CTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTC
     GAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAG

9903 ACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGT
     TGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACA

9979 CTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATA
     GACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATAT
```

```
10055 CCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTG
      GGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCAC

10131 GAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTA
      CTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAAT

10207 ACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATA
      TGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTAT

10283 TCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTGTTTTTTTTTTTA
      AGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAACAAAAAAAAAAAT

10359 CCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAATAAAAACA
      GGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACACGGTTATTTTTGT

10435 AGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTTGTCTTCTAACCA
      TCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACAGAAGATTGGT

10511 AAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTAT
      TTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATATAAAAAATA

10587 ATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGTTGCTCTTTCGCT
      TATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGCGA

10663 GTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCGCGA
      CAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGCGCT

10739 TTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGTATGTACAAT
      AAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACATACATGTTA

10815 TTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACTTTTTAAG
      AATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGAAAAATTC

10891 GTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTCCTTG
      CATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACAAGGAAC

10967 TGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAACTATAATAATAA
      ACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTTGATATTATTATT

11043 GAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCAT
      CTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGTA

11119 TTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAGT
      AAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCA

11195 CACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCT
      GTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGA

11271 TCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATG
      AGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTAC
```

```
11347  TCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAG
       AGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTC

11423  GCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTC
       CGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAG

11499  ACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATT
       TGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAA

11575  CGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTA
       GCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGAT

                                      5' P
11651  TCGACGGGACCACCTTATGTTATTTCATCATG
       AGCTGCCCTGGTGGAATACAATAAAGTAGTAC
```

pUAST-Kek5Δ45-GFP

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGAC

  77 GGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCT
     CCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGA

 153 CGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAA
     GCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTT

 229 GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCG
     CCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGC

 305 ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTA
     TGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAAT

 381 CCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTC
     GGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAG

 457 GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
     CCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCA

 533 AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
     TTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGT

 609 GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
     CTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAA

 685 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
     ACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG

 761 GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
     CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACT

 837 TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
     AGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTC

 913 GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
     CTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGA

 989 GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
     CTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTG

1065 TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
     AGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGG

1141 ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
     TGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGA

1217 TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA
     AATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGT

1293 ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
     TGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGT

1369 ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
     TGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAG

1445 AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
     TCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGC

1521 TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
     ATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAG

1597 TTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
     AACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGA

1673 TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
     AGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTA

1749 CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
     GAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTA
```

```
1825 AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
     TTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACA

1901 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
     GAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTC

1977 TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
     ACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGC

2053 TCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA
     AGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATT

2129 GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG
     CGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATAC

2205 CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACG
     GCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGC

2281 ACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTA
     TGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGAT

2357 CGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCC
     GCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGG

2433 GGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAAT
     CCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTA

2509 CCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACA
     GGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGT

2585 AATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAG
     TTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTC

2661 CAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGA
     GTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCT

                           3'P
2737 AGCTCACGATGAGAATGGCCAGACCATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACG
     TCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGC

2813 TATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGA
     ATACGAACGTTATTCACGCTCACTTTCCTTATCATAAGACTCACAGCATAACTCAGACTCACTCTGTCGCTATACT

2889 TTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAG
     AACAACTAATTGGGAATCGTACAGGCACCCCAAACTTAATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTC

2965 TTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGG
     AAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCC

3041 AGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGG
     TCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCC

                              UAS sites
3117 AGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATT
     TCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAA

3193 CAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAA
     GTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTT

3269 CAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACT
     GTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGA
```

```
3345 GAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACA GAT
     CTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGT CTA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
         attB1
3420 CTG ACA AGT TTG TAC AAA AAA GCA GGC TCCAGGAAA ATG ATC CTT CTG CTG CTG GGT
     GAC TGT TCA AAC ATG TTT TTT CGT CCG AGGTCCTTT TAC TAG GAA GAC GAC GAC CCA
                                                1▶ M   I   L   L   L   L   G
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3477 GTG CTA GTG GTT CTA ATG GCC CTA CCG CCG CCC ACC GCA GGC ACC ACC GAT TGG ATG
     CAC GAT CAC CAA GAT TAC CGG GAT GGC GGC GGG TGG CGT CCG TGG TGG CTA ACC TAC
   8▶ V   L   V   V   L   M   A   L   P   P   P   T   A   G   T   T   D   W   M
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3534 CAG AGC TGC GGT ACA TGC CAC TGT CAG TGG AAT TCG GGC AAG AAG AGC GCC GAC TGC
     GTC TCG ACG CCA TGT ACG GTG ACA GTC ACC TTA AGC CCG TTC TTC TCG CGG CTG ACG
  27▶ Q   S   C   G   T   C   H   C   Q   W   N   S   G   K   K   S   A   D   C
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3591 AAG AAC AAG GCG CTA ACC AAA ATT CCG CAG GAC ATG AGC AAC GAG ATG CAG GTG CTG
     TTC TTG TTC CGC GAT TGG TTT TAA GGC GTC CTG TAC TCG TTG CTC TAC GTC CAC GAC
  46▶ K   N   K   A   L   T   K   I   P   Q   D   M   S   N   E   M   Q   V   L
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3648 GAC TTT GCC CAC AAT CAA ATA CCC GAG CTG CGG CGC GAA GAG TTC CTA CTG GCC GGT
     CTG AAA CGG GTG TTA GTT TAT GGG CTC GAC GCC GCG CTT CTC AAG GAT GAC CGG CCA
  65▶ D   F   A   H   N   Q   I   P   E   L   R   R   E   E   F   L   L   A   G
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3705 CTG CCC AAT GTG CAC AAG ATC TTT TTG CGC AAC TGC ACC ATC CAG GAG GTG CAT CGC
     GAC GGG TTA CAC GTG TTC TAG AAA AAC GCG TTG ACG TGG TAG GTC CTC CAC GTA GCG
  84▶ L   P   N   V   H   K   I   F   L   R   N   C   T   I   Q   E   V   H   R
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3762 GAG GCC TTC AAG GGT CTG CAT ATC CTA ATC GAG CTG GAC CTG TCG GGC AAT CGG ATA
     CTC CGG AAG TTC CCA GAC GTA TAG GAT TAG CTC GAC CTG GAC AGC CCG TTA GCC TAT
 103▶ E   A   F   K   G   L   H   I   L   I   E   L   D   L   S   G   N   R   I
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3819 CGG GAA CTG CAT CCG GGC ACT TTC GCC GGC CTG GAG AAG CTG CGC AAT GTG ATC ATC
     GCC CTT GAC GTA GGC CCG TGA AAG CGG CCG GAC CTC TTC GAC GCG TTA CAC TAG TAG
 122▶ R   E   L   H   P   G   T   F   A   G   L   E   K   L   R   N   V   I   I
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3876 AAC AAC AAC GAG ATC GAG GTG CTG CCC AAC CAT CTG TTC GTC AAC CTG AGC TTC CTG
     TTG TTG TTG CTC TAG CTC CAC GAC GGG TTG GTA GAC AAG CAG TTG GAC TCG AAG GAC
 141▶ N   N   N   E   I   E   V   L   P   N   H   L   F   V   N   L   S   F   L
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3933 TCG CGC ATC GAG TTC CGG AAC AAT CGA TTG CGC CAG GTG CAG CTG CAC GTC TTC GCT
     AGC GCG TAG CTC AAG GCC TTG TTA GCT AAC GCG GTC CAC GTC GAC GTG CAG AAG CGA
 160▶ S   R   I   E   F   R   N   N   R   L   R   Q   V   Q   L   H   V   F   A
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
3990 GGC ACA ATG GCG CTG AGC GCC ATT TCG CTG GAA CAG AAC CGC CTC TCA CAT CTG CAC
     CCG TGT TAC CGC GAC TCG CGG TAA AGC GAC CTT GTC TTG GCG GAG AGT GTA GAC GTG
 179▶ G   T   M   A   L   S   A   I   S   L   E   Q   N   R   L   S   H   L   H
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4047 AAG GAG ACA TTC AAG GAT CTG CAG AAG CTG ATG CAT CTA TCG CTG CAG GGT AAC GCA
     TTC CTC TGT AAG TTC CTA GAC GTC TTC GAC TAC GTA GAT AGC GAC GTC CCA TTG CGT
 198▶ K   E   T   F   K   D   L   Q   K   L   M   H   L   S   L   Q   G   N   A
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4104 TGG AAC TGC AGC TGC GAG CTG CAG GAC TTT CGC GAC TTT GCG ATC AGC AAA CGG CTC
     ACC TTG ACG TCG ACG CTC GAC GTC CTG AAA GCG CTG AAA CGC TAG TCG TTT GCC GAG
 217▶ W   N   C   S   C   E   L   Q   D   F   R   D   F   A   I   S   K   R   L
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4161 TAC ACA CCG CCC ACC GAT TGC CAG GAG CCG CCA CAG CTG CGC GGC AAG CTG TGG AGC
     ATG TGT GGC GGG TGG CTA ACG GTC CTC GGC GGT GTC GAC GCG CCG TTC GAC ACC TCG
 236▶ Y   T   P   P   T   D   C   Q   E   P   P   Q   L   R   G   K   L   W   S
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
4218 GAG GTG CCA TCG GAG AAC TTC GCC TGC CGG CCG CGC ATT TTG GGT TCC GTG CGC TCC
     CTC CAC GGT AGC CTC TTG AAG CGG ACG GCC GGC GCG TAA AAC CCA AGG CAC GCG AGG
 255▶ E   V   P   S   E   N   F   A   C   R   P   R   I   L   G   S   V   R   S
```

```
4275 TTC ATC GAG GCC AAT CAC GAC AAT ATC TCG CTA CCC TGC CGC ATT GTC GGC AGT CCG
     AAG TAG CTC CGG TTA GTG CTG TTA TAG AGC GAT GGG ACG GCG TAA CAG CCG TCA GGC
 274▸ F   I   E   A   N   H   D   N   I   S   L   P   C   R   I   V   G   S   P

4332 CGT CCC AAT GTC ACC TGG GTG TAC AAC AAG CGG CCA TTG CAG CAG TAC GAC CCG CGT
     GCA GGG TTA CAG TGG ACC CAC ATG TTG TTC GCC GGT AAC GTC GTC ATG CTG GGC GCA
 293▸ R   P   N   V   T   W   V   Y   N   K   R   P   L   Q   Q   Y   D   P   R

4389 GTG CGT GTC CTC ACC TCC GTG GAA CAG ATG CCG GAG CAG CCC TCC CAG GTG CTC ACC
     CAC GCA CAG GAG TGG AGG CAC CTT GTC TAC GGC CTC GTC GGG AGG GTC CAC GAG TGG
 312▸ V   R   V   L   T   S   V   E   Q   M   P   E   Q   P   S   Q   V   L   T

4446 TCG GAG CTG CGC ATC GTG GGC GTA CGG GCC TCC GAC AAG GGT GCC TAC ACC TGT GTG
     AGC CTC GAC GCG TAG CAC CCG CAT GCC CGG AGG CTG TTC CCA CGG ATG TGG ACA CAC
 331▸ S   E   L   R   I   V   G   V   R   A   S   D   K   G   A   Y   T   C   V

4503 GCG GAT AAC CGG GGC GGA CGG GCG GAG GCC GAG TTC CAG CTG CTC GTG AGC GGT GAC
     CGC CTA TTG GCC CCG CCT GCC CGC CTC CGG CTC AAG GTC GAC GAG CAC TCG CCA CTG
 350▸ A   D   N   R   G   G   R   A   E   A   E   F   Q   L   L   V   S   G   D

4560 TAT GCC GGC GCG GTA TCC GCC TCC GAT GGC ATG GGC ATG GGC GCC ATT GGG GCA CCA
     ATA CGG CCG CGC CAT AGG CGG AGG CTA CCG TAC CCG TAC CCG CGG TAA CCC CGT GGT
 369▸ Y   A   G   A   V   S   A   S   D   G   M   G   M   G   A   I   G   A   P

                                             TM
4617 ACC ATT GAT CCG CAA ACG AAC ATG TTT CTC ATC ATC TGT CTA ATC ATT ACG ACG CTG
     TGG TAA CTA GGC GTT TGC TTG TAC AAA GAG TAG TAG ACA GAT TAG TAA TGC TGC GAC
 388▸ T   I   D   P   Q   T   N   M   F   L   I   I   C   L   I   I   T   T   L

4674 CTG CTC CTG CTG CTC GTG GCG GTG CTG ACG CTC TTC TGG TAC TGC CGT CGC ATC AAG
     GAC GAG GAC GAC GAG CAC CGC CAC GAC TGC GAG AAG ACC ATG ACG GCA GCG TAG TTC
 407▸ L   L   L   L   L   V   A   V   L   T   L   F   W   Y   C   R   R   I   K

4731 ACC TAT CAA AAG GAC ACC ACC ATG ATG AGC GGC GAC GGG CTG ATC TCT TCC AAG ATG
     TGG ATA GTT TTC CTG TGG TGG TAC TAC TCG CCG CTG CCC GAC TAG AGA AGG TTC TAC
 426▸ T   Y   Q   K   D   T   T   M   M   S   G   D   G   L   I   S   S   K   M

4788 GAC AAG ACG CAC AAC GGC TCC ATG CTC GAG GGT TCC GTC ATC ATG GAG ATG CAG AAG
     CTG TTC TGC GTG TTG CCG AGG TAC GAG CTC CCA AGG CAG TAG TAC CTC TAC GTC TTC
 445▸ D   K   T   H   N   G   S   M   L   E   G   S   V   I   M   E   M   Q   K

4845 AGC CTG CTC AAC GAG GTC AAT CCA GTC GAG AAG CCG CCA CGG CGC ACG GAC ATC GAG
     TCG GAC GAG TTG CTC CAG TTA GGT CAG CTC TTC GGC GGT GCC GCG TGC CTG TAG CTC
 464▸ S   L   L   N   E   V   N   P   V   E   K   P   P   R   R   T   D   I   E

4902 AGC GTG GAT GGT GGC GAT GAC GTG CTC GAG ATC AAG AAG ACG CTG CTC GAC GAC ACC
     TCG CAC CTA CCA CCG CTA CTG CAC GAG CTC TAG TTC TTC TGC GAC GAG CTG CTG TGG
 483▸ S   V   D   G   G   D   D   V   L   E   I   K   K   T   L   L   D   D   T

4959 GTC TAT ATG GCC AAT CAC TCG CGC GAC GAA GAA GCC GTC TCA GTG GCC ATG TCG GAT
     CAG ATA TAC CGG TTA GTG AGC GCG CTG CTT CTT CGG CAG AGT CAC CGG TAC AGC CTA
 502▸ V   Y   M   A   N   H   S   R   D   E   E   A   V   S   V   A   M   S   D

5016 ACG ACG ACC ACG CCC CGA TCT CGA CAC ACC TAC GTG GAT GAT GCG TAT GCC AAT AGC
     TGC TGC TGG TGC GGG GCT AGA GCT GTG TGG ATG CAC CTA CTA CGC ATA CGG TTA TCG
 521▸ T   T   T   T   P   R   S   R   H   T   Y   V   D   D   A   Y   A   N   S

5073 TTG CCA CCG GAT CTG CTG GCC TTT CCC GCT CGC GTG CCG CCC ACC TCG CCC TCG ATG
     AAC GGT GGC CTA GAC GAC CGG AAA GGG CGA GCG CAC GGC GGG TGG AGC GGG AGC TAC
 540▸ L   P   P   D   L   L   A   F   P   A   R   V   P   P   T   S   P   S   M

5130 CAA TCG TCG CAG TCG AAC ATA CCC GAC CAG GTG ATC TAC GGC ATC CGT TCG CCA CCG
     GTT AGC AGC GTC AGC TTG TAT GGG CTG GTC CAC TAG ATG CCG TAG GCA AGC GGT GGC
 559▸ Q   S   S   Q   S   N   I   P   D   Q   V   I   Y   G   I   R   S   P   P

5187 TCG CTA ACC AGT CCG GTC TAC ACG CAT ATG ACG CCG CAC GGC ATC TAC GGC ACC AAG
     AGC GAT TGG TCA GGC CAG ATG TGC GTA TAC TGC GGC GTG CCG TAG ATG CCG TGG TTC
 578▸ S   L   T   S   P   V   Y   T   H   M   T   P   H   G   I   Y   G   T   K
```

```
5244 ACG ATG ACG GCT CCG CAT AAC GGC TTT ATG ACG CTG CAG CAT CCC AAG TCG CGC AAC
     TGC TAC TGC CGA GGC GTA TTG CCG AAA TAC TGC GAC GTC GTA GGG TTC AGC GCG TTG
 597▸ T   M   T   A   P   H   N   G   F   M   T   L   Q   H   P   K   S   R   N

5301 CTG GCG CTC ATT GCC ACC ACC AAC AGC AGT CGC CAG CAC CAG CAC CAC CAT CAG CTG
     GAC CGC GAG TAA CGG TGG TGG TTG TCG TCA GCG GTC GTG GTC GTG GTG GTA GTC GAC
 616▸ L   A   L   I   A   T   T   N   S   S   R   Q   H   Q   H   H   H   Q   L

5358 CAG CAG CAG CAG CAG CAC CAC CAC CAC CAC CAG CGG CAA CAA CAA CAG CAG CAG CAG
     GTC GTC GTC GTC GTC GTG GTG GTG GTG GTG GTC GCC GTT GTT GTT GTC GTC GTC GTC
 635▸ Q   Q   Q   Q   Q   H   H   H   H   H   Q   R   Q   Q   Q   Q   Q   Q   Q

5415 CAG CAA CAT CCG CTG GCC ACC ACA TCG CCC TTC CTG CCC GCA CCC GTC GTC TAT TCG
     GTC GTT GTA GGC GAC CGG TGG TGT AGC GGG AAG GAC GGG CGT GGG CAG CAG ATA AGC
 654▸ Q   Q   H   P   L   A   T   T   S   P   F   L   P   A   P   V   V   Y   S

5472 CCG GCC ACG GGT GTG GTC ATG AAA CAG GGA TAT ATG ACC ATT CCG CGC AAG CCG CGC
     GGC CGG TGC CCA CAC CAG TAC TTT GTC CCT ATA TAC TGG TAA GGC GCG TTC GGC GCG
 673▸ P   A   T   G   V   V   M   K   Q   G   Y   M   T   I   P   R   K   P   R

5529 GCT CCC AGC TGG GCG CCC AGT ACT TCC GGT GCC GCT GGC CAC GGA TCC ATT CAG CTA
     CGA GGG TCG ACC CGC GGG TCA TGA AGG CCA CGG CGA CCG GTG CCT AGG TAA GTC GAT
 692▸ A   P   S   W   A   P   S   T   S   G   A   A   G   H   G   S   I   Q   L

5586 AGT GAA TTC CAG AGC CCC ACA TCG CCG AAT CCC AGC GAG ACT GGC ACa GCC ACC ACC
     TCA CTT AAG GTC TCG GGG TGT AGC GGC TTA GGG TCG CTC TGA CCG TGt CGG TGG TGG
 711▸ S   E   F   Q   S   P   T   S   P   N   P   S   E   T   G   T   A   T   T

                         IC3-PDZ merge                                 attB2
5643 GCG GAA CTG CAG GCa CTC TTC GAC GAC GAG GGC GAG GAT GGC ACC GAG GTC GAC CCA
     CGC CTT GAC GTC CGt GAG AAG CTG CTG CTC CCG CTC CTA CCG TGG CTC CAG CTG GGT
 730▸ A   E   L   Q   A   L   F   D   D   E   G   E   D   G   T   E   V   D   P

5700 GCT TTC TTG TAC AAA GTG GTG GTA CCG CGG GCC CGG GAT CCA CCG GTC GCC ACC
     CGA AAG AAC ATG TTT CAC CAC CAT GGC GCC CGG GCC CTA GGT GGC CAG CGG TGG
 749▸ A   F   L   Y   K   V   V   V   P   R   A   R   D   P   P   V   A   T

     EGFPN1
5754 ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG
     TAC CAC TCG TTC CCG CTC CTC GAC AAG TGG CCC CAC CAC GGG TAG GAC CAG CTC GAC
 767▸ M   V   S   K   G   E   E   L   F   T   G   V   V   P   I   L   V   E   L

5811 GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC GAG GGC GAT GCC
     CTG CCG CTG CAT TTG CCG GTG TTC AAG TCG CAC AGG CCG CTC CCG CTC CCG CTA CGG
 786▸ D   G   D   V   N   G   H   K   F   S   V   S   G   E   G   E   G   D   A

5868 ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC
     TGG ATG CCG TTC GAC TGG GAC TTC AAG TAG ACG TGG TGG CCG TTC GAC GGG CAC GGG
 805▸ T   Y   G   K   L   T   L   K   F   I   C   T   T   G   K   L   P   V   P

5925 TGG CCC ACC CTC GTG ACC ACC CTG ACC TAC GGC GTG CAG TGC TTC AGC CGC TAC CCC
     ACC GGG TGG GAG CAC TGG TGG GAC TGG ATG CCG CAC GTC ACG AAG TCG GCG ATG GGG
 824▸ W   P   T   L   V   T   T   L   T   Y   G   V   Q   C   F   S   R   Y   P

5982 GAC CAC ATG AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG
     CTG GTG TAC TTC GTC GTG CTG AAG AAG TTC AGG CGG TAC GGG CTT CCG ATG CAG GTC
 843▸ D   H   M   K   Q   H   D   F   F   K   S   A   M   P   E   G   Y   V   Q

6039 GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG
     CTC GCG TGG TAG AAG AAG TTC CTG CTG CCG TTG ATG TTC TGG GCG CGG CTC CAC TTC
 862▸ E   R   T   I   F   F   K   D   D   G   N   Y   K   T   R   A   E   V   K

6096 TTC GAG GGC GAC ACC CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG
     AAG CTC CCG CTG TGG GAC CAC TTG GCG TAG CTC GAC TTC CCG TAG CTG AAG TTC CTC
 881▸ F   E   G   D   T   L   V   N   R   I   E   L   K   G   I   D   F   K   E
```

```
6153 GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT
     CTG CCG TTG TAG GAC CCC GTG TTC GAC CTC ATG TTG ATG TTG TCG GTG TTG CAG ATA
 900▸ D   G   N   I   L   G   H   K   L   E   Y   N   Y   N   S   H   N   V   Y

6210 ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC AAC
     TAG TAC CGG CTG TTC GTC TTC TTG CCG TAG TTC CAC TTG AAG TTC TAG GCG GTG TTG
 919▸ I   M   A   D   K   Q   K   N   G   I   K   V   N   F   K   I   R   H   N

6267 ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC
     TAG CTC CTG CCG TCG CAC GTC GAG CGG CTG GTG ATG GTC GTC TTG TGG GGG TAG CCG
 938▸ I   E   D   G   S   V   Q   L   A   D   H   Y   Q   Q   N   T   P   I   G

6324 GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC CAG TCC GCC CTG AGC
     CTG CCG GGG CAC GAC GAC GGG CTG TTG GTG ATG GAC TCG TGG GTC AGG CGG GAC TCG
 957▸ D   G   P   V   L   L   P   D   N   H   Y   L   S   T   Q   S   A   L   S

6381 AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG ACC GCC GCC
     TTT CTG GGG TTG CTC TTC GCG CTA GTG TAC CAG GAC GAC CTC AAG CAC TGG CGG CGG
 976▸ K   D   P   N   E   K   R   D   H   M   V   L   L   E   F   V   T   A   A

                                                                  SV40 Poly A
6438 GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG TAA AGCGGCCGCGACTCTAGAGGATCTTTGT
     CCC TAG TGA GAG CCG TAC CTG CTC GAC ATG TTC ATT TCGCCGGCGCTGAGATCTCCTAGAAACA
 995▸ G   I   T   L   G   M   D   E   L   Y   K   •

6502 GAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATA
     CTTCCTTGGAATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATAT

6578 AAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTG
     TTTAAAAATTCACATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGAC

6654 ATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGA
     TACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACT

6730 TGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCT
     ACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGA

6806 TCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAA
     AGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTT

6882 AGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTA
     TCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAAT

6958 TAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTG
     ATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAAC

7034 TGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATA
     ACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTAT

7110 ATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATA
     TAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTAT

7186 AAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
     TTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTT

7262 TTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTC
     AAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAG
```

```
                  white gene
7338  TGGATCGGATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAA
      ACCTAGCCTAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7414  AAACATATTTTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGAT
      TTTGTATAAAAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7490  CCAAGGGTATTTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGT
      GGTTCCCATAAACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7566  TTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTG
      AACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7642  TAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTA
      ATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAAT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7718  CGCCATCAATTAAACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGG
      GCGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7794  GGTGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGC
      CCACCACTAACCAAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7870  TATTAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTT
      ATAATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
7946  AAGTTCACTTGATCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAAT
      TTCAAGTGAACTAGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8022  CCCAAACCGCAATCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAA
      GGGTTTGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8098  TTTTTAAGATCATATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAA
      AAAAATTCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8174  CAACCAGATATTTTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCG
      GTTGGTCTATAAAATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8250  TTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGA
      AAAACAAGCCCCAATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8326  GCGGCTTCGCAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGC
      CGCCGAAGCGTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8402  CGGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGG
      GCCTCCTGAGGCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8478  GGCGGTCAATCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACAC
      CCGCCAGTTAGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8554  ATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTA
      TATGGCCGCGGGTCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAAT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
8630 ATAAAGGGTCCAATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGC
     TATTTCCCAGGTTAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCG

8706 AGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCAT
     TCAAGGCCACGGCCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTA

8782 CCGGGATGCGACTGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGA
     GGCCCTACGCTGACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCT

8858 TGACCTCTTTATCGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTG
     ACTGGAGAAATAGCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGAC

8934 ACCTATCGGCAGCGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCA
     TGGATAGCCGTCGCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGT

9010 TCGGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAAC
     AGCCACACGGGCCGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTG

9086 CGATCCGCCGCTTCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTG
     GCTAGGCGGCGAAGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCAC

9162 CTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCT
     GACTTCTTCGACAGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGA

9238 TTGACAAGATCCTTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTC
     AACTGTTCTAGGAAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAG

9314 CTAGTGAGTTCGATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCA
     GATCACTCAAGCTACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGT

9390 GTGTCCTACCAACTACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCC
     CACAGGATGGTTGATGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGG

9466 CGTGATCGGATCGCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCA
     GCACTAGCCTAGCGGTTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGT

9542 CCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCG
     GGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGC

9618 GGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACG
     CCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGC

9694 GTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAA
     CACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTT

9770 TTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATT
     AACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAA

9846 TCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGC
     AGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACG
```

```
 9922 CATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTC
      GTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAG

 9998 CTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCAT
      GAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTA

10074 ATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGG
      TAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCC

10150 AGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCAC
      TCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTG

10226 AGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGC
      TCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACG

10302 CTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCT
      GACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGA

10378 CGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTC
      GCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAG

10454 GGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAA
      CCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTT

10530 TGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCC
      ACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGG

10606 TGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCG
      ACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGC

10682 GGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTG
      CCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAAC

10758 TTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTG
      AAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACAC

10834 CCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTT
      GGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAA

10910 GTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGA
      CAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCT

10986 TATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGT
      ATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCA

11062 TGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTT
      ACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAA

11138 AGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATT
      TCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAA
```

```
11214 TGTATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTA
      ACATACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAAT

11290 TGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGT
      ACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCA

11366 GTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAA
      CAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTT

11442 CTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTC
      GATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAG

11518 TAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGAC
      ATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTG

11594 CTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCA
      GATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGT

11670 CTGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCC
      GACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGG

11746 TCCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGA
      AGGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCT

11822 GTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAAT
      CAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTA

11898 TTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTT
      AAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAA

11974 CAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGT
      GTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCA

                                                 5' P
12050 AAGCTTCGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      TTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
```

pUAST-Kek5ΔIC+PC-GFP

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACT
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGA

  76 GGGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACC
     CCCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGG

 151 CTCGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAA
     GAGCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTT

 226 GAAGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
     CTTCCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTT

 301 CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
     GGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGG

 376 GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCT
     CGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGA

 451 CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
     GTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAA

 526 ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
     TAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTC

 601 GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
     CTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTC

 676 GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAA
     CTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTT

 751 ACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
     TGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCT

 826 AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
     TCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTC

 901 ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGA
     TAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACT

 976 GTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC
     CATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAG

1051 CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
     GTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTA

1126 GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
     CTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTC

1201 AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCC
     TTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGG

1276 AGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
     TCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAG

1351 ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT
     TAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAA

1426 CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTC
     GCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAG

1501 TCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG
     AGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCAC

1576 TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
     ATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCA

1651 GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTA
     CGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACAT

1726 ACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
     TGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTC
```

```
1801 GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTA
     CGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAAT

1876 TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGG
     AACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCC

1951 GGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAA
     CCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATT

2026 AAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCT
     TTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGA

2101 CCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGT
     GGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACA

2176 TGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGA
     ACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACT

2251 AATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGA
     TTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCT

2326 GAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGA
     CTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCT

2401 TTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATG
     AAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATAC

2476 CAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGT
     GTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACA

2551 ATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTA
     TAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCAT

2626 AACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGG
     TTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCC

                                                                    3'P
2701 CCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATAACA
     GGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGT

2776 TAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTC
     ATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAAG

2851 TGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAA
     ACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTT

2926 TTAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTC
     AATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAAATAAAAATTATTAAACGCTCATGCGTTTCGAAG

3001 TGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAG
     ACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTC

                                                                       UAS sites
3076 CGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCG
     GCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGC

3151 CCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAA
     GGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATT

3226 GCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAAT
     CGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTA
```

```
3301 CAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACA
     GTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGT

                                                      attB1
3376 AGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACA GAT CTG ACA AGT TTG TAC AAA AAA
     TCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGT CTA GAC TGT TCA AAC ATG TTT TTT

3441 GCA GGC TCCAGGAAA ATG ATC CTT CTG CTG CTG GGT GTG CTA GTG GTT CTA ATG GCC
     CGT CCG AGGTCCTTT TAC TAG GAA GAC GAC GAC CCA CAC GAT CAC CAA GAT TAC CGG
                     1▶ M   I   L   L   L   L   G   V   L   V   V   L   M   A

3498 CTA CCG CCG CCC ACC GCA GGC ACC ACC GAT TGG ATG CAG AGC TGC GGT ACA TGC CAC
     GAT GGC GGC GGG TGG CGT CCG TGG TGG CTA ACC TAC GTC TCG ACG CCA TGT ACG GTG
  15▶ L   P   P   P   T   A   G   T   T   D   W   M   Q   S   C   G   T   C   H

3555 TGT CAG TGG AAT TCG GGC AAG AAG AGC GCC GAC TGC AAG AAC AAG GCG CTA ACC AAA
     ACA GTC ACC TTA AGC CCG TTC TTC TCG CGG CTG ACG TTC TTG TTC CGC GAT TGG TTT
  34▶ C   Q   W   N   S   G   K   K   S   A   D   C   K   N   K   A   L   T   K

3612 ATT CCG CAG GAC ATG AGC AAC GAG ATG CAG GTG CTG GAC TTT GCC CAC AAT CAA ATA
     TAA GGC GTC CTG TAC TCG TTG CTC TAC GTC CAC GAC CTG AAA CGG GTG TTA GTT TAT
  53▶ I   P   Q   D   M   S   N   E   M   Q   V   L   D   F   A   H   N   Q   I

3669 CCC GAG CTG CGG CGC GAA GAG TTC CTA CTG GCC GGT CTG CCC AAT GTG CAC AAG ATC
     GGG CTC GAC GCC GCG CTT CTC AAG GAT GAC CGG CCA GAC GGG TTA CAC GTG TTC TAG
  72▶ P   E   L   R   R   E   E   F   L   L   A   G   L   P   N   V   H   K   I

3726 TTT TTG CGC AAC TGC ACC ATC CAG GAG GTG CAT CGC GAG GCC TTC AAG GGT CTG CAT
     AAA AAC GCG TTG ACG TGG TAG GTC CTC CAC GTA GCG CTC CGG AAG TTC CCA GAC GTA
  91▶ F   L   R   N   C   T   I   Q   E   V   H   R   E   A   F   K   G   L   H

3783 ATC CTA ATC GAG CTG GAC CTG TCG GGC AAT CGG ATA CGG GAA CTG CAT CCG GGC ACT
     TAG GAT TAG CTC GAC CTG GAC AGC CCG TTA GCC TAT GCC CTT GAC GTA GGC CCG TGA
 110▶ I   L   I   E   L   D   L   S   G   N   R   I   R   E   L   H   P   G   T

3840 TTC GCC GGC CTG GAG AAG CTG CGC AA T GTG ATC ATC AAC AAC AAC GAG ATC GAG
     AAG CGG CCG GAC CTC TTC GAC GCG TT A CAC TAG TAG TTG TTG TTG CTC TAG CTC
 129▶ F   A   G   L   E   K   L   R   N    V   I   I   N   N   N   E   I   E

3894 GTG CTG CCC AAC CAT CTG TTC GTC AAC CTG AGC TTC CTG TCG CGC ATC GAG TTC CGG
     CAC GAC GGG TTG GTA GAC AAG CAG TTG GAC TCG AAG GAC AGC GCG TAG CTC AAG GCC
 147▶ V   L   P   N   H   L   F   V   N   L   S   F   L   S   R   I   E   F   R

3951 AAC AAT CGA TTG CGC CAG GTG CAG CTG CAC GTC TTC GCT GGC ACA ATG GCG CTG AGC
     TTG TTA GCT AAC GCG GTC CAC GTC GAC GTG CAG AAG CGA CCG TGT TAC CGC GAC TCG
 166▶ N   N   R   L   R   Q   V   Q   L   H   V   F   A   G   T   M   A   L   S

4008 GCC ATT TCG CTG GAA CAG AAC CGC CTC TCA CAT CTG CAC AAG GAG ACA TTC AAG GAT
     CGG TAA AGC GAC CTT GTC TTG GCG GAG AGT GTA GAC GTG TTC CTC TGT AAG TTC CTA
 185▶ A   I   S   L   E   Q   N   R   L   S   H   L   H   K   E   T   F   K   D

4065 CTG CAG AAG CTG ATG CAT CTA TCG CTG CAG GGT AAC GCA TGG AAC TGC AGC TGC GAG
     GAC GTC TTC GAC TAC GTA GAT AGC GAC GTC CCA TTG CGT ACC TTG ACG TCG ACG CTC
 204▶ L   Q   K   L   M   H   L   S   L   Q   G   N   A   W   N   C   S   C   E

4122 CTG CAG GAC TTT CGC GAC TTT GCG ATC AGC AAA CGG CTC TAC ACA CCG CCC ACC GAT
     GAC GTC CTG AAA GCG CTG AAA CGC TAG TCG TTT GCC GAG ATG TGT GGC GGG TGG CTA
 223▶ L   Q   D   F   R   D   F   A   I   S   K   R   L   Y   T   P   P   T   D

4179 TGC CAG GAG CCG CCA CAG CTG CGC GGC AAG CTG TGG AGC GAG GTG CCA TCG GAG AAC
     ACG GTC CTC GGC GGT GTC GAC GCG CCG TTC GAC ACC TCG CTC CAC GGT AGC CTC TTG
 242▶ C   Q   E   P   P   Q   L   R   G   K   L   W   S   E   V   P   S   E   N
```

```
4236 TTC GCC TGC CGG CCG CGC ATT TTG GGT TCC GTG CGC TCC TTC ATC GAG GCC AAT CAC
     AAG CGG ACG GCC GGC GCG TAA AAC CCA AGG CAC GCG AGG AAG TAG CTC CGG TTA GTG
 261▸ F   A   C   R   P   R   I   L   G   S   V   R   S   F   I   E   A   N   H
-------------------------------------------------------------------------------
4293 GAC AAT ATC TCG CTA CCC TGC CGC ATT GTC GGC AGT CCG CGT CCC AAT GTC ACC TGG
     CTG TTA TAG AGC GAT GGG ACG GCG TAA CAG CCG TCA GGC GCA GGG TTA CAG TGG ACC
 280▸ D   N   I   S   L   P   C   R   I   V   G   S   P   R   P   N   V   T   W
-------------------------------------------------------------------------------
4350 GTG TAC AAC AAG CGG CCA TTG CAG CAG TAC GAC CCG CGT GTG CGT GTC CTC ACC TCC
     CAC ATG TTG TTC GCC GGT AAC GTC GTC ATG CTG GGC GCA CAC GCA CAG GAG TGG AGG
 299▸ V   Y   N   K   R   P   L   Q   Q   Y   D   P   R   V   R   V   L   T   S
-------------------------------------------------------------------------------
4407 GTG GAA CAG ATG CCG GAG CAG CCC TCC CAG GTG CTC ACC TCG GAG CTG CGC ATC GTG
     CAC CTT GTC TAC GGC CTC GTC GGG AGG GTC CAC GAG TGG AGC CTC GAC GCG TAG CAC
 318▸ V   E   Q   M   P   E   Q   P   S   Q   V   L   T   S   E   L   R   I   V
-------------------------------------------------------------------------------
4464 GGC GTA CGG GCC TCC GAC AAG GGT GCC TAC ACC TGT GTG GCG GAT AAC CGG GGC GGA
     CCG CAT GCC CGG AGG CTG TTC CCA CGG ATG TGG ACA CAC CGC CTA TTG GCC CCG CCT
 337▸ G   V   R   A   S   D   K   G   A   Y   T   C   V   A   D   N   R   G   G
-------------------------------------------------------------------------------
4521 CGG GCG GAG GCC GAG TTC CAG CTG CTC GTG AGC GGT GAC TAT GCC GGC GCG GTA TCC
     GCC CGC CTC CGG CTC AAG GTC GAC GAG CAC TCG CCA CTG ATA CGG CCG CGC CAT AGG
 356▸ R   A   E   A   E   F   Q   L   L   V   S   G   D   Y   A   G   A   V   S
-------------------------------------------------------------------------------
4578 GCC TCC GAT GGC ATG GGC ATG GGC GCC ATT GGG GCA CCA ACC ATT GAT CCG CAA ACG
     CGG AGG CTA CCG TAC CCG TAC CCG CGG TAA CCC CGT GGT TGG TAA CTA GGC GTT TGC
 375▸ A   S   D   G   M   G   M   G   A   I   G   A   P   T   I   D   P   Q   T
-------------------------------------------------------------------------------
                     TM
4635 AAC ATG TTT CTC ATC ATC TGT CTA ATC ATT ACG ACG CTG CTG CTC CTG CTG CTC GTG
     TTG TAC AAA GAG TAG TAG ACA GAT TAG TAA TGC TGC GAC GAC GAG GAC GAC GAG CAC
 394▸ N   M   F   L   I   I   C   L   I   I   T   T   L   L   L   L   L   L   V
-------------------------------------------------------------------------------
                                                             TM-PDZ merge
4692 GCG GTG CTG ACG CTC TTC TGG TAC TGC CGT CGC ATC AAG ACa CTC TTC GAt GAC GAG
     CGC CAC GAC TGC GAG AAG ACC ATG ACG GCA GCG TAG TTC TGt GAG AAG CTa CTG CTC
 413▸ A   V   L   T   L   F   W   Y   C   R   R   I   K   T   L   F   D   D   E
-------------------------------------------------------------------------------
                                 attB2
4749 GGC GAG GAT GGC ACC GAG GTC GAC CCA GCT TTC TTG TAC AAA GTG GTG GTA
     CCG CTC CTA CCG TGG CTC CAG CTG GGT CGA AAG AAC ATG TTT CAC CAC CAT
 432▸ G   E   D   G   T   E   V   D   P   A   F   L   Y   K   V   V   V
-------------------------------------------------------------------------------
                                             EGFPN1
4800 CCG CGG GCC CGG GAT CCA CCG GTC GCC ACC ATG GTG AGC AAG GGC GAG GAG CTG TTC
     GGC GCC CGG GCC CTA GGT GGC CAG CGG TGG TAC CAC TCG TTC CCG CTC CTC GAC AAG
 449▸ P   R   A   R   D   P   P   V   A   T   M   V   S   K   G   E   E   L   F
-------------------------------------------------------------------------------
4857 ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC
     TGG CCC CAC CAC GGG TAG GAC CAG CTC GAC CTG CCG CTG CAT TTG CCG GTG TTC AAG
 468▸ T   G   V   V   P   I   L   V   E   L   D   G   D   V   N   G   H   K   F
-------------------------------------------------------------------------------
4914 AGC GTG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC
     TCG CAC AGG CCG CTC CCG CTC CCG CTA CGG TGG ATG CCG TTC GAC TGG GAC TTC AAG
 487▸ S   V   S   G   E   G   E   G   D   A   T   Y   G   K   L   T   L   K   F
-------------------------------------------------------------------------------
4971 ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC CTG ACC
     TAG ACG TGG TGG CCG TTC GAC GGG CAC GGG ACC GGG TGG GAG CAC TGG TGG GAC TGG
 506▸ I   C   T   T   G   K   L   P   V   P   W   P   T   L   V   T   T   L   T
-------------------------------------------------------------------------------
5028 TAC GGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC
     ATG CCG CAC GTC ACG AAG TCG GCG ATG GGG CTG GTG TAC TTC GTC GTG CTG AAG AAG
 525▸ Y   G   V   Q   C   F   S   R   Y   P   D   H   M   K   Q   H   D   F   F
-------------------------------------------------------------------------------
5085 AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC
     TTC AGG CGG TAC GGG CTT CCG ATG CAG GTC CTC GCG TGG TAG AAG AAG TTC CTG CTG
 544▸ K   S   A   M   P   E   G   Y   V   Q   E   R   T   I   F   F   K   D   D
-------------------------------------------------------------------------------
```

```
5142 GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CGC
     CCG TTG ATG TTC TGG GCG CGG CTC CAC TTC AAG CTC CCG CTG TGG GAC CAC TTG GCG
 563▸ G   N   Y   K   T   R   A   E   V   K   F   E   G   D   T   L   V   N   R
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5199 ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG CTG
     TAG CTC GAC TTC CCG TAG CTG AAG TTC CTC CTG CCG TTG TAG GAC CCC GTG TTC GAC
 582▸ I   E   L   K   G   I   D   F   K   E   D   G   N   I   L   G   H   K   L
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5256 GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC
     CTC ATG TTG ATG TTG TCG GTG TTG CAG ATA TAG TAC CGG CTG TTC GTC TTC TTG CCG
 601▸ E   Y   N   Y   N   S   H   N   V   Y   I   M   A   D   K   Q   K   N   G
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5313 ATC AAG GTG AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC
     TAG TTC CAC TTG AAG TTC TAG GCG GTG TTG TAG CTC CTG CCG TCG CAC GTC GAG CGG
 620▸ I   K   V   N   F   K   I   R   H   N   I   E   D   G   S   V   Q   L   A
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5370 GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC
     CTG GTG ATG GTC GTC TTG TGG GGG TAG CCG CTG CCG GGG CAC GAC GAC GGG CTG TTG
 639▸ D   H   Y   Q   Q   N   T   P   I   G   D   G   P   V   L   L   P   D   N
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5427 CAC TAC CTG AGC ACC CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC
     GTG ATG GAC TCG TGG GTC AGG CGG GAC TCG TTT CTG GGG TTG CTC TTC GCG CTA GTG
 658▸ H   Y   L   S   T   Q   S   A   L   S   K   D   P   N   E   K   R   D   H
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5484 ATG GTC CTG CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG
     TAC CAG GAC GAC CTC AAG CAC TGG CGG CGG CCC TAG TGA GAG CCG TAC CTG CTC GAC
 677▸ M   V   L   L   E   F   V   T   A   A   G   I   T   L   G   M   D   E   L
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                                  SV40 Poly A
5541 TAC AAG TAA AGCGGCCGCGACTCTAGAGGATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGG
     ATG TTC ATT TCGCCGGCGCTGAGATCTCCTAGAAACACTTCCTTGGAATGAAGACACCACACTGTATTAACC
 696▸ Y   K   •
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5613 ACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGA
     TGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCACATATTACACAATTTGATGACT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5688 TTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATG
     AAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCGTCACCACCTTACGGAAATTAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5763 AGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTC
     TCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5838 CTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTG
     GAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5913 TGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAA
     ACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
5988 TTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTC
     AATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
6063 CACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAG
     GTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
6138 GGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACCACATTTGTAGAGGTT
     CCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGGTGTAAACATCTCCAA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
6213 TTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAAC
     AATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTG
```

```
6288 TTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCA
     AACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGT

                                                                          white gene
6363 CTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGGATCCACTAGAAGGCCT
     GACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCCTAGGTGATCTTCCGGA

6438 TAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATTTTTTTTTTATTTTT
     ATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAAAAAAAAAATAAAAA

6513 TACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTATTTGAAGTACCAG
     ATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATAAACTTCATGGTC

6588 GTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTCTCTGTCCACA
     CAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAGAGACAGGTGT

6663 GAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTAGCGTCAAT
     CTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGATCGCAGTTA

6738 GTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTAAACACAA
     CAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAATTTGTGTT

6813 AGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGGTTTTGG
     TCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACCAAAACC

6888 GTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCGGAGGC
     CACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGCCTCCG

6963 AGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGATCTA
     TCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACTAGAT

7038 TAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAATCG
     ATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTTAGC

7113 CACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCATA
     GTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGTAT

7188 TCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAGATATTT
     AGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAAA

7263 TAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGG
     ATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCC

7338 TTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCA
     AATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGT

7413 GAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTC
     CTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAG

7488 CGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAA
     GCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTT
```

```
7563 TCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGC
     AGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCG

7638 GCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGG
     CGGGTCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCC

7713 GTCCAATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCC
     CAGGTTAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGG

7788 GGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGG
     CCACGGCCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCC

7863 ATGCGACTGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGAC
     TACGCTGACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTG

7938 CTCTTTATCGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACC
     GAGAAATAGCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGG

8013 TATCGGCAGCGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATC
     ATAGCCGTCGCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAG

8088 GGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACC
     CCACACGGGCCGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGG

8163 GATCCGCCGCTTCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTG
     CTAGGCGGCGAAGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCAC

8238 CTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTC
     GACTTCTTCGACAGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAG

8313 TTTGACAAGATCCTTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTT
     AAACTGTTCTAGGAAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAA

8388 TCCTAGTGAGTTCGATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGC
     AGGATCACTCAAGCTACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACG

8463 CCAGTGTCCTACCAACTACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGA
     GGTCACAGGATGGTTGATGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCT

8538 GTCCCGTGATCGGATCGCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTT
     CAGGGCACTAGCCTAGCGGTTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAA

8613 GGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCA
     CCGGTGGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGT

8688 GTTCCGGGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCA
     CAAGGCCCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGT

8763 GACAACGGTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTT
     CTGTTGCCACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAA
```

```
 8838 AGTTAGAATTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAAT
      TCAATCTTAACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTA

 8913 AAATAAATTTCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGC
      TTTATTTAAAGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACG

 8988 AGATGGTTGCCATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACG
      TCTACCAACGGTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGC

 9063 GAGCCATCTTCCTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGA
      CTCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCT

 9138 ATACATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCA
      TATGTAAACGTATAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGT

 9213 GTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTA
      CAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAAT

 9288 CCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTG
      GGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGAC

 9363 CACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGC
      GTGAAGAAGTTGACGGACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACG

 9438 GCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTC
      CGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAG

 9513 TTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAG
      AAGAACTTGAGCCCGAGCCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTC

 9588 GGTCTGCTGATTAACCAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCC
      CCAGACGACTAATTGGTTACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGG

 9663 AGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCC
      TCAAGCCCGTTCCAGTAGGACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGG

 9738 ATTCTCATCGTGAGCTTCCGGGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACA
      TAAGAGTAGCACTCGAAGGCCCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGT

 9813 TATATCCGAAATAACTGCTTGTTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAA
      ATATAGGCTTTATTGACGAACAAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTT

 9888 AGCTAATGTAATTATATTTGTGCCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACAT
      TCGATTACATTAATATAAACACGGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTA

 9963 CCCCACAAGTAGACTTTGGATTTGTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGT
      GGGGTGTTCATCTGAAACCTAAACAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCA

10038 TTATCGCTACATAAAACACCGGGATATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCA
      AATAGCGATGTATTTTGTGGCCCTATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGT
```

```
10113 CCTCCACCACACCACGTTTCGTAGTTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGT
      GGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACA

10188 TCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCG
      AGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGC

10263 TCTCAATAGAAATTAGTAATAAATATTTGTATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCA
      AGAGTTATCTTTAATCATTATTTATAAACATACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGT

10338 CAGCTATATTTATTCTAATTTAATATTATGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGT
      GTCGATATAAATAAGATTAAATTATAATACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCA

10413 TACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAAT
      ATGTTAAACTTGACTTTCACTGTAGGTCACAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTA

10488 AGTGTTGTTTATATATATCAAAAATAACAACTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTC
      TCACAACAAATATATATAGTTTTTATTGTTGATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAG

10563 ACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCA
      TGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGT

10638 AGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTT
      TCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAA

10713 ACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATC
      TGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAG

10788 ACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCAC
      TGAAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTG

10863 TGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAAT
      ACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTA

10938 TTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATAT
      AAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATA

11013 TTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAA
      AAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTT

11088 CAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGACGGGACCACCTTATGTTATTT
      GTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAA

            5' P
11163 CATCATG
      GTAGTAC
```

pUAST-AMIGO1-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTGG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGACC

  78 GGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCTCG
     CCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGAGC

 155 TTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAAGGC
     AACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTTCCG

 232 GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
     CAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTC

 309 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
     CTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCT

 386 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
     ATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACAT

 463 GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
     CCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAG

 540 GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
     CAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCC

 617 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
     ATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGAC

 694 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
     GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGC

 771 GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
     CACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGC

 848 GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
     CCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGAT

 925 GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
     CTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTA

1002 GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
     CGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATC

1079 ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
     TATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

1156 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
     TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGG

1233 AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
     TCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGA

1310 ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
     TGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATG

1387 ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
     TACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTC

1464 TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
     ACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGA

1541 GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGA
     CCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCT

1618 TAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
     ATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCT

1695 TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
     AGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGG

1772 AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
     TCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTA
```

```
1849 ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
     TGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTA

1926 GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
     CATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAA

2003 ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
     TAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACT

2080 AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
     TTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAG

2157 AGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
     TCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCAC

2234 CACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGG
     GTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCC

2311 TGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGA
     ACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACT

2388 AGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGA
     TCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACT

2465 TGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGA
     ACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCT

2542 AATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATA
     TTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTAT

2619 AACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCG
     TTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGC

                                                                           3'P
2696 TCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATA
     AGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTAT

2773 ACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATT
     TGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAA

2850 CTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAAT
     GACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTTA

2927 TAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGC
     ATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACG

3004 ATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG
     TACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTC

                                                                   UAS sites
3081 TACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGT
     ATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCA

3158 ATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCT
     TATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGA

3235 AAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGT
     TTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCA

3312 AACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCT
     TTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGA
```

```
                         Polylinker          attB1  (hybrid)
3389 GAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGTACAAAAAAGTTGGAAA ATG CAC CCC CAC
     CTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAAACATGTTTTTTCAACCTTT TAC GTG GGG GTG

                                                                1▸ M   H   P   H

3461 CGT GAC CCG AGA GGC CTC TGG CTC CTG CTG CCG TCC TTG TCC CTG CTG CTT TTT GAG
     GCA CTG GGC TCT CCG GAG ACC GAG GAC GAC GGC AGG AAC AGG GAC GAC GAA AAA CTC

   5▸ R   D   P   R   G   L   W   L   L   L   P   S   L   S   L   L   L   F   E

3518 GTG GCC AGA GCT GGC CGA GCC GTG GTT AGC TGT CCT GCC GCC TGC TTG TGC GCC AGC
     CAC CGG TCT CGA CCG GCT CGG CAC CAA TCG ACA GGA CGG CGG ACG AAC ACG CGG TCG

  24▸ V   A   R   A   G   R   A   V   V   S   C   P   A   A   C   L   C   A   S

3575 AAC ATC CTC AGC TGC TCC AAG CAG CAG CTG CCC AAT GTG CCC CAT TCC CTG CCC AGT
     TTG TAG GAG TCG ACG AGG TTC GTC GTC GAC GGG TTA CAC GGG GTA AGG GAC GGG TCA

  43▸ N   I   L   S   C   S   K   Q   Q   L   P   N   V   P   H   S   L   P   S

3632 TAC ACA GCA CTA CTG GAC CTC AGT CAC AAC AAC CTG AGC CGC CTG CGG GCC GAG TGG
     ATG TGT CGT GAT GAC CTG GAG TCA GTG TTG TTG GAC TCG GCG GAC GCC CGG CTC ACC

  62▸ Y   T   A   L   L   D   L   S   H   N   N   L   S   R   L   R   A   E   W

3689 ACC CCC ACG CGC CTG ACC CAA CTG CAC TCC CTG CTG CTG AGC CAC AAC CAC CTG AAC
     TGG GGG TGC GCG GAC TGG GTT GAC GTG AGG GAC GAC GAC TCG GTG TTG GTG GAC TTG

  81▸ T   P   T   R   L   T   Q   L   H   S   L   L   L   S   H   N   H   L   N

3746 TTC ATC TCC TCT GAG GCC TTT TCC CCG GTA CCC AAC CTG CGC TAC CTG GAC CTC TCC
     AAG TAG AGG AGA CTC CGG AAA AGG GGC CAT GGG TTG GAC GCG ATG GAC CTG GAG AGG

 100▸ F   I   S   S   E   A   F   S   P   V   P   N   L   R   Y   L   D   L   S

3803 TCC AAC CAG CTG CGT ACA CTG GAT GAG TTC CTG TTC AGT GAC CTG CAA GTA CTG GAG
     AGG TTG GTC GAC GCA TGT GAC CTA CTC AAG GAC AAG TCA CTG GAC GTT CAT GAC CTC

 119▸ S   N   Q   L   R   T   L   D   E   F   L   F   S   D   L   Q   V   L   E

3860 GTG CTG CTG CTC TAC AAT AAC CAC ATC ATG GCG GTG GAC CGG TGC GCC TTC GAT GAC
     CAC GAC GAC GAG ATG TTA TTG GTG TAG TAC CGC CAC CTG GCC ACG CGG AAG CTA CTG

 138▸ V   L   L   L   Y   N   N   H   I   M   A   V   D   R   C   A   F   D   D

3917 ATG GCC CAG CTG CAG AAA CTC TAC TTG AGC CAG AAC CAG ATC TCT CGC TTC CCT CTG
     TAC CGG GTC GAC GTC TTT GAG ATG AAC TCG GTC TTG GTC TAG AGA GCG AAG GGA GAC

 157▸ M   A   Q   L   Q   K   L   Y   L   S   Q   N   Q   I   S   R   F   P   L

3974 GAA CTG GTC AAG GAA GGA GCC AAG CTA CCC AAA CTA ACG CTC CTG GAT CTC TCT TCT
     CTT GAC CAG TTC CTT CCT CGG TTC GAT GGG TTT GAT TGC GAG GAC CTA GAG AGA AGA

 176▸ E   L   V   K   E   G   A   K   L   P   K   L   T   L   L   D   L   S   S

4031 AAC AAG CTG AAG AAC TTG CCA TTG CCT GAC CTG CAG AAG CTG CCG GCC TGG ATC AAG
     TTG TTC GAC TTC TTG AAC GGT AAC GGA CTG GAC GTC TTC GAC GGC CGG ACC TAG TTC

 195▸ N   K   L   K   N   L   P   L   P   D   L   Q   K   L   P   A   W   I   K

4088 AAT GGG CTG TAC CTA CAT AAC AAC CCC CTG AAC TGC GAC TGT GAG CTC TAC CAG CTG
     TTA CCC GAC ATG GAT GTA TTG TTG GGG GAC TTG ACG CTG ACA CTC GAG ATG GTC GAC

 214▸ N   G   L   Y   L   H   N   N   P   L   N   C   D   C   E   L   Y   Q   L
```

```
4145 TTT TCA CAC TGG CAG TAT CGG CAG CTG AGC TCC GTG ATG GAC TTT CAA GAG GAT CTG
     AAA AGT GTG ACC GTC ATA GCC GTC GAC TCG AGG CAC TAC CTG AAA GTT CTC CTA GAC

 233▶ F   S   H   W   Q   Y   R   Q   L   S   S   V   M   D   F   Q   E   D   L
4202 TAC TGC ATG AAC TCC AAG AAG CTG CAC AAT GTC TTC AAC CTG AGT TTC CTC AAC TGT
     ATG ACG TAC TTG AGG TTC TTC GAC GTG TTA CAG AAG TTG GAC TCA AAG GAG TTG ACA

 252▶ Y   C   M   N   S   K   K   L   H   N   V   F   N   L   S   F   L   N   C
4259 GGC GAG TAC AAG GAG CGT GCC TGG GAG GCC CAC CTG GGT GAC ACC TTG ATC ATC AAG
     CCG CTC ATG TTC CTC GCA CGG ACC CTC CGG GTG GAC CCA CTG TGG AAC TAG TAG TTC

 271▶ G   E   Y   K   E   R   A   W   E   A   H   L   G   D   T   L   I   I   K
4316 TGT GAC ACC AAG CAG CAA GGG ATG ACC AAG GTG TGG GTG ACA CCA AGT AAT GAA CGG
     ACA CTG TGG TTC GTC GTT CCC TAC TGG TTC CAC ACC CAC TGT GGT TCA TTA CTT GCC

 290▶ C   D   T   K   Q   Q   G   M   T   K   V   W   V   T   P   S   N   E   R
4373 GTG CTA GAT GAG GTG ACC AAT GGC ACA GTG AGT GTG TCT AAG GAT GGC AGT CTT CTT
     CAC GAT CTA CTC CAC TGG TTA CCG TGT CAC TCA CAC AGA TTC CTA CCG TCA GAA GAA

 309▶ V   L   D   E   V   T   N   G   T   V   S   V   S   K   D   G   S   L   L
4430 TTC CAG CAG GTG CAG GTC GAG GAC GGT GGT GTG TAT ACC TGC TAT GCC ATG GGA GAG
     AAG GTC GTC CAC GTC CAG CTC CTG CCA CCA CAC ATA TGG ACG ATA CGG TAC CCT CTC

 328▶ F   Q   Q   V   Q   V   E   D   G   G   V   Y   T   C   Y   A   M   G   E
4487 ACT TTC AAT GAG ACA CTG TCT GTG GAA TTG AAA GTG CAC AAT TTC ACC TTG CAC GGA
     TGA AAG TTA CTC TGT GAC AGA CAC CTT AAC TTT CAC GTG TTA AAG TGG AAC GTG CCT

 347▶ T   F   N   E   T   L   S   V   E   L   K   V   H   N   F   T   L   H   G
                         TM
4544 CAC CAT GAC ACC CTC AAC ACA GCC TAT ACC ACC CTA GTG GGC TGT ATC CTT AGT GTG
     GTG GTA CTG TGG GAG TTG TGT CGG ATA TGG TGG GAT CAC CCG ACA TAG GAA TCA CAC

 366▶ H   H   D   T   L   N   T   A   Y   T   T   L   V   G   C   I   L   S   V
4601 GTC CTG GTC CTC ATA TAC CTA TAC CTC ACC CCT TGC CGC TGC TGG TGC CGG GGT GTA
     CAG GAC CAG GAG TAT ATG GAT ATG GAG TGG GGA ACG GCG ACG ACC ACG GCC CCA CAT

 385▶ V   L   V   L   I   Y   L   Y   L   T   P   C   R   C   W   C   R   G   V
4658 GAG AAG CCT TCC AGC CAT CAA GGA GAC AGC CTC AGC TCT TCC ATG CTT AGT ACC ACA
     CTC TTC GGA AGG TCG GTA GTT CCT CTG TCG GAG TCG AGA AGG TAC GAA TCA TGG TGT

 404▶ E   K   P   S   S   H   Q   G   D   S   L   S   S   S   M   L   S   T   T
4715 CCC AAC CAT GAT CCT ATG GCT GGT GGG GAC AAA GAT GAT GGT TTT GAC CGG CGG GTG
     GGG TTG GTA CTA GGA TAC CGA CCA CCC CTG TTT CTA CTA CCA AAA CTG GCC GCC CAC

 423▶ P   N   H   D   P   M   A   G   G   D   K   D   D   G   F   D   R   R   V
4772 GCT TTC CTG GAA CCT GCT GGA CCT GGG CAG GGT CAA AGC GGC AAG CTC AAG CCA GGC
     CGA AAG GAC CTT GGA CGA CCT GGA CCC GTC CCA GTT TCG CCG TTC GAG TTC GGT CCG

 442▶ A   F   L   E   P   A   G   P   G   Q   G   Q   S   G   K   L   K   P   G
4829 AAC ACC CTG CCA GTG CCT GAG GCC ACA GGC AAG GGC CAA CGG AGG ATG TCG GAT CCA
     TTG TGG GAC GGT CAC GGA CTC CGG TGT CCG TTC CCG GTT GCC TCC TAC AGC CTA GGT

 461▶ N   T   L   P   V   P   E   A   T   G   K   G   Q   R   R   M   S   D   P
```

```
4886 GAA TCA GTC AGC TCG GTC TTC TCT GAT ACG CCC ATT GTG GTG CCA ACT TTC TTG TAC
     CTT AGT CAG TCG AGC CAG AAG AGA CTA TGC GGG TAA CAC CAC GGT TGA AAG AAC ATG

 480▸ E   S   V   S   S   V   F   S   D   T   P   I   V   V   P   T   F   L   Y
-------------------------------------------------------------------------------
                                    V5  epitope
                 attB2  (hybrid)

4943 AAA GTG GTG GTA CCG  GGT AAG CCT ATC CCT AAC CCT CTC CTC GGT CTC GAT TCT ACG

     TTT CAC CAC CAT GGC  CCA TTC GGA TAG GGA TTG GGA GAG GAG CCA GAG CTA AGA TGC

 499▸ K   V   V   V   P    G   K   P   I   P   N   P   L   L   G   L   D   S   T
-------------------------------------------------------------------------------
                6xHis                        SV40  Poly  A
5000 CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA TCTAGAGGATCTTTGTGAAGGAACCTTACTTCTGTGG
     GCA TGG CCA GTA GTA GTG GTA GTG GTA ACT AGATCTCCTAGAAACACTTCCTTGGAATGAAGACACC

 518▸ R   T   G   H   H   H   H   H   H   •
-------------------------------------------------------------------------------
5067 TGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTG
     ACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCACATATTACAC

-------------------------------------------------------------------------------
5144 TTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGGAAT
     AATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCGTCACCACCTTA

-------------------------------------------------------------------------------
5221 GCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACA
     CGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGT

-------------------------------------------------------------------------------
5298 TTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTC
     AAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAG

-------------------------------------------------------------------------------
5375 ATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAG
     TACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTC

-------------------------------------------------------------------------------
5452 AAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTAC
     TTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAAGAATG

-------------------------------------------------------------------------------
5529 TCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAG
     AGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTC

-------------------------------------------------------------------------------
5606 GGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACCACATTTGTAGAGGTTTT
     CCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAA

-------------------------------------------------------------------------------
5683 ACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGT
     TGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACA

-------------------------------------------------------------------------------
5760 TTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCAT
     AATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTA

-------------------------------------------------------------------------------
                                                                 white gene
5837 TCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGGATCCACTAGAAGGCCTTAGTATGT
     AGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCCTAGGTGATCTTCCGGAATCATACA
```

```
5914 ATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATTTTTTTTTTATTTTTTACTGCACTG
     TACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAAAAAAAAAATAAAAAATGACGTGAC

5991 GACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTATTTGAAGTACCAGGTTCTTTCGATT
     CTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATAAACTTCATGGTCCAAGAAAGCTAA

6068 ACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTC
     TGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAG

6145 TCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGC
     AGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACG

6222 ACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTAAACACAAAGTGCTGTGCCAAAACTC
     TGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAG

6299 CTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGT
     GAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACCAAAACCCACCCATTCGTCCCCTTTCA

6376 GTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGC
     CACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCG

6453 ATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGATCTATAGGAACTGCGATTGCAACATCAA
     TAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACTAGATATCCTTGACGCTAACGTTGTAGTT

6530 ATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAATCGCACAAACAAATAGTGACACGAAACAG
     TAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTC

6607 ATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCATATCATGATCAAGACATCTAAAGGCATTCA
     TAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGT

6684 TTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAGATATTTTAAGCTGATCCTAGATGCACAAAAAATAAA
     AAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAAAATTCGACTAGGATCTACGTGTTTTTTATTT

6761 TAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATATTT
     ATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCCAATCTACTCGTATTGCGAACATCAACTATAAA

6838 GAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAA
     CTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTT

6915 AACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCAC
     TTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTG

6992 CTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGA
     GATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCT

7069 CACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTC
     GTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGGTCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAG

7146 GCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTAT
     CGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTTAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATA
```

```
7223 CCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATC
     GGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAG

7300 GCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGG
     CGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCC

7377 CCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCC
     GGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGG

7454 ATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCT
     TACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGA

7531 CAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGG
     GTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACC

7608 CATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACC
     GTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGG

7685 GCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTC
     CGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAG

7762 TTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCG
     AAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGC

7839 AAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCT
     TTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGA

7916 ATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCG
     TAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGC

7993 GACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATG
     CTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATAC

8070 GAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTG
     CTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGAC

8147 GTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGAC
     CAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTG

8224 TTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAG
     AATAAGTCTGTTGCCACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATC

8301 ATTTTAGTTAGAATTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAA
     TAAAATCAATCTTAACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTT

8378 AATAAATAAATTTCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTG
     TTATTTATTTAAAGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAAC

8455 CAGATGGTTGCCATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGG
     GTCTACCAACGGTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCC
```

```
8532 AGCCATCTTCCTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATA
     TCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTAT

8609 CATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTT
     GTAAACGTATAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAA

8686 TATGAGGGAGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTT
     ATACTCCCTCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAA

8763 TTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTC
     AAGAGTGTCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAG

8840 AACTGCCTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTC
     TTGACGGACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAG

8917 GACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGG
     CTGGAGCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCC

8994 GCTCGGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAAC
     CGAGCCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTG

9071 CAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCAT
     GTTACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTA

9148 CCTGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCC
     GGACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGG

9225 GGGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTG
     CCCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAAC

9302 TTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGC
     AAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACACG

9379 CAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTTGT
     GTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACA

9456 CTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATAT
     GAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATA

9533 ATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGTTGCT
     TAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGA

9610 CTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTC
     GAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAG

9687 CGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGTATG
     GCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACATAC

9764 TACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACTTT
     ATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGAAA
```

```
 9841  TTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTC
       AATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACAAG

 9918  CTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAACTATAATAA
       GAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTTGATATTATT

 9995  TAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGC
       ATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCG

10072  ATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAG
       TAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTC

10149  TCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCT
       AGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGA

10226  TCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATGT
       AGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTACA

10303  CTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAGGC
       GACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTCCG

10380  TGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTCACA
       ACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAGTGT

10457  GCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTC
       CGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAAGCAG

10534  CGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGAC
       GCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGATAGCTG

                                 5' P
10611  GGGACCACCTTATGTTATTTCATCATG
       CCCTGGTGGAATACAATAAAGTAGTAC
```

pUAST-AMIGO2-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGAC

  77 GGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCT
     CCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGA

 153 CGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAA
     GCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTT

 229 GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCG
     CCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGC

 305 ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTA
     TGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAAT

 381 CCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTC
     GGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAG

 457 GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
     CCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCA

 533 AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
     TTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGT

 609 GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
     CTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAA

 685 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
     ACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG

 761 GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
     CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACT

 837 TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
     AGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTC

 913 GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
     CTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGA

 989 GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
     CTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTG

1065 TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
     AGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGG

1141 ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
     TGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGA

1217 TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA
     AATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGT

1293 ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
     TGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGT

1369 ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
     TGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAG

1445 AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
     TCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGC

1521 TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
     ATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAG

1597 TTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
     AACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGA

1673 TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
     AGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTA

1749 CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
     GAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTA
```

```
1825 AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
     TTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACA

1901 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
     GAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTC

1977 TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
     ACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGC

2053 TCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA
     AGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATT

2129 GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG
     CGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATAC

2205 CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACG
     GCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGC

2281 ACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTA
     TGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGAT

2357 CGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCC
     GCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGG

2433 GGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAAT
     CCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTA

2509 CCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACA
     GGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGT

2585 AATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAAACAAG
     TTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTTGTTC

2661 CAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGA
     GTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCT

                            3'P
2737 AGCTCACGATGAGAATGGCCAGACCATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACG
     TCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGC

2813 TATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGA
     ATACGAACGTTATTCACGCTCACTTTCCTTATCATAAGACTCACAGCATAACTCAGACTCACTCTGTCGCTATACT

2889 TTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAG
     AACAACTAATTGGGAATCGTACAGGCACCCCAAACTTAATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTC

2965 TTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGG
     AAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCC

3041 AGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGG
     TCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCC

                               UAS sites
3117 AGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATT
     TCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAA

3193 CAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAA
     GTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTT

3269 CAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACT
     GTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGA
```

```
                                                                Polylinker
3345 GAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACAGATC
     CTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGTCTAG

       attB1  (hybrid)
3421 TGACAAGTTTGTACAAAAAAGTTGGAAA ATG TCG TTA CGT GTA CAC ACT CTG CCC ACC CTG CTT
     ACTGTTCAAACATGTTTTTTCAACCTTT TAC AGC AAT GCA CAT GTG TGA GAC GGG TGG GAC GAA

                                1▸ M   S   L   R   V   H   T   L   P   T   L   L
3485 GGA GCC GTC GTC AGA CCG GGC TGC AGG GAG CTG CTG TGT TTG CTG ATG ATC ACA GTG
     CCT CGG CAG CAG TCT GGC CCG ACG TCC CTC GAC GAC ACA AAC GAC TAC TAG TGT CAC

  13▸ G   A   V   V   R   P   G   C   R   E   L   L   C   L   L   M   I   T   V
3542 ACT GTG GGC CCT GGT GCC TCT GGG GTG TGC CCC ACC GCT TGC ATC TGT GCC ACT GAC
     TGA CAC CCG GGA CCA CGG AGA CCC CAC ACG GGG TGG CGA ACG TAG ACA CGG TGA CTG

  32▸ T   V   G   P   G   A   S   G   V   C   P   T   A   C   I   C   A   T   D
3599 ATC GTC AGC TGC ACC AAC AAA AAC CTG TCC AAG GTG CCT GGG AAC CTT TTC AGA CTG
     TAG CAG TCG ACG TGG TTG TTT TTG GAC AGG TTC CAC GGA CCC TTG GAA AAG TCT GAC

  51▸ I   V   S   C   T   N   K   N   L   S   K   V   P   G   N   L   F   R   L
3656 ATT AAG AGA CTG GAC CTG AGT TAT AAC AGA ATT GGG CTT CTG GAT TCT GAG TGG ATT
     TAA TTC TCT GAC CTG GAC TCA ATA TTG TCT TAA CCC GAA GAC CTA AGA CTC ACC TAA

  70▸ I   K   R   L   D   L   S   Y   N   R   I   G   L   L   D   S   E   W   I
3713 CCA GTA TCG TTT GCA AAG CTG AAC ACC CTA ATT CTT CGT CAT AAC AAC ATC ACC AGC
     GGT CAT AGC AAA CGT TTC GAC TTG TGG GAT TAA GAA GCA GTA TTG TTG TAG TGG TCG

  89▸ P   V   S   F   A   K   L   N   T   L   I   L   R   H   N   N   I   T   S
3770 ATT TCC ACG GGC AGT TTT TCC ACA ACT CCA AAT TTG AAG TGT CTT GAC TTA TCG TCC
     TAA AGG TGC CCG TCA AAA AGG TGT TGA GGT TTA AAC TTC ACA GAA CTG AAT AGC AGG

 108▸ I   S   T   G   S   F   S   T   T   P   N   L   K   C   L   D   L   S   S
3827 AAT AAG CTG AAG ACG GTG AAA AAT GCT GTA TTC CAA GAG TTG AAG GTT CTG GAA GTG
     TTA TTC GAC TTC TGC CAC TTT TTA CGA CAT AAG GTT CTC AAC TTC CAA GAC CTT CAC

 127▸ N   K   L   K   T   V   K   N   A   V   F   Q   E   L   K   V   L   E   V
3884 CTT CTG CTT TAC AAC AAT CAC ATA TCC TAT CTC GAT CCT TCA GCG TTT GGA GGG CTC
     GAA GAC GAA ATG TTG TTA GTG TAT AGG ATA GAG CTA GGA AGT CGC AAA CCT CCC GAG

 146▸ L   L   L   Y   N   N   H   I   S   Y   L   D   P   S   A   F   G   G   L
3941 TCC CAG TTG CAG AAA CTC TAC TTA AGT GGA AAT TTT CTC ACA CAG TTT CCG ATG GAT
     AGG GTC AAC GTC TTT GAG ATG AAT TCA CCT TTA AAA GAG TGT GTC AAA GGC TAC CTA

 165▸ S   Q   L   Q   K   L   Y   L   S   G   N   F   L   T   Q   F   P   M   D
3998 TTG TAT GTT GGA AGG TTC AAG CTG GCA GAA CTG ATG TTT TTA GAT GTT TCT TAT AAC
     AAC ATA CAA CCT TCC AAG TTC GAC CGT CTT GAC TAC AAA AAT CTA CAA AGA ATA TTG

 184▸ L   Y   V   G   R   F   K   L   A   E   L   M   F   L   D   V   S   Y   N
4055 CGA ATT CCT TCC ATG CCA ATG CAC CAC ATA AAT TTA GTG CCA GGA AAA CAG CTG AGA
     GCT TAA GGA AGG TAC GGT TAC GTG GTG TAT TTA AAT CAC GGT CCT TTT GTC GAC TCT

 203▸ R   I   P   S   M   P   M   H   H   I   N   L   V   P   G   K   Q   L   R
```

```
4112 GGC ATC TAC CTT CAT GGA AAC CCA TTT GTC TGT GAC TGT TCC CTG TAC TCC TTG CTG
     CCG TAG ATG GAA GTA CCT TTG GGT AAA CAG ACA CTG ACA AGG GAC ATG AGG AAC GAC

 222▶ G   I   Y   L   H   G   N   P   F   V   C   D   C   S   L   Y   S   L   L

4169 GTC TTT TGG TAT CGT AGG CAC TTT AGC TCA GTG ATG GAT TTT AAG AAC GAT TAC ACC
     CAG AAA ACC ATA GCA TCC GTG AAA TCG AGT CAC TAC CTA AAA TTC TTG CTA ATG TGG

 241▶ V   F   W   Y   R   R   H   F   S   S   V   M   D   F   K   N   D   Y   T

4226 TGT CGC CTG TGG TCT GAC TCC AGG CAC TCG CGT CAG GTA CTT CTG CTC CAG GAT AGC
     ACA GCG GAC ACC AGA CTG AGG TCC GTG AGC GCA GTC CAT GAA GAC GAG GTC CTA TCG

 260▶ C   R   L   W   S   D   S   R   H   S   R   Q   V   L   L   L   Q   D   S

4283 TTT ATG AAT TGC TCT GAC AGC ATC ATC AAT GGT TCC TTT CGT GCG CTT GGC TTT ATT
     AAA TAC TTA ACG AGA CTG TCG TAG TAG TTA CCA AGG AAA GCA CGC GAA CCG AAA TAA

 279▶ F   M   N   C   S   D   S   I   I   N   G   S   F   R   A   L   G   F   I

4340 CAT GAG GCT CAG GTC GGG GAA AGA CTG ATG GTC CAC TGT GAC AGC AAG ACA GGT AAT
     GTA CTC CGA GTC CAG CCC CTT TCT GAC TAC CAG GTG ACA CTG TCG TTC TGT CCA TTA

 298▶ H   E   A   Q   V   G   E   R   L   M   V   H   C   D   S   K   T   G   N

4397 GCA AAT ACG GAT TTC ATC TGG GTG GGT CCA GAT AAC AGA CTG CTA GAG CCG GAT AAA
     CGT TTA TGC CTA AAG TAG ACC CAC CCA GGT CTA TTG TCT GAC GAT CTC GGC CTA TTT

 317▶ A   N   T   D   F   I   W   V   G   P   D   N   R   L   L   E   P   D   K

4454 GAG ATG GAA AAC TTT TAC GTG TTT CAC AAT GGA AGT CTG GTT ATA GAA AGC CCT CGT
     CTC TAC CTT TTG AAA ATG CAC AAA GTG TTA CCT TCA GAC CAA TAT CTT TCG GGA GCA

 336▶ E   M   E   N   F   Y   V   F   H   N   G   S   L   V   I   E   S   P   R

4511 TTT GAG GAT GCT GGA GTG TAT TCT TGT ATC GCA ATG AAT AAG CAA CGC CTG TTA AAT
     AAA CTC CTA CGA CCT CAC ATA AGA ACA TAG CGT TAC TTA TTC GTT GCG GAC AAT TTA

 355▶ F   E   D   A   G   V   Y   S   C   I   A   M   N   K   Q   R   L   L   N

4568 GAA ACT GTG GAC GTC ACA ATA AAT GTG AGC AAT TTC ACT GTA AGC AGA TCC CAT GCT
     CTT TGA CAC CTG CAG TGT TAT TTA CAC TCG TTA AAG TGA CAT TCG TCT AGG GTA CGA

 374▶ E   T   V   D   V   T   I   N   V   S   N   F   T   V   S   R   S   H   A

                          TM
4625 CAT GAG GCA TTT AAC ACA GCT TTT ACC ACT CTT GCT GCT TGC GTG GCC AGT ATC GTT
     GTA CTC CGT AAA TTG TGT CGA AAA TGG TGA GAA CGA CGA ACG CAC CGG TCA TAG CAA

 393▶ H   E   A   F   N   T   A   F   T   T   L   A   A   C   V   A   S   I   V

4682 TTG GTA CTT TTG TAC CTC TAT CTG ACT CCA TGC CCC TGC AAG TGT AAA ACC AAG AGA
     AAC CAT GAA AAC ATG GAG ATA GAC TGA GGT ACG GGG ACG TTC ACA TTT TGG TTC TCT

 412▶ L   V   L   L   Y   L   Y   L   T   P   C   P   C   K   C   K   T   K   R

4739 CAG AAA AAT ATG CTA CAC CAA AGC AAT GCC CAT TCA TCG ATT CTC AGT CCT GGC CCC
     GTC TTT TTA TAC GAT GTG GTT TCG TTA CGG GTA AGT AGC TAA GAG TCA GGA CCG GGG

 431▶ Q   K   N   M   L   H   Q   S   N   A   H   S   S   I   L   S   P   G   P

4796 GCT AGT GAT GCC TCC GCT GAT GAA CGG AAG GCA GGT GCA GGT AAA AGA GTG GTG TTT
     CGA TCA CTA CGG AGG CGA CTA CTT GCC TTC CGT CCA CGT CCA TTT TCT CAC CAC AAA

 450▶ A   S   D   A   S   A   D   E   R   K   A   G   A   G   K   R   V   V   F
```

```
4853 TTG GAA CCC CTG AAG GAT ACT GCA GCA GGG CAG AAC GGG AAA GTC AGG CTC TTT CCC
     AAC CTT GGG GAC TTC CTA TGA CGT CGT CCC GTC TTG CCC TTT CAG TCC GAG AAA GGG

 469▸ L   E   P   L   K   D   T   A   A   G   Q   N   G   K   V   R   L   F   P

4910 AGC GAG GCA GTG ATA GCT GAG GGC ATC CTA AAG TCC ACG AGG GGG AAA TCT GAC TCA
     TCG CTC CGT CAC TAT CGA CTC CCG TAG GAT TTC AGG TGC TCC CCC TTT AGA CTG AGT

 488▸ S   E   A   V   I   A   E   G   I   L   K   S   T   R   G   K   S   D   S

                                                    (was G)
4967 GAT TCA GTC AAT TCA GTG TTT TCT GAC ACA CCT TTT GTC GCG TCC ACT CCA ACT TTC
     CTA AGT CAG TTA AGT CAC AAA AGA CTG TGT GGA AAA CAG CGC AGG TGA GGT TGA AAG

 507▸ D   S   V   N   S   V   F   S   D   T   P   F   V   A   S   T   P   T   F

                                  V5  epitope
                       attB2  (hybrid)
5024 TTG TAC AAA GTG GTG GTA CCG  GGT AAG CCT ATC CCT AAC CCT CTC CTC GGT CTC GAT
     AAC ATG TTT CAC CAC CAT GGC  CCA TTC GGA TAG GGA TTG GGA GAG GAG CCA GAG CTA
                                  ───────────────────────────────►
 526▸ L   Y   K   V   V   V   P    G   K   P   I   P   N   P   L   L   G   L   D

                         6xHis                            SV40  Poly  A
5081 TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA TCTAGAGGATCTTTGTGAAGGAACCTTA
     AGA TGC GCA TGG CCA GTA GTA GTG GTA GTG GTA ACT AGATCTCCTAGAAACACTTCCTTGGAAT
                     ◄──────────────────────────────
 545▸ S   T   R   T   G   H   H   H   H   H   H   •

5145 CTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGT
     GAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCA

5221 GTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGC
     CATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCG

5297 AGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGC
     TCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACG

5373 TGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTA
     ACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGAT

5449 AGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTG
     TCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGAC

5525 CACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACAT
     GTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTA

5601 ACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGC
     TGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACATGGAAATCG

5677 TTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACC
     AAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGG
```

```
5753 ACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAA
     TGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTT

5829 TTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAA
     AACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATT

                                                                            white
5905 AGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGGATCC
     TCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCCTAGG

5981 ACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATTTTT
     TGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAAAAA

6057 TTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTATTT
     AAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATAAA

6133 GAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTCT
     CTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAGA

6209 CTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTA
     GACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGAT

6285 GCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTA
     CGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAAT

6361 AACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGG
     TTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACC

6437 TTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCG
     AAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGC

6513 GAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGA
     CTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACT

6589 TCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAA
     AGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTT

6665 TCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCA
     AGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGT

6741 TATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAGATATT
     ATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAA

6817 TTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGG
     AATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCC

6893 TTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAG
     AATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTC

6969 AGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCG
     TCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGC
```

```
7045 GTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCA
     CAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGT

7121 GCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCC
     CGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGG

7197 AGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCA
     TCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGT

7273 ATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCC
     TAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGG

7349 GGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGAC
     CCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTG

7425 TGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTAT
     ACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATA

7501 CGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAG
     GCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTC

7577 CGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCG
     GCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGC

7653 GCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCT
     CGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGA

7729 TCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTG
     AGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGAC

7805 TCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCC
     AGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGG

7881 TTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCG
     AAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGC

7957 ATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAA
     TACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTT

8033 CTACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATC
     GATGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAG

8109 GCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGG
     CGGTTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACC

8185 AGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTG
     TCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACAC

8261 GCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCC
     CGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGG
```

```
8337 AGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTAGTTAGAATTGCCTGATTCC
     TCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGG

8413 ACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCG
     TGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGC

8489 CGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGG
     GCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACC

8565 CCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACC
     GGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGG

8641 AACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTA
     TTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAAT

8717 CTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTC
     GATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAG

8793 GACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGT
     CTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCA

8869 CTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTC
     GAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAG

8945 ACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGT
     TGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACA

9021 CTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATA
     GACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATAT

9097 CCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTG
     GGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCAC

9173 GAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTA
     CTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAAT

9249 ACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATA
     TGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTAT

9325 TCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTGTTTTTTTTTTTA
     AGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAACAAAAAAAAAAAT

9401 CCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAATAAAAACA
     GGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACACGGTTATTTTTGT

9477 AGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTTGTCTTCTAACCA
     TCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACAGAAGATTGGT

9553 AAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTAT
     TTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATATAAAAAATA
```

```
 9629 ATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGTTGCTCTTTCGCT
      TATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGCGA
-------------------------------------------------------------------------------------
 9705 GTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCGCGA
      CAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGCGCT
-------------------------------------------------------------------------------------
 9781 TTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGTATGTACAAT
      AAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACATACATGTTA
-------------------------------------------------------------------------------------
 9857 TTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACTTTTTAAG
      AATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGAAAAATTC
-------------------------------------------------------------------------------------
 9933 GTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTCCTTG
      CATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACAAGGAAC
-------------------------------------------------------------------------------------
10009 TGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAACTATAATAATAA
      ACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTTGATATTATTATT
-------------------------------------------------------------------------------------
10085 GAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCAT
      CTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGTA
-------------------------------------------------------------------------------------
10161 TTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAGT
      AAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCA
-------------------------------------------------------------------------------------
10237 CACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCT
      GTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGA
-------------------------------------------------------------------------------------
10313 TCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATG
      AGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTAC
-------------------------------------------------------------------------------------
10389 TCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAG
      AGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTC
-------------------------------------------------------------------------------------
10465 GCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTC
      CGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAG
-------------------------------------------------------------------------------------
10541 ACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATT
      TGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAA
-------------------------------------------------------------------------------------
10617 CGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTA
      GCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGAT
-------------------------------------------------------------------------------------
                                     5' P
10693 TCGACGGGACCACCTTATGTTATTTCATCATG
      AGCTGCCCTGGTGGAATACAATAAAGTAGTAC
-------------------------------------------------------------------------------------
```

pUAST-AMIGO3-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTGG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGACC

  78 GGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCTCG
     CCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGAGC

 155 TTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAAGGC
     AACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTTCCG

 232 GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
     CAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTC

 309 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
     CTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCT

 386 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
     ATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACAT

 463 GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
     CCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAG

 540 GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
     CAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCC

 617 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
     ATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGAC

 694 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
     GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGC

 771 GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
     CACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGC

 848 GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
     CCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGAT

 925 GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
     CTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTA

1002 GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
     CGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATC

1079 ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
     TATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

1156 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
     TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGG

1233 AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
     TCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGA

1310 ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
     TGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATG

1387 ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
     TACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTC

1464 TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
     ACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGA

1541 GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGA
     CCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCT

1618 TAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
     ATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCT

1695 TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
     AGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGG

1772 AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
     TCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTA
```

```
1849  ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
      TGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTA

1926  GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
      CATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAA

2003  ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
      TAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACT

2080  AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
      TTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAG

2157  AGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
      TCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCAC

2234  CACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGG
      GTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCC

2311  TGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGA
      ACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACT

2388  AGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGA
      TCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACT

2465  TGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGA
      ACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCT

2542  AATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATA
      TTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTAT

2619  AACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCG
      TTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGC

                                                                             3'P
2696  TCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATA
      AGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTAT

2773  ACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATT
      TGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAA

2850  CTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAAT
      GACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTTA

2927  TAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGC
      ATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACG

3004  ATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG
      TACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTC

                                                                           UAS sites
3081  TACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGT
      ATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCA

3158  ATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCT
      TATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGA

3235  AAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGT
      TTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCA

3312  AACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCT
      TTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGA
```

```
                        Polylinker           attB1   (hybrid)                 (was C)
3389 GAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGTACAAAAAAGTTGGAAA ATG ACG TGG TTG
     CTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAAACATGTTTTTTCAACCTTT TAC TGC ACC AAC

                                                               1▸  M   T   W   L
3461 GTG CTG CTG GGG ACA CTG CTC TGC ATG CTG CGC GTT GGG TTA GGC ACC CCG GAC TCC
     CAC GAC GAC CCC TGT GAC GAG ACG TAC GAC GCG CAA CCC AAT CCG TGG GGC CTG AGG

  5▸  V   L   L   G   T   L   L   C   M   L   R   V   G   L   G   T   P   D   S
3518 GAG GGT TTC CCG CCC CGT GCG CTC CAC AAC TGC CCC TAC AAA TGT ATC TGC GCT GCC
     CTC CCA AAG GGC GGG GCA CGC GAG GTG TTG ACG GGG ATG TTT ACA TAG ACG CGA CGG

 24▸  E   G   F   P   P   R   A   L   H   N   C   P   Y   K   C   I   C   A   A
                                         PstI
3575 GAC CTG CTA AGC TGC ACT GGC CTA GGG CTG CAG GAC GTG CCA GCC GAG TTA CCT GCC
     CTG GAC GAT TCG ACG TGA CCG GAT CCC GAC GTC CTG CAC GGT CGG CTC AAT GGA CGG

 43▸  D   L   L   S   C   T   G   L   G   L   Q   D   V   P   A   E   L   P   A
3632 GCT ACT GCG GAC CTC GAC CTG AGC CAC AAC GCG CTC CAG CGC CTG CGC CCC GGC TGG
     CGA TGA CGC CTG GAG CTG GAC TCG GTG TTG CGC GAG GTC GCG GAC GCG GGG CCG ACC

 62▸  A   T   A   D   L   D   L   S   H   N   A   L   Q   R   L   R   P   G   W
3689 TTG GCG CCC CTC TTC CAG CTG CGC GCC CTG CAC CTA GAC CAC AAC GAA CTA GAT GCG
     AAC CGC GGG GAG AAG GTC GAC GCG CGG GAC GTG GAT CTG GTG TTG CTT GAT CTA CGC

 81▸  L   A   P   L   F   Q   L   R   A   L   H   L   D   H   N   E   L   D   A
3746 CTG GGT CGC GGC GTC TTC GTC AAC GCC AGC GGC CTG AGG CTG CTC GAT CTA TCA TCT
     GAC CCA GCG CCG CAG AAG CAG TTG CGG TCG CCG GAC TCC GAC GAG CTA GAT AGT AGA

100▸  L   G   R   G   V   F   V   N   A   S   G   L   R   L   L   D   L   S   S
3803 AAC ACG TTG CGG GCG CTT GGC CGC CAC GAC CTC GAC GGG CTG GGG GCG CTG GAG AAG
     TTG TGC AAC GCC CGC GAA CCG GCG GTG CTG GAG CTG CCC GAC CCC CGC GAC CTC TTC

119▸  N   T   L   R   A   L   G   R   H   D   L   D   G   L   G   A   L   E   K
3860 CTG CTT CTG TTC AAT AAC CGC TTG GTG CAC TTG GAC GAG CAT GCC TTC CAC GGC CTG
     GAC GAA GAC AAG TTA TTG GCG AAC CAC GTG AAC CTG CTC GTA CGG AAG GTG CCG GAC

138▸  L   L   L   F   N   N   R   L   V   H   L   D   E   H   A   F   H   G   L
3917 CGC GCG CTC AGC CAT CTC TAC CTG GGC TGC AAC GAA CTC GCC TCG TTC TCC TTC GAC
     GCG CGC GAG TCG GTA GAG ATG GAC CCG ACG TTG CTT GAG CGG AGC AAG AGG AAG CTG

157▸  R   A   L   S   H   L   Y   L   G   C   N   E   L   A   S   F   S   F   D
3974 CAC CTG CAC GGT CTG AGC GCC ACC CAC CTG CTT ACT CTG GAC CTC TCC TCC AAC CGG
     GTG GAC GTG CCA GAC TCG CGG TGG GTG GAC GAA TGA GAC CTG GAG AGG AGG TTG GCC

176▸  H   L   H   G   L   S   A   T   H   L   L   T   L   D   L   S   S   N   R
4031 CTG GGA CAC ATC TCC GTA CCT GAG CTG GCC GCG CTG CCG GCC TTC CTC AAG AAC GGC
     GAC CCT GTG TAG AGG CAT GGA CTC GAC CGG CGC GAC GGC CGG AAG GAG TTC TTG CCG

195▸  L   G   H   I   S   V   P   E   L   A   A   L   P   A   F   L   K   N   G
```

```
4088 CTC TAC TTG CAC AAC AAC CCT TTG CCT TGC GAC TGC CGC CTC TAC CAC CTG CTA CAG
     GAG ATG AAC GTG TTG TTG GGA AAC GGA ACG CTG ACG GCG GAG ATG GTG GAC GAT GTC

 214▸ L   Y   L   H   N   N   P   L   P   C   D   C   R   L   Y   H   L   L   Q

4145 CGC TGG CAC CAG CGG GGC CTG AGC GCC GTG CGC GAC TTT GCG CGC GAG TAC GTA TGC
     GCG ACC GTG GTC GCC CCG GAC TCG CGG CAC GCG CTG AAA CGC GCG CTC ATG CAT ACG

 233▸ R   W   H   Q   R   G   L   S   A   V   R   D   F   A   R   E   Y   V   C

4202 TTG GCC TTC AAG GTA CCC GCG TCC CGC GTG CGC TTC TTC CAG CAC AGC CGC GTC TTT
     AAC CGG AAG TTC CAT GGG CGC AGG GCG CAC GCG AAG AAG GTC GTG TCG GCG CAG AAA

 252▸ L   A   F   K   V   P   A   S   R   V   R   F   F   Q   H   S   R   V   F

4259 GAG AAC TGC TCG TCG GCC CCA GCT CTT GGC CTA GAG CGG CCG GAA GAG CAC CTG TAC
     CTC TTG ACG AGC AGC CGG GGT CGA GAA CCG GAT CTC GCC GGC CTT CTC GTG GAC ATG

 271▸ E   N   C   S   S   A   P   A   L   G   L   E   R   P   E   E   H   L   Y

4316 GCG CTG GTG GGT CGG TCC CTG AGG CTT TAC TGC AAC ACC AGC GTC CCG GCC ATG CGC
     CGC GAC CAC CCA GCC AGG GAC TCC GAA ATG ACG TTG TGG TCG CAG GGC CGG TAC GCG

 290▸ A   L   V   G   R   S   L   R   L   Y   C   N   T   S   V   P   A   M   R

4373 ATT GCC TGG GTT TCG CCG CAG CAG GAG CTT CTC AGG GCG CCA GGA TCC CGC GAT GGC
     TAA CGG ACC CAA AGC GGC GTC GTC CTC GAA GAG TCC CGC GGT CCT AGG GCG CTA CCG

 309▸ I   A   W   V   S   P   Q   Q   E   L   L   R   A   P   G   S   R   D   G

4430 AGC ATC GCG GTG CTG GCC GAC GGC AGC TTG GCC ATA GGC AAC GTA CAG GAG CAG CAT
     TCG TAG CGC CAC GAC CGG CTG CCG TCG AAC CGG TAT CCG TTG CAT GTC CTC GTC GTA

 328▸ S   I   A   V   L   A   D   G   S   L   A   I   G   N   V   Q   E   Q   H

4487 GCG GGA CTC TTC GTG TGC CTG GCC ACT GGG CCC CGC CTG CAC CAC AAC CAG ACG CAC
     CGC CCT GAG AAG CAC ACG GAC CGG TGA CCC GGG GCG GAC GTG GTG TTG GTC TGC GTG

 347▸ A   G   L   F   V   C   L   A   T   G   P   R   L   H   H   N   Q   T   H

                                                                     TM
4544 GAG TAC AAC GTG AGC GTG CAC TTT CCG CGC CCA GAG CCC GAG GCT TTC AAC ACA GGC
     CTC ATG TTG CAC TCG CAC GTG AAA GGC GCG GGT CTC GGG CTC CGA AAG TTG TGT CCG

 366▸ E   Y   N   V   S   V   H   F   P   R   P   E   P   E   A   F   N   T   G

4601 TTC ACC ACA CTG CTG GGC TGT GCC GTG GGC CTT GTG CTC GTG CTG CTC TAC CTG TTC
     AAG TGG TGT GAC GAC CCG ACA CGG CAC CCG GAA CAC GAG CAC GAC GAG ATG GAC AAG

 385▸ F   T   T   L   L   G   C   A   V   G   L   V   L   V   L   L   Y   L   F

4658 GCC CCA CCC TGC CGC TGC TGC CGC CGT GCC TGC CGC TGC CGC CGC TGG CCC CAA ACA
     CGG GGT GGG ACG GCG ACG ACG GCG GCA CGG ACG GCG ACG GCG GCG ACC GGG GTT TGT

 404▸ A   P   P   C   R   C   C   R   R   A   C   R   C   R   R   W   P   Q   T

4715 CCC AGC CCG CTC CAA GAG CTG AGC GCA CAG TCC TCA GTA CTC AGC ACC ACA CCG CCA
     GGG TCG GGC GAG GTT CTC GAC TCG CGT GTC AGG AGT CAT GAG TCG TGG TGT GGC GGT

 423▸ P   S   P   L   Q   E   L   S   A   Q   S   S   V   L   S   T   T   P   P

4772 GAC GCA CCC AGC CGC AAG GCC AGC GTC CAC AAG CAC GTA GTC TTT CTG GAG CCA GGC
     CTG CGT GGG TCG GCG TTC CGG TCG CAG GTG TTC GTG CAT CAG AAA GAC CTC GGT CCG

 442▸ D   A   P   S   R   K   A   S   V   H   K   H   V   V   F   L   E   P   G
```

```
4829 CGG AGG GGC CTC AAT GGC CGC GTG CAG CTG GCA GTA GCT GAG GAA TTC GAT CTC TAC
     GCC TCC CCG GAG TTA CCG GCG CAC GTC GAC CGT CAT CGA CTC CTT AAG CTA GAG ATG

 461▸ R   R   G   L   N   G   R   V   Q   L   A   V   A   E   E   F   D   L   Y
------------------------------------------------------------------------------------
4886 AAC CCT GGA GGC CTG CAG CTG AAG GCT GGC TCT GAG TCC GCC AGC TCC ATA GGC TCC
     TTG GGA CCT CCG GAC GTC GAC TTC CGA CCG AGA CTC AGG CGG TCG AGG TAT CCG AGG

 480▸ N   P   G   G   L   Q   L   K   A   G   S   E   S   A   S   S   I   G   S
------------------------------------------------------------------------------------
                                                                  V5  epitope
                                                      attB2  (hybrid)
4943 GAG GGT CCC ATG ACA ACC CCA ACT TTC TTG TAC AAA GTG GTG GTA CCG  GGT AAG CCT

     CTC CCA GGG TAC TGT TGG GGT TGA AAG AAC ATG TTT CAC CAC CAT GGC  CCA TTC GGA

 499▸ E   G   P   M   T   T   P   T   F   L   Y   K   V   V   V   P    G   K   P
------------------------------------------------------------------------------------
                                                       6xHis
5000 ATC CCT AAC CCT CTC CTC GGT CTC GAT TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC
     TAG GGA TTG GGA GAG GAG CCA GAG CTA AGA TGC GCA TGG CCA GTA GTA GTG GTA GTG

 518▸ I   P   N   P   L   L   G   L   D   S   T   R   T   G   H   H   H   H   H
------------------------------------------------------------------------------------
                SV40  Poly  A
5057 CAT TGA TCTAGAGGATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGA
     GTA ACT AGATCTCCTAGAAACACTTCCTTGGAATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCT

 537▸ H   •
------------------------------------------------------------------------------------
5132 TTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTT
     AAATTTCGAGATTCCATTTATATTTTAAAAATTCACATATTACACAATTTGATGACTAAGATTAACAAACACATAAA

------------------------------------------------------------------------------------
5209 TAGATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGA
     ATCTAAGGTTGGATACCTTGACTACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCT

------------------------------------------------------------------------------------
5286 AGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAG
     TCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATC

------------------------------------------------------------------------------------
5363 AAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGC
     TTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACG

------------------------------------------------------------------------------------
5440 TTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTAT
     AAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATA

------------------------------------------------------------------------------------
5517 AAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATA
     TTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTAT

------------------------------------------------------------------------------------
5594 ACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCC
     TGATACGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGG

------------------------------------------------------------------------------------
5671 TTGACTAGAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCC
     AACTGATCTCTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGG

------------------------------------------------------------------------------------
5748 TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGC
     ACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCG

------------------------------------------------------------------------------------
```

```
5825  AATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGT
      TTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACA

                              white gene
5902  ATCTTATCATGTCTGGATCGGATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAAT
      TAGAATAGTACAGACCTAGCCTAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTA

5979  GTAGATTTAAAAAAACATATTTTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAA
      CATCTAAATTTTTTTGTATAAAAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTT

6056  TTTACTTCGATCCAAGGGTATTTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAG
      AAATGAAGCTAGGTTCCCATAAACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTC

6133  TCAGCGCTGTTTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGC
      AGTCGCGACAAACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCG

6210  CACCGTTTGTAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAA
      GTGGCAAACATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTT

6287  GCGGTTTACGCCATCAATTAAACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAG
      CGCCAAATGCGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTC

6364  TGATTGGGGTGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATC
      ACTAACCCCACCACTAACCAAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAG

6441  AGGAGCTATTAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATAC
      TCCTCGATAATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATG

6518  ATCTTAAGTTCACTTGATCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAA
      TAGAATTCAAGTGAACTAGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTT

6595  AAATCCCAAACCGCAATCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGA
      TTTAGGGTTTGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCT

6672  CAATTTTTAAGATCATATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATAT
      GTTAAAAATTCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATA

6749  AACAACCAGATATTTTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTC
      TTGTTGGTCTATAAAATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAG

6826  GTTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGA
      CAAAACAAGCCCCAATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCT

6903  GCGGCTTCGCAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCC
      CGCCGAAGCGTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGG

6980  GGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGG
      CCTCCTGAGGCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCC

7057  CGGTCAATCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATA
      GCCAGTTAGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTAT
```

```
7134 CCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAA
     GGCCGCGGGTCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATT

7211 AGGGTCCAATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTC
     TCCCAGGTTAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAG

7288 CGGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGA
     GCCACGGCCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCT

7365 TGCGACTGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTC
     ACGCTGACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAG

7442 TTTATCGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCG
     AAATAGCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGC

7519 GCAGCGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGC
     CGTCGCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACG

7596 CCGGCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCG
     GGCCGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGC

7673 CTTCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCT
     GAAGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGA

7750 GTCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCC
     CAGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGG

7827 TTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGA
     AAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCT

7904 TGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACT
     ACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGA

7981 ACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCC
     TGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGG

8058 AAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAA
     TTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTT

8135 GCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGAT
     CGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTA

8212 CCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGA
     GGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCT

8289 AACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCT
     TTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGA

8366 TCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATG
     AGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTAC
```

```
8443 TTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCTT
     AACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGAA

8520 TTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGACCT
     AAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTGGA

8597 TTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTA
     AAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAAGAT

8674 ATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCGCT
     TACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGCGA

8751 GTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATT
     CACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCGCTAA

8828 GCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGTGGCCAA
     CGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCACCGGTT

8905 TGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGG
     ACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCAGGCGGCC

8982 TTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAAATGGTTGTCG
     AATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATATGGAGTTTACCAACAGC

9059 TACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTGGAGCCGGGCGAAATTAG
     ATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCACCTCGGCCCGCTTTAATC

9136 CTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCCGCCGATC
     GACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAATTGAAGAGGCGGCGGCTAG

9213 TGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATATCTGGCTCTAAGACTTCGG
     ACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTATAGACCGAGATTCTGAAGCC

9290 GCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTGTTTTTTTTTTTACCATTATTACCATCGTGTTT
     CGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAACAAAAAAAAAAATGGTAATAATGGTAGCACAAA

9367 ACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAATAAAAACAAGATATGACCTATAGAATACA
     TGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACACGGTTATTTTTGTTCTATACTGGATATCTTATGT

9444 AGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTTGTCTTCTAACCAAAAGACTTACACACCTGCATAC
     TCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACAGAAGATTGGTTTTCTGAATGTGTGGACGTATG

9521 CTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTATATACATACTTTTCAAATCGCGCG
     GAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATATAAAAAATATATGTATGAAAAGTTTAGCGCGC

9598 CCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGTTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACA
     GGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGT

9675 CATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAG
     GTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTC
```

```
 9752 TGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGTATGTACAATTTATTTGCTCCAATATATTTGTATAT
      ACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACATACATGTTAAATAAACGAGGTTATATAAACATATA
-------------------------------------------------------------------------------------
 9829 ATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTG
      TAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCAC
-------------------------------------------------------------------------------------
 9906 ATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGG
      TAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCC
-------------------------------------------------------------------------------------
 9983 CATAATAGTGTTGTTTATATATATCAAAAATAACAACTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGA
      GTATTATCACAACAAATATATATAGTTTTTATTGTTGATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCT
-------------------------------------------------------------------------------------
10060 TTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCT
      AAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGA
-------------------------------------------------------------------------------------
10137 CAAGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTT
      GTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAA
-------------------------------------------------------------------------------------
10214 ACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCAC
      TGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTG
-------------------------------------------------------------------------------------
10291 TTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCG
      AAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGC
-------------------------------------------------------------------------------------
10368 AATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTC
      TTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAG
-------------------------------------------------------------------------------------
10445 TTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTT
      AATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAA
-------------------------------------------------------------------------------------
10522 TTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACG
      AAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGC
-------------------------------------------------------------------------------------
                                                                             5' P
10599 TGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      ACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
-------------------------------------------------------------------------------------
```

pUAST-LINGO1-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGAC

  77 GGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCT
     CCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGA

 153 CGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAA
     GCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTT

 229 GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCG
     CCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGC

 305 ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTA
     TGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAAT

 381 CCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTC
     GGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAG

 457 GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
     CCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCA

 533 AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
     TTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGT

 609 GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
     CTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAA

 685 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
     ACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG

 761 GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
     CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACT

 837 TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
     AGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTC

 913 GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
     CTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGA

 989 GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
     CTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTG

1065 TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
     AGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGG

1141 ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
     TGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGA

1217 TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA
     AATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGT

1293 ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
     TGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGT

1369 ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
     TGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAG

1445 AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
     TCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGC

1521 TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
     ATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAG

1597 TTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
     AACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGA

1673 TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
     AGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTA

1749 CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
     GAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTA
```

```
1825 AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
     TTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACA

1901 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
     GAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTC

1977 TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
     ACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGC

2053 TCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA
     AGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATT

2129 GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG
     CGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATAC

2205 CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACG
     GCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGC

2281 ACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTA
     TGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGAT

2357 CGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCC
     GCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGG

2433 GGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAAT
     CCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTA

2509 CCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACA
     GGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGT

2585 AATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAG
     TTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTC

2661 CAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGA
     GTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCT

                            3'P
2737 AGCTCACGATGAGAATGGCCAGACCATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACG
     TCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGC

2813 TATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGA
     ATACGAACGTTATTCACGCTCACTTTCCTTATCATAAGACTCACAGCATAACTCAGACTCACTCTGTCGCTATACT

2889 TTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAG
     AACAACTAATTGGGAATCGTACAGGCACCCCAAACTTAATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTC

2965 TTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGG
     AAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCC

3041 AGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGG
     TCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCC

                               UAS sites
3117 AGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATT
     TCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAA

3193 CAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAA
     GTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTT

3269 CAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACT
     GTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGA
```

```
                                                                           Polylinker
3345 GAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACAGATC
     CTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGTCTAG
------------------------------------------------------------------------------------

        attB1  (hybrid)                                  (was C)
3421 TGACAAGTTTGTACAAAAAAGTTGGAAA ATG CAG GTG AGT AAG AGG ATG CTG GCG GGG GGC GTG
     ACTGTTCAAACATGTTTTTTCAACCTTT TAC GTC CAC TCA TTC TCC TAC GAC CGC CCC CCG CAC
     ---------------------------1▸ M   Q   V   S   K   R   M   L   A   G   G   V --
3485 AGG AGC ATG CCC AGC CCC CTC CTG GCC TGC TGG CAG CCC ATC CTC CTG CTG GTG CTG
     TCC TCG TAC GGG TCG GGG GAG GAC CGG ACG ACC GTC GGG TAG GAG GAC GAC CAC GAC
  13▸ R   S   M   P   S   P   L   L   A   C   W   Q   P   I   L   L   L   V   L
3542 GGC TCA GTG CTG TCA GGC TCG GCC ACG GGC TGC CCG CCC CGC TGC GAG TGC TCC GCC
     CCG AGT CAC GAC AGT CCG AGC CGG TGC CCG ACG GGC GGG GCG ACG CTC ACG AGG CGG
  32▸ G   S   V   L   S   G   S   A   T   G   C   P   P   R   C   E   C   S   A
3599 CAG GAC CGC GCT GTG CTG TGC CAC CGC AAG CGC TTT GTG GCA GTC CCC GAG GGC ATC
     GTC CTG GCG CGA CAC GAC ACG GTG GCG TTC GCG AAA CAC CGT CAG GGG CTC CCG TAG
  51▸ Q   D   R   A   V   L   C   H   R   K   R   F   V   A   V   P   E   G   I
3656 CCC ACC GAG ACG CGC CTG CTG GAC CTA GGC AAG AAC CGC ATC AAA ACG CTC AAC CAG
     GGG TGG CTC TGC GCG GAC GAC CTG GAT CCG TTC TTG GCG TAG TTT TGC GAG TTG GTC
  70▸ P   T   E   T   R   L   L   D   L   G   K   N   R   I   K   T   L   N   Q
3713 GAC GAG TTC GCC AGC TTC CCG CAC CTG GAG GAG CTG GAG CTC AAC GAG AAC ATC GTG
     CTG CTC AAG CGG TCG AAG GGC GTG GAC CTC CTC GAC CTC GAG TTG CTC TTG TAG CAC
  89▸ D   E   F   A   S   F   P   H   L   E   E   L   E   L   N   E   N   I   V
3770 AGC GCC GTG GAG CCC GGC GCC TTC AAC AAC CTC TTC AAC CTC CGG ACG CTG GGT CTC
     TCG CGG CAC CTC GGG CCG CGG AAG TTG TTG GAG AAG TTG GAG GCC TGC GAC CCA GAG
 108▸ S   A   V   E   P   G   A   F   N   N   L   F   N   L   R   T   L   G   L
3827 CGC AGC AAC CGC CTG AAG CTC ATC CCG CTA GGC GTC TTC ACT GGC CTC AGC AAC CTG
     GCG TCG TTG GCG GAC TTC GAG TAG GGC GAT CCG CAG AAG TGA CCG GAG TCG TTG GAC
 127▸ R   S   N   R   L   K   L   I   P   L   G   V   F   T   G   L   S   N   L
3884 ACC AAG CTG GAC ATC AGC GAG AAC AAG ATC GTT ATC CTA CTG GAC TAC ATG TTT CAG
     TGG TTC GAC CTG TAG TCG CTC TTG TTC TAG CAA TAG GAT GAC CTG ATG TAC AAA GTC
 146▸ T   K   L   D   I   S   E   N   K   I   V   I   L   L   D   Y   M   F   Q
3941 GAC CTG TAC AAC CTC AAG TCA CTG GAG GTT GGC GAC AAT GAC CTC GTC TAC ATC TCT
     CTG GAC ATG TTG GAG TTC AGT GAC CTC CAA CCG CTG TTA CTG GAG CAG ATG TAG AGA
 165▸ D   L   Y   N   L   K   S   L   E   V   G   D   N   D   L   V   Y   I   S
3998 CAC CGC GCC TTC AGC GGC CTC AAC AGC CTG GAG CAG CTG ACG CTG GAG AAA TGC AAC
     GTG GCG CGG AAG TCG CCG GAG TTG TCG GAC CTC GTC GAC TGC GAC CTC TTT ACG TTG
 184▸ H   R   A   F   S   G   L   N   S   L   E   Q   L   T   L   E   K   C   N
4055 CTG ACC TCC ATC CCC ACC GAG GCG CTG TCC CAC CTG CAC GGC CTC ATC GTC CTG AGG
     GAC TGG AGG TAG GGG TGG CTC CGC GAC AGG GTG GAC GTG CCG GAG TAG CAG GAC TCC
 203▸ L   T   S   I   P   T   E   A   L   S   H   L   H   G   L   I   V   L   R
4112 CTC CGG CAC CTC AAC ATC AAT GCC ATC CGG GAC TAC TCC TTC AAG AGG CTG TAC CGA
     GAG GCC GTG GAG TTG TAG TTA CGG TAG GCC CTG ATG AGG AAG TTC TCC GAC ATG GCT
 222▸ L   R   H   L   N   I   N   A   I   R   D   Y   S   F   K   R   L   Y   R
4169 CTC AAG GTC TTG GAG ATC TCC CAC TGG CCC TAC TTG GAC ACC ATG ACA CCC AAC TGC
     GAG TTC CAG AAC CTC TAG AGG GTG ACC GGG ATG AAC CTG TGG TAC TGT GGG TTG ACG
 241▸ L   K   V   L   E   I   S   H   W   P   Y   L   D   T   M   T   P   N   C
4226 CTC TAC GGC CTC AAC CTG ACG TCC CTG TCC ATC ACA CAC TGC AAT CTG ACC GCT GTG
     GAG ATG CCG GAG TTG GAC TGC AGG GAC AGG TAG TGT GTG ACG TTA GAC TGG CGA CAC
 260▸ L   Y   G   L   N   L   T   S   L   S   I   T   H   C   N   L   T   A   V
```

```
4283 CCC TAC CTG GCC GTC CGC CAC CTA GTC TAT CTC CGC TTC CTC AAC CTC TCC TAC AAC
     GGG ATG GAC CGG CAG GCG GTG GAT CAG ATA GAG GCG AAG GAG TTG GAG AGG ATG TTG
 279▸ P   Y   L   A   V   R   H   L   V   Y   L   R   F   L   N   L   S   Y   N

4340 CCC ATC AGC ACC ATT GAG GGC TCC ATG TTG CAT GAG CTG CTC CGG CTG CAG GAG ATC
     GGG TAG TCG TGG TAA CTC CCG AGG TAC AAC GTA CTC GAC GAG GCC GAC GTC CTC TAG
 298▸ P   I   S   T   I   E   G   S   M   L   H   E   L   L   R   L   Q   E   I

4397 CAG CTG GTG GGC GGG CAG CTG GCC GTG GTG GAG CCC TAT GCC TTC CGC GGC CTC AAC
     GTC GAC CAC CCG CCC GTC GAC CGG CAC CAC CTC GGG ATA CGG AAG GCG CCG GAG TTG
 317▸ Q   L   V   G   G   Q   L   A   V   V   E   P   Y   A   F   R   G   L   N

4454 TAC CTG CGC GTG CTC AAT GTC TCT GGC AAC CAG CTG ACC ACA CTG GAG GAA TCA GTC
     ATG GAC GCG CAC GAG TTA CAG AGA CCG TTG GTC GAC TGG TGT GAC CTC CTT AGT CAG
 336▸ Y   L   R   V   L   N   V   S   G   N   Q   L   T   T   L   E   E   S   V

4511 TTC CAC TCG GTG GGC AAC CTG GAG ACA CTC ATC CTG GAC TCC AAC CCG CTG GCC TGC
     AAG GTG AGC CAC CCG TTG GAC CTC TGT GAG TAG GAC CTG AGG TTG GGC GAC CGG ACG
 355▸ F   H   S   V   G   N   L   E   T   L   I   L   D   S   N   P   L   A   C

4568 GAC TGT CGG CTC CTG TGG GTG TTC CGG CGC CGC TGG CGG CTC AAC TTC AAC CGG CAG
     CTG ACA GCC GAG GAC ACC CAC AAG GCC GCG GCG ACC GCC GAG TTG AAG TTG GCC GTC
 374▸ D   C   R   L   L   W   V   F   R   R   R   W   R   L   N   F   N   R   Q

4625 CAG CCC ACG TGC GCC ACG CCC GAG TTT GTC CAG GGC AAG GAG TTC AAG GAC TTC CCT
     GTC GGG TGC ACG CGG TGC GGG CTC AAA CAG GTC CCG TTC CTC AAG TTC CTG AAG GGA
 393▸ Q   P   T   C   A   T   P   E   F   V   Q   G   K   E   F   K   D   F   P

4682 GAT GTG CTA CTG CCC AAC TAC TTC ACC TGC CGC CGC GCC CGC ATC CGG GAC CGC AAG
     CTA CAC GAT GAC GGG TTG ATG AAG TGG ACG GCG GCG CGG GCG TAG GCC CTG GCG TTC
 412▸ D   V   L   L   P   N   Y   F   T   C   R   R   A   R   I   R   D   R   K

4739 GCC CAG CAG GTG TTT GTG GAC GAG GGC CAC ACG GTG CAG TTT GTG TGC CGG GCC GAT
     CGG GTC GTC CAC AAA CAC CTG CTC CCG GTG TGC CAC GTC AAA CAC ACG GCC CGG CTA
 431▸ A   Q   Q   V   F   V   D   E   G   H   T   V   Q   F   V   C   R   A   D

4796 GGC GAC CCG CCG CCC GCC ATC CTC TGG CTC TCA CCC CGA AAG CAC CTG GTC TCA GCC
     CCG CTG GGC GGC GGG CGG TAG GAG ACC GAG AGT GGG GCT TTC GTG GAC CAG AGT CGG
 450▸ G   D   P   P   P   A   I   L   W   L   S   P   R   K   H   L   V   S   A

4853 AAG AGC AAT GGG CGG CTC ACA GTC TTC CCT GAT GGC ACG CTG GAG GTG CGC TAC GCC
     TTC TCG TTA CCC GCC GAG TGT CAG AAG GGA CTA CCG TGC GAC CTC CAC GCG ATG CGG
 469▸ K   S   N   G   R   L   T   V   F   P   D   G   T   L   E   V   R   Y   A

4910 CAG GTA CAG GAC AAC GGC ACG TAC CTG TGC ATC GCG GCC AAC GCG GGC GGC AAC GAC
     GTC CAT GTC CTG TTG CCG TGC ATG GAC ACG TAG CGC CGG TTG CGC CCG CCG TTG CTG
 488▸ Q   V   Q   D   N   G   T   Y   L   C   I   A   A   N   A   G   G   N   D

4967 TCC ATG CCC GCC CAC CTG CAT GTG CGC AGC TAC TCG CCC GAC TGG CCC CAT CAG CCC
     AGG TAC GGG CGG GTG GAC GTA CAC GCG TCG ATG AGC GGG CTG ACC GGG GTA GTC GGG
 507▸ S   M   P   A   H   L   H   V   R   S   Y   S   P   D   W   P   H   Q   P

5024 AAC AAG ACC TTC GCT TTC ATC TCC AAC CAG CCG GGC GAG GGA GAG GCC AAC AGC ACC
     TTG TTC TGG AAG CGA AAG TAG AGG TTG GTC GGC CCG CTC CCT CTC CGG TTG TCG TGG
 526▸ N   K   T   F   A   F   I   S   N   Q   P   G   E   G   E   A   N   S   T

                                                                     TM
5081 CGC GCC ACT GTG CCT TTC CCC TTC GAC ATC AAG ACC CTC ATC ATC GCC ACC ACC ATG
     GCG CGG TGA CAC GGA AAG GGG AAG CTG TAG TTC TGG GAG TAG TAG CGG TGG TGG TAC
 545▸ R   A   T   V   P   F   P   F   D   I   K   T   L   I   I   A   T   T   M

5138 GGC TTC ATC TCT TTC CTG GGC GTC GTC CTC TTC TGC CTG GTG CTG CTG TTT CTC TGG
     CCG AAG TAG AGA AAG GAC CCG CAG CAG GAG AAG ACG GAC CAC GAC GAC AAA GAG ACC
 564▸ G   F   I   S   F   L   G   V   V   L   F   C   L   V   L   L   F   L   W

5195 AGC CGG GGC AAG GGC AAC ACA AAG CAC AAC ATC GAG ATC GAG TAT GTG CCC CGA AAG
     TCG GCC CCG TTC CCG TTG TGT TTC GTG TTG TAG CTC TAG CTC ATA CAC GGG GCT TTC
 583▸ S   R   G   K   G   N   T   K   H   N   I   E   I   E   Y   V   P   R   K
```

```
5252 TCG GAC GCA GGC ATC AGC TCC GCC GAC GCG CCC CGC AAG TTC AAC ATG AAG ATG ATA
     AGC CTG CGT CCG TAG TCG AGG CGG CTG CGC GGG GCG TTC AAG TTG TAC TTC TAC TAT
 602▸ S   D   A   G   I   S   S   A   D   A   P   R   K   F   N   M   K   M   I

                                                      V5  epitope
                                         attB2  (hybrid)
5309 CCA ACT TTC TTG TAC AAA GTG GTG GTA CCG  GGT AAG CCT ATC CCT AAC CCT CTC CTC

     GGT TGA AAG AAC ATG TTT CAC CAC CAT GGC  CCA TTC GGA TAG GGA TTG GGA GAG GAG
 621▸ P   T   F   L   Y   K   V   V   V   P    G   K   P   I   P   N   P   L   L

                                         6xHis                          SV40  Poly  A
5366 GGT CTC GAT TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA TCTAGAGGATCTTTGT
     CCA GAG CTA AGA TGC GCA TGG CCA GTA GTA GTG GTA GTG GTA ACT AGATCTCCTAGAAACA
 640▸ G   L   D   S   T   R   T   G   H   H   H   H   H   H   •

5427 GAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATA
     CTTCCTTGGAATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATAT

5503 AAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTG
     TTTAAAAATTCACATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGAC

5579 ATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGA
     TACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACT

5655 TGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCT
     ACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGA

5731 TCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAA
     AGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTT

5807 AGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTA
     TCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAAT

5883 TAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTG
     ATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAAC

5959 TGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATA
     ACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTAT

6035 ATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATA
     TAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTAT

6111 AAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
     TTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTT

6187 TTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTC
     AAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAG

                  white gene
6263 TGGATCGGATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAA
     ACCTAGCCTAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTT
```

```
6339 AAACATATTTTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGAT
     TTTGTATAAAAAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTA

6415 CCAAGGGTATTTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGT
     GGTTCCCATAAACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACA

6491 TTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTG
     AACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAAC

6567 TAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTA
     ATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAAT

6643 CGCCATCAATTAAACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGG
     GCGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACC

6719 GGTGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGC
     CCACCACTAACCAAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCG

6795 TATTAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTT
     ATAATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAA

6871 AAGTTCACTTGATCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAAT
     TTCAAGTGAACTAGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTA

6947 CCCAAACCGCAATCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAA
     GGGTTTGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTT

7023 TTTTTAAGATCATATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAA
     AAAAATTCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATT

7099 CAACCAGATATTTTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCG
     GTTGGTCTATAAAATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGC

7175 TTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGA
     AAAACAAGCCCCAATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCT

7251 GCGGCTTCGCAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGC
     CGCCGAAGCGTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCG

7327 CGGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGG
     GCCTCCTGAGGCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACC

7403 GGCGGTCAATCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACAC
     CCGCCAGTTAGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTG

7479 ATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTA
     TATGGCCGCGGGTCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAAT

7555 ATAAAGGGTCCAATTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGC
     TATTTCCCAGGTTAATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCG
```

```
7631 AGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCAT
     TCAAGGCCACGGCCTTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTA

7707 CCGGGATGCGACTGCTCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGA
     GGCCCTACGCTGACGAGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCT

7783 TGACCTCTTTATCGGCTCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTG
     ACTGGAGAAATAGCCGAGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGAC

7859 ACCTATCGGCAGCGAGTGGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCA
     TGGATAGCCGTCGCTCACCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGT

7935 TCGGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAAC
     AGCCACACGGGCCGTCCCACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTG

8011 CGATCCGCCGCTTCTGATCTGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTG
     GCTAGGCGGCGAAGACTAGACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCAC

8087 CTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCT
     GACTTCTTCGACAGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGA

8163 TTGACAAGATCCTTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTC
     AACTGTTCTAGGAAGACTACCGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAG

8239 CTAGTGAGTTCGATGTGTTTATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCA
     GATCACTCAAGCTACACAAATAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGT

8315 GTGTCCTACCAACTACAATCCGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCC
     CACAGGATGGTTGATGTTAGGCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGG

8391 CGTGATCGGATCGCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCA
     GCACTAGCCTAGCGGTTCTATACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGT

8467 CCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCG
     GGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGC

8543 GGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACG
     CCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGC

8619 GTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAA
     CACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTT

8695 TTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATT
     AACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAA

8771 TCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGC
     AGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACG

8847 CATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTC
     GTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAG
```

```
 8923 CTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCAT
      GAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTA

 8999 ATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGG
      TAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCC

 9075 AGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCAC
      TCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTG

 9151 AGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGC
      TCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACG

 9227 CTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCT
      GACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGA

 9303 CGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTC
      GCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAG

 9379 GGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAA
      CCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTT

 9455 TGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCC
      ACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGG

 9531 TGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCG
      ACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGC

 9607 GGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTG
      CCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAAC

 9683 TTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTG
      AAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACAC

 9759 CCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTT
      GGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAA

 9835 GTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGA
      CAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCT

 9911 TATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGT
      ATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCA

 9987 TGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTT
      ACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAA

10063 AGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATT
      TCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAA

10139 TGTATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTA
      ACATACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAAT
```

```
10215 TGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGT
      ACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10291 GTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAA
      CAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10367 CTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTC
      GATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10443 TAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGAC
      ATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10519 CTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCA
      GATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10595 CTGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCC
      GACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10671 TCCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGA
      AGGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10747 GTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAAT
      CAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10823 TTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTT
      AAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10899 CAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGT
      GTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
                                                 5' P
10975 AAGCTTCGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      TTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

pUAST-sDER-GFP

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTGGG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGACCC

  79 GCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCTCGTT
     CGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGAGCAA

 157 GGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAAGGCGTT
     CCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTTCCGCAA

 235 TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACT
     AAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTCCTGA

 313 ATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT
     TATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGA

 391 GTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGT
     CAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCA

 469 TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
     AGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACT

 547 GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
     CAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCCATACATCC

 625 CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCT
     GCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGACGCGAGACGA

 703 GAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
     CTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGCCACCAAAAAA

 781 TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGC
     ACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGCCCCAGACTGCG

 859 TCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAA
     AGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTT

 937 TTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
     AATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACT

1015 GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACG
     CCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGC

1093 GGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAAT
     CCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTA

1171 AAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
     TTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAAC

1249 CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTC
     GGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACCACAG

1327 ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTG
     TGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACAC

1405 CAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTAT
     GTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATA

1483 GGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTC
     CCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAG

1561 ATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAG
     TAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTC

1639 AACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
     TTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTC

1717 TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAC
     AAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTG

1795 AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
     TCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTAT
```

```
1873 TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGG
     AATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCC

1951 GGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAA
     CCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATTTTT

2029 TAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGA
     ATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCT

2107 GACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTG
     CTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCAC

2185 TCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACCG
     AGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGGC

2263 AATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGAGAGCGGGTGGGAGAC
     TTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCTCTCGCCCACCCTCTG

2341 AGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGATTTGAAAAGTATGTATAT
     TCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCTAAACTTTTCATACATATA

2419 AAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATGCAGGTGTGTAAGTCTTTTGGT
     TTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATACGTCCACACATTCAGAAAACCA

2497 TAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGTATTCTATAGGTCATATCTTGTTTT
     ATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACATAAGATATCCAGTATAGAACAAAA

2575 TATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTAAACACGATGGTAATAATGGTAAAAAAA
     ATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCATTTGTGCTACCATTATTACCATTTTTTT

2653 AAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCA
     TTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGT

                                3 ' P
2731 CCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTT
     GGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAA

2809 AACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATAT
     TTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAAGACTCACAGCATAACTCAGACTCACTCTGTCGCTATA

2887 GATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAG
     CTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTTAATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTC

2965 TTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAG
     AAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTC

3043 TACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTA
     ATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCAT

                         UAS sites
3121 CTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTC
     GACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAG

3199 AAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGC
     TTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACG

3277 AGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCC
     TCATTTCACGTTCAATTTCACTTAGTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGG
```

```
                                                                          Polylinker            attB1   (hyl
3355 AAGAAGTAATTATTGAATACAAGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGT
     TTCTTCATTAATAACTTATGTTCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAAACA

------------------------------------------------------------------------------------

                                      was G
3433 ACAAAAAAGTTGGAAA ATG CTG CTa CGA CGG CGC AAC GGC CCC TGC CCC TTC CCC CTG CTG
     TGTTTTTTCAACCTTT TAC GAC GAt GCT GCC GCG TTG CCG GGG ACG GGG AAG GGG GAC GAC
                   1▸ M   L   L   R   R   R   N   G   P   C   P   F   P   L   L
------------------------------------------------------------------------------------
3494 CTT CTG CTC CTG GCC CAC TGC ATT TGC ATT TGG CCC GCG TCG GCG GCC CGC GAT CGC
     GAA GAC GAG GAC CGG GTG ACG TAA ACG TAA ACC GGG CGC AGC CGC CGG GCG CTA GCG
  16▸ L   L   L   L   A   H   C   I   C   I   W   P   A   S   A   A   R   D   R
------------------------------------------------------------------------------------
3551 TAC GCC CGC CAG AAC AAT CGC CAG CGC CAT CAG GAT ATA GAT CGC GAT CGG GAT CGA
     ATG CGG GCG GTC TTG TTA GCG GTC GCG GTA GTC CTA TAT CTA GCG CTA GCC CTA GCT
  35▸ Y   A   R   Q   N   N   R   Q   R   H   Q   D   I   D   R   D   R   D   R
------------------------------------------------------------------------------------
3608 GAT CGA TTC CTA TAC CGC AGC AGT TCG GCC CAA AAT CGA CAG AGG GGC GGG GCC AAC
     CTA GCT AAG GAT ATG GCG TCG TCA AGC CGG GTT TTA GCT GTC TCC CCG CCC CGG TTG
  54▸ D   R   F   L   Y   R   S   S   S   A   Q   N   R   Q   R   G   G   A   N
------------------------------------------------------------------------------------
3665 TTC GCC CTG GGA CTG GGA GCC AAC GGA GTC ACC ATT CCC ACC AGT CTG GAG GAT AAG
     AAG CGG GAC CCT GAC CCT CGG TTG CCT CAG TGG TAA GGG TGG TCA GAC CTC CTA TTC
  73▸ F   A   L   G   L   G   A   N   G   V   T   I   P   T   S   L   E   D   K
------------------------------------------------------------------------------------
3722 AAC AAG AAC GAG TTC GTC AAG GGG AAA ATC TGC ATC GGC ACT AAA TCT CGG CTC TCC
     TTG TTC TTG CTC AAG CAG TTC CCC TTT TAG ACG TAG CCG TGA TTT AGA GCC GAG AGG
  92▸ N   K   N   E   F   V   K   G   K   I   C   I   G   T   K   S   R   L   S
------------------------------------------------------------------------------------
3779 GTG CCC TCC AAC AAG GAA CAT CAT TAC CGA AAC CTC AGA GAT CGG TAC ACG AAC TGT
     CAC GGG AGG TTG TTC CTT GTA GTA ATG GCT TTG GAG TCT CTA GCC ATG TGC TTG ACA
 111▸ V   P   S   N   K   E   H   H   Y   R   N   L   R   D   R   Y   T   N   C
------------------------------------------------------------------------------------
3836 ACG TAT GTG GAT GGC AAC TTG AAA CTG ACC TGG CTA CCC AAC GAG AAT TTG GAC CTC
     TGC ATA CAC CTA CCG TTG AAC TTT GAC TGG ACC GAT GGG TTG CTC TTA AAC CTG GAG
 130▸ T   Y   V   D   G   N   L   K   L   T   W   L   P   N   E   N   L   D   L
------------------------------------------------------------------------------------
3893 AGC TTC CTA GAC AAC ATA CGG GAG GTC ACC GGC TAT ATT CTG ATC AGT CAT GTG GAC
     TCG AAG GAT CTG TTG TAT GCC CTC CAG TGG CCG ATA TAA GAC TAG TCA GTA CAC CTG
 149▸ S   F   L   D   N   I   R   E   V   T   G   Y   I   L   I   S   H   V   D
------------------------------------------------------------------------------------
3950 GTT AAG AAA GTG GTG TTT CCC AAA CTA CAA ATC ATT CGC GGA CGC ACG CTG TTC AGC
     CAA TTC TTT CAC CAC AAA GGG TTT GAT GTT TAG TAA GCG CCT GCG TGC GAC AAG TCG
 168▸ V   K   K   V   V   F   P   K   L   Q   I   I   R   G   R   T   L   F   S
------------------------------------------------------------------------------------
4007 TTA TCC GTG GAG GAG GAG AAG TAT GCC TTG TTC GTC ACT TAT TCC AAA ATG TAC ACG
     AAT AGG CAC CTC CTC CTC TTC ATA CGG AAC AAG CAG TGA ATA AGG TTT TAC ATG TGC
 187▸ L   S   V   E   E   E   K   Y   A   L   F   V   T   Y   S   K   M   Y   T
------------------------------------------------------------------------------------
4064 CTG GAG ATT CCC GAT CTA CGC GAT GTC TTA AAT GGC CAA GTG GGC TTC CAC AAC AAC
     GAC CTC TAA GGG CTA GAT GCG CTA CAG AAT TTA CCG GTT CAC CCG AAG GTG TTG TTG
 206▸ L   E   I   P   D   L   R   D   V   L   N   G   Q   V   G   F   H   N   N
------------------------------------------------------------------------------------
4121 TAC AAT CTC TGC CAC ATG CGA ACG ATC CAG TGG TCG GAG ATT GTA TCC AAC GGC ACG
     ATG TTA GAG ACG GTG TAC GCT TGC TAG GTC ACC AGC CTC TAA CAT AGG TTG CCG TGC
 225▸ Y   N   L   C   H   M   R   T   I   Q   W   S   E   I   V   S   N   G   T
------------------------------------------------------------------------------------
4178 GAT GCA TAC TAC AAC TAC GAC TTT ACT GCT CCG GAG CGC GAG TGT CCC AAG TGC CAC
     CTA CGT ATG ATG TTG ATG CTG AAA TGA CGA GGC CTC GCG CTC ACA GGG TTC ACG GTG
 244▸ D   A   Y   Y   N   Y   D   F   T   A   P   E   R   E   C   P   K   C   H
------------------------------------------------------------------------------------
4235 GAG AGC TGC ACG CAC GGA TGT TGG GGC GAG GGT CCC AAG AAT TGC CAG AAG TTC AGC
     CTC TCG ACG TGC GTG CCT ACA ACC CCG CTC CCA GGG TTC TTA ACG GTC TTC AAG TCG
 263▸ E   S   C   T   H   G   C   W   G   E   G   P   K   N   C   Q   K   F   S
```

```
4292 AAG CTC ACC TGC TCG CCA CAG TGT GCC GGA GGT CGT TGC TAT GGA CCA AAG CCG CGG
     TTC GAG TGG ACG AGC GGT GTC ACA CGG CCT CCA GCA ACG ATA CCT GGT TTC GGC GCC
 282▸ K   L   T   C   S   P   Q   C   A   G   G   R   C   Y   G   P   K   P   R

4349 GAG TGT TGT CAC CTC TTC TGC GCC GGA GGA TGC ACT GGT CCC ACG CAA AAG GAT TGC
     CTC ACA ACA GTG GAG AAG ACG CGG CCT CCT ACG TGA CCA GGG TGC GTT TTC CTA ACG
 301▸ E   C   C   H   L   F   C   A   G   G   C   T   G   P   T   Q   K   D   C

4406 ATC GCC TGC AAG AAC TTC TTC GAC GAG GCA GTA TCA AAG GAG GAA TGC CCG CCC ATG
     TAG CGG ACG TTC TTG AAG AAG CTG CTC CGT CAT AGT TTC CTC CTT ACG GGC GGG TAC
 320▸ I   A   C   K   N   F   F   D   E   A   V   S   K   E   E   C   P   P   M

4463 CGC AAG TAC AAT CCC ACC ACC TAT GTT CTT GAA ACG AAT CCT GAG GGA AAG TAT GCC
     GCG TTC ATG TTA GGG TGG TGG ATA CAA GAA CTT TGC TTA GGA CTC CCT TTC ATA CGG
 339▸ R   K   Y   N   P   T   T   Y   V   L   E   T   N   P   E   G   K   Y   A

4520 TAT GGT GCC ACC TGC GTC AAG GAG TGT CCC GGT CAT CTG TTG CGG GAT AAT GGC GCC
     ATA CCA CGG TGG ACG CAG TTC CTC ACA GGG CCA GTA GAC AAC GCC CTA TTA CCG CGG
 358▸ Y   G   A   T   C   V   K   E   C   P   G   H   L   L   R   D   N   G   A

4577 TGC GTG CGC AGC TGT CCC CAG GAC AAG ATG GAC AAG GGG GGC GAG TGT GTG CCC TGC
     ACG CAC GCG TCG ACA GGG GTC CTG TTC TAC CTG TTC CCC CCG CTC ACA CAC GGG ACG
 377▸ C   V   R   S   C   P   Q   D   K   M   D   K   G   G   E   C   V   P   C

4634 AAT GGA CCG TGC CCC AAA ACC TGC CCG GGC GTT ACT GTC CTG CAT GCC GGC AAC ATT
     TTA CCT GGC ACG GGG TTT TGG ACG GGC CCG CAA TGA CAG GAC GTA CGG CCG TTG TAA
 396▸ N   G   P   C   P   K   T   C   P   G   V   T   V   L   H   A   G   N   I

4691 GAC TCG TTC CGG AAT TGT ACG GTG ATC GAT GGC AAC ATT CGC ATT TTG GAT CAG ACC
     CTG AGC AAG GCC TTA ACA TGC CAC TAG CTA CCG TTG TAA GCG TAA AAC CTA GTC TGG
 415▸ D   S   F   R   N   C   T   V   I   D   G   N   I   R   I   L   D   Q   T

4748 TTC TCG GGC TTC CAG GAT GTC TAT GCC AAC TAC ACG ATG GGA CCA CGA TAC ATA CCG
     AAG AGC CCG AAG GTC CTA CAG ATA CGG TTG ATG TGC TAC CCT GGT GCT ATG TAT GGC
 434▸ F   S   G   F   Q   D   V   Y   A   N   Y   T   M   G   P   R   Y   I   P

4805 CTG GAT CCC GAG CGA CGG GAG GTG TTC TCC ACG GTG AAG GAG ATC ACC GGG TAT CTG
     GAC CTA GGG CTC GCT GCC CTC CAC AAG AGG TGC CAC TTC CTC TAG TGG CCC ATA GAC
 453▸ L   D   P   E   R   R   E   V   F   S   T   V   K   E   I   T   G   Y   L

4862 AAT ATC GAG GGA ACC CAC CCG CAG TTC CGG AAT CTG TCG TAC TTT CGC AAT CTG GAA
     TTA TAG CTC CCT TGG GTG GGC GTC AAG GCC TTA GAC AGC ATG AAA GCG TTA GAC CTT
 472▸ N   I   E   G   T   H   P   Q   F   R   N   L   S   Y   F   R   N   L   E

4919 ACA ATT CAT GGC CGC CAG CTG ATG GAG AGC ATG TTT GCC GCT TTG GCG ATC GTT AAG
     TGT TAA GTA CCG GCG GTC GAC TAC CTC TCG TAC AAA CGG CGA AAC CGC TAG CAA TTC
 491▸ T   I   H   G   R   Q   L   M   E   S   M   F   A   A   L   A   I   V   K

4976 TCA TCC CTG TAC AGC CTG GAG ATG CGC AAT CTG AAG CAG ATT AGT TCC GGC AGT GTG
     AGT AGG GAC ATG TCG GAC CTC TAC GCG TTA GAC TTC GTC TAA TCA AGG CCG TCA CAC
 510▸ S   S   L   Y   S   L   E   M   R   N   L   K   Q   I   S   S   G   S   V

5033 GTC ATC CAG CAT AAT AGA GAC CTC TGC TAC GTA AGC AAT ATC CGT TGG CCG GCC ATT
     CAG TAG GTC GTA TTA TCT CTG GAG ACG ATG CAT TCG TTA TAG GCA ACC GGC CGG TAA
 529▸ V   I   Q   H   N   R   D   L   C   Y   V   S   N   I   R   W   P   A   I

5090 CAG AAG GAG CCC GAA CAG AAG GTG TGG GTC AAC GAG AAT CTC AGG GCG GAT CTA TGC
     GTC TTC CTC GGG CTT GTC TTC CAC ACC CAG TTG CTC TTA GAG TCC CGC CTA GAT ACG
 548▸ Q   K   E   P   E   Q   K   V   W   V   N   E   N   L   R   A   D   L   C

5147 GAG AAA AAT GGA ACC ATT TGC TCG GAT CAG TGC AAC GAG GAC GGC TGC TGG GGA GCT
     CTC TTT TTA CCT TGG TAA ACG AGC CTA GTC ACG TTG CTC CTG CCG ACG ACC CCT CGA
 567▸ E   K   N   G   T   I   C   S   D   Q   C   N   E   D   G   C   W   G   A

5204 GGC ACG GAT CAG TGC CTT ACC TGC AAG AAC TTC AAT TTC AAT GGC ACC TGC ATC GCC
     CCG TGC CTA GTC ACG GAA TGG ACG TTC TTG AAG TTA AAG TTA CCG TGG ACG TAG CGG
 586▸ G   T   D   Q   C   L   T   C   K   N   F   N   F   N   G   T   C   I   A
```

```
5261 GAC TGT GGT TAT ATA TCC AAT GCC TAC AAG TTT GAC AAT AGA ACG TGC AAG ATA TGC
     CTG ACA CCA ATA TAT AGG TTA CGG ATG TTC AAA CTG TTA TCT TGC ACG TTC TAT ACG
 605▶ D   C   G   Y   I   S   N   A   Y   K   F   D   N   R   T   C   K   I   C

5318 CAT CCA GAG TGC CGG ACT TGC AAT GGA GCT GGA GCA GAT CAC TGC CAG GAG TGC GTC
     GTA GGT CTC ACG GCC TGA ACG TTA CCT CGA CCT CGT CTA GTG ACG GTC CTC ACG CAG
 624▶ H   P   E   C   R   T   C   N   G   A   G   A   D   H   C   Q   E   C   V

5375 CAT GTG AGG GAC GGT CAG CAC TGT GTG TCC GAG TGC CCG AAG AAC AAG TAC AAC GAT
     GTA CAC TCC CTG CCA GTC GTG ACA CAC AGG CTC ACG GGC TTC TTG TTC ATG TTG CTA
 643▶ H   V   R   D   G   Q   H   C   V   S   E   C   P   K   N   K   Y   N   D

5432 CGT GGT GTC TGC CGA GAG TGC CAC GCC ACC TGC GAT GGA TGC ACT GGG CCC AAG GAC
     GCA CCA CAG ACG GCT CTC ACG GTG CGG TGG ACG CTA CCT ACG TGA CCC GGG TTC CTG
 662▶ R   G   V   C   R   E   C   H   A   T   C   D   G   C   T   G   P   K   D

5489 ACC ATC GGC ATT GGA GCG TGT ACG ACG TGC AAT TTG GCC ATT ATC AAC AAT GAC GCC
     TGG TAG CCG TAA CCT CGC ACA TGC TGC ACG TTA AAC CGG TAA TAG TTG TTA CTG CGG
 681▶ T   I   G   I   G   A   C   T   T   C   N   L   A   I   I   N   N   D   A

5546 ACA GTA AAA CGC TGC CTG CTG AAG GAC GAC AAG TGC CCC GAT GGG TAT TTC TGG GAG
     TGT CAT TTT GCG ACG GAC GAC TTC CTG CTG TTC ACG GGG CTA CCC ATA AAG ACC CTC
 700▶ T   V   K   R   C   L   L   K   D   D   K   C   P   D   G   Y   F   W   E

5603 TAT GTG CAT CCG CAA GAG CAG GGA TCG CTG AAG CCA TTG GCC GGC AGA GCA GTT TGC
     ATA CAC GTA GGC GTT CTC GTC CCT AGC GAC TTC GGT AAC CGG CCG TCT CGT CAA ACG
 719▶ Y   V   H   P   Q   E   Q   G   S   L   K   P   L   A   G   R   A   V   C

5660 CGA AAG TGC CAT CCC CTT TGC GAG CTG TGC ACG AAC TAC GGA TAC CAT GAA CAG GTG
     GCT TTC ACG GTA GGG GAA ACG CTC GAC ACG TGC TTG ATG CCT ATG GTA CTT GTC CAC
 738▶ R   K   C   H   P   L   C   E   L   C   T   N   Y   G   Y   H   E   Q   V

5717 TGC TCC AAG TGC ACC CAC TAC AAG CGA CGG GAG CAG TGC GAG ACC GAG TGT CCG GCC
     ACG AGG TTC ACG TGG GTG ATG TTC GCT GCC CTC GTC ACG CTC TGG CTC ACA GGC CGG
 757▶ C   S   K   C   T   H   Y   K   R   R   E   Q   C   E   T   E   C   P   A

5774 GAT CAC TAC ACG GAT GAG GAG CAG CGC GAG TGC TTC CAG CGC CAC CCG GAA TGC AAT
     CTA GTG ATG TGC CTA CTC CTC GTC GCG CTC ACG AAG GTC GCG GTG GGC CTT ACG TTA
 776▶ D   H   Y   T   D   E   E   Q   R   E   C   F   Q   R   H   P   E   C   N

5831 GGT TGC ACG GGT CCG GGT GCC GAC GAT TGC AAG TCT TGC CGC AAC TTT AAG TTG TTC
     CCA ACG TGC CCA GGC CCA CGG CTG CTA ACG TTC AGA ACG GCG TTG AAA TTC AAC AAG
 795▶ G   C   T   G   P   G   A   D   D   C   K   S   C   R   N   F   K   L   F

5888 GAC GCG AAT GAG ACG GGT CCC TAT GTG AAC TCC ACG ATG TTC AAT TGC ACC TCG AAG
     CTG CGC TTA CTC TGC CCA GGG ATA CAC TTG AGG TGC TAC AAG TTA ACG TGG AGC TTC
 814▶ D   A   N   E   T   G   P   Y   V   N   S   T   M   F   N   C   T   S   K

5945 TGT CCC TTG GAG ATG CGA CAT GTG AAC TAT CAG TAC ACG GCC ATT GGA CCC TAC TGC
     ACA GGG AAC CTC TAC GCT GTA CAC TTG ATA GTC ATG TGC CGG TAA CCT GGG ATG ACG
 833▶ C   P   L   E   M   R   H   V   N   Y   Q   Y   T   A   I   G   P   Y   C

                                                                   attB2  (hybri
6002 GCA GCT AGT CCG CCG AGG AGC AGC AAG CCA ACT TTC TTG TAC AAA GTG GTG GTA CCG
     CGT CGA TCA GGC GGC TCC TCG TCG TTC GGT TGA AAG AAC ATG TTT CAC CAC CAT GGC
 852▶ A   A   S   P   P   R   S   S   K   P   T   F   L   Y   K   V   V   V   P

                                       EGFP
6059 CGG GCC CGG GAT CCA CCG GTC GCC ACC ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC
     GCC CGG GCC CTA GGT GGC CAG CGG TGG TAC CAC TCG TTC CCG CTC CTC GAC AAG TGG
 871▶ R   A   R   D   P   P   V   A   T   M   V   S   K   G   E   E   L   F   T

6116 GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC
     CCC CAC CAC GGG TAG GAC CAG CTC GAC CTG CCG CTG CAT TTG CCG GTG TTC AAG TCG
 890▶ G   V   V   P   I   L   V   E   L   D   G   D   V   N   G   H   K   F   S
```

```
6173 GTG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC
     CAC AGG CCG CTC CCG CTC CCG CTA CGG TGG ATG CCG TTC GAC TGG GAC TTC AAG TAG
 909▸ V   S   G   E   G   E   G   D   A   T   Y   G   K   L   T   L   K   F   I

6230 TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC CTG ACC TAC
     ACG TGG TGG CCG TTC GAC GGG CAC GGG ACC GGG TGG GAG CAC TGG TGG GAC TGG ATG
 928▸ C   T   T   G   K   L   P   V   P   W   P   T   L   V   T   T   L   T   Y

6287 GGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC AAG
     CCG CAC GTC ACG AAG TCG GCG ATG GGG CTG GTG TAC TTC GTC GTG CTG AAG AAG TTC
 947▸ G   V   Q   C   F   S   R   Y   P   D   H   M   K   Q   H   D   F   F   K

6344 TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC
     AGG CGG TAC GGG CTT CCG ATG CAG GTC CTC GCG TGG TAG AAG AAG TTC CTG CTG CCG
 966▸ S   A   M   P   E   G   Y   V   Q   E   R   T   I   F   F   K   D   D   G

6401 AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CGC ATC
     TTG ATG TTC TGG GCG CGG CTC CAC TTC AAG CTC CCG CTG TGG GAC CAC TTG GCG TAG
 985▸ N   Y   K   T   R   A   E   V   K   F   E   G   D   T   L   V   N   R   I

6458 GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG
     CTC GAC TTC CCG TAG CTG AAG TTC CTC CTG CCG TTG TAG GAC CCC GTG TTC GAC CTC
1004▸ E   L   K   G   I   D   F   K   E   D   G   N   I   L   G   H   K   L   E

6515 TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC
     ATG TTG ATG TTG TCG GTG TTG CAG ATA TAG TAC CGG CTG TTC GTC TTC TTG CCG TAG
1023▸ Y   N   Y   N   S   H   N   V   Y   I   M   A   D   K   Q   K   N   G   I

6572 AAG GTG AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC
     TTC CAC TTG AAG TTC TAG GCG GTG TTG TAG CTC CTG CCG TCG CAC GTC GAG CGG CTG
1042▸ K   V   N   F   K   I   R   H   N   I   E   D   G   S   V   Q   L   A   D

6629 CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC
     GTG ATG GTC GTC TTG TGG GGG TAG CCG CTG CCG GGG CAC GAC GAC GGG CTG TTG GTG
1061▸ H   Y   Q   Q   N   T   P   I   G   D   G   P   V   L   L   P   D   N   H

6686 TAC CTG AGC ACC CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG
     ATG GAC TCG TGG GTC AGG CGG GAC TCG TTT CTG GGG TTG CTC TTC GCG CTA GTG TAC
1080▸ Y   L   S   T   Q   S   A   L   S   K   D   P   N   E   K   R   D   H   M

6743 GTC CTG CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC
     CAG GAC GAC CTC AAG CAC TGG CGG CGG CCC TAG TGA GAG CCG TAC CTG CTC GAC ATG
1099▸ V   L   L   E   F   V   T   A   A   G   I   T   L   G   M   D   E   L   Y

                                     SV40 Poly A
6800 AAG TAA AGC GGC CGC GAC TCT AGA G GATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGG
     TTC ATT TCG CCG GCG CTG AGA TCT C CTAGAAACACTTCCTTGGAATGAAGACACCACACTGTATTAACC
1118▸ K   •

6869 ACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTC
     TGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCACATATTACACAATTTGATGACTAAG

6947 TAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAA
     ATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTT

7025 ACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAA
     TGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTT

7103 AGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATA
     TCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTAT

7181 GAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATT
     CTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAA
```

```
7259 CTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACATACTGTTTTTCTTACTCCACACAGGCATAGAGTGT
     GACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAAAGAATGAGGTGTGTCCGTATCTCACA

7337 CTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGA
     GACGATAATTATTGATACGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACT

7415 TGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCA
     ACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGT

7493 CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTAC
     GTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATG

7571 AAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTC
     TTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAG

                                    white gene
7649 ATCAATGTATCTTATCATGTCTGGATCGGATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTT
     TAGTTACATAGAATAGTACAGACCTAGCCTAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAA

7727 GGAGAATGTAGATTTAAAAAAACATATTTTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGT
     CCTCTTACATCTAAATTTTTTTGTATAAAAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCA

7805 TTTAAATTTACTTCGATCCAAGGGTATTTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACT
     AAATTTAAATGAAGCTAGGTTCCCATAAACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGA

7883 CAAAGTCAGCGCTGTTTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTT
     GTTTCAGTCGCGACAAACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAA

7961 TCGCCACCGTTTGTAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTC
     AGCGGTGGCAAACATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAG

8039 CAAGCGGTTTACGCCATCAATTAAACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTG
     GTTCGCCAAATGCGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAAC

8117 AGTGATTGGGGTGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGAT
     TCACTAACCCCACCACTAACCAAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTA

8195 CAGGAGCTATTAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATAC
     GTCCTCGATAATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATG

8273 ATCTTAAGTTCACTTGATCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAA
     TAGAATTCAAGTGAACTAGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTT

8351 AATCCCAAACCGCAATCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACA
     TTAGGGTTTGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGT

8429 ATTTTTAAGATCATATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAAC
     TAAAAATTCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTG

8507 AACCAGATATTTTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTT
     TTGGTCTATAAAATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAA
```

```
8585 TGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGC
     ACAAGCCCCAATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCG

8663 TTCGCAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGA
     AAGCGTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCT

8741 CTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAA
     GAGGCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTT

8819 TCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCC
     AGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGG

8897 CAGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAA
     GTCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTT

8975 TTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGGA
     AATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCCT

9053 AAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTC
     TTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACGAG

9131 AATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGCTCC
     TTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCGAGG

9209 CTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGTGGCC
     GATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCACCGG

9287 CGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGGTGAAA
     GCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCCACTTT

9365 GGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATCTGCGAT
     CCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAGACGCTA

9443 GAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAAGGGCAAG
     CTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTTCCCGTTC

9521 ACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGGCCGAGGGC
     TGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTACCGGCTCCCG

9599 AGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTTATTAAGGGTAT
     TCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAATAATTCCCATA

9677 CTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGGCGGACTTTTA
     GATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAGGCCGCCTGAAAAT

9755 CGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGATATGCGACAATTTTGC
     GCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTATACGCTGTTAAAACG

9833 TATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAA
     ATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTT
```

```
 9911 TGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGGA
      ACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCCT

 9989 ACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTACAA
      TGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATGTT

10067 TTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGATGT
      AAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCTACA

10145 ATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAGAT
      TATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTACTCTA

10223 GCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCAAGTGG
      CGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACC

10301 GCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATG
      CGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTAC

10379 TAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACC
      ATTCAGAACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGG

10457 TCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATT
      AGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAA

10535 GCCGAATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGA
      CGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCT

10613 GTGCTGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCC
      CACGACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGG

10691 TGCGCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTC
      ACGCGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAG

10769 TTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGT
      AAGAACTTGAGCCCGAGCCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCA

10847 CTGCTGATTAACCAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCG
      GACGACTAATTGGTTACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGC

10925 GGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATC
      CCGTTCCAGTAGGACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAG

11003 GTGAGCTTCCGGGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAAT
      CACTCGAAGGCCCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTA

11081 AACTGCTTGTTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTAT
      TTGACGAACAAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATA

11159 ATTTGTGCCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTG
      TAAACACGGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAAC
```

```
11237 GATTTGTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCG
      CTAAACAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGC

11315 GGATATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAG
      CCTATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATC

11393 TTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTA
      AACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAAT

11471 GTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGT
      CAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACA

11549 ATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACT
      TACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGA

11627 TTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGTT
      AAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACAA

11705 CCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAACTATAATAA
      GGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTTGATATTATT

11783 TAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCA
      ATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGT

11861 TTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAGTC
      AAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCAG

11939 ACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCTTCT
      TGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGAAGA

12017 TTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATGTCTGA
      AAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTACAGACT

12095 CCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAGGCTGCAC
      GGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTCCGACGTG

12173 CCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTCACAGCTTTG
      GGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAGTGTCGAAAC

12251 CAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCACAC
      GTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAAGCAGGCGTGTG

12329 AACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGACGGGACCAC
      TTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGATAGCTGCCCTGGTG

                        5' P
12407 CTTATGTTATTTCATCATG
      GAATACAATAAAGTAGTAC
```

pUAST-sDER-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTGG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGACC

  78 GGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCTCG
     CCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGAGC

 155 TTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAAGGC
     AACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTTCCG

 232 GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
     CAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTC

 309 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
     CTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCT

 386 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
     ATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACAT

 463 GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
     CCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAG

 540 GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
     CAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCC

 617 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
     ATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGAC

 694 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
     GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGC

 771 GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
     CACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGC

 848 GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
     CCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGAT

 925 GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
     CTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTA

1002 GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
     CGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATC

1079 ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
     TATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

1156 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
     TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGG

1233 AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
     TCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGA

1310 ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
     TGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATG

1387 ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
     TACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTC

1464 TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
     ACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGA

1541 GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGA
     CCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCT

1618 TAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
     ATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCT

1695 TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
     AGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGG

1772 AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
     TCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTA
```

```
1849 ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
     TGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTA

1926 GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
     CATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAA

2003 ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
     TAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACT

2080 AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
     TTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAG

2157 AGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
     TCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCAC

2234 CACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGG
     GTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCC

2311 TGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGA
     ACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACT

2388 AGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGA
     TCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACT

2465 TGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGA
     ACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCT

2542 AATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATA
     TTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTAT

2619 AACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCG
     TTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGC

                                                                     3'P
2696 TCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATA
     AGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTAT

2773 ACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATT
     TGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAA

2850 CTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAAT
     GACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTTA

2927 TAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGC
     ATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACG

3004 ATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG
     TACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTC

                                                                   UAS sites
3081 TACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGT
     ATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCA

3158 ATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCT
     TATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGA

3235 AAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGT
     TTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCA

3312 AACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCT
     TTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGA
```

```
                 Polylinker           attB1  (hybrid)                 was G
3389 GAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGTACAAAAAAGTTGGAAA ATG CTG CTa CGA
     CTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAAACATGTTTTTTCAACCTTT TAC GAC GAt GCT

                                                              1▶ M   L   L   R
3461 CGG CGC AAC GGC CCC TGC CCC TTC CCC CTG CTG CTT CTG CTC CTG GCC CAC TGC ATT
     GCC GCG TTG CCG GGG ACG GGG AAG GGG GAC GAC GAA GAC GAG GAC CGG GTG ACG TAA

   5▶ R   R   N   G   P   C   P   F   P   L   L   L   L   L   L   A   H   C   I
3518 TGC ATT TGG CCC GCG TCG GCG GCC CGC GAT CGC TAC GCC CGC CAG AAC AAT CGC CAG
     ACG TAA ACC GGG CGC AGC CGC CGG GCG CTA GCG ATG CGG GCG GTC TTG TTA GCG GTC

  24▶ C   I   W   P   A   S   A   A   R   D   R   Y   A   R   Q   N   N   R   Q
3575 CGC CAT CAG GAT ATA GAT CGC GAT CGG GAT CGA GAT CGA TTC CTA TAC CGC AGC AGT
     GCG GTA GTC CTA TAT CTA GCG CTA GCC CTA GCT CTA GCT AAG GAT ATG GCG TCG TCA

  43▶ R   H   Q   D   I   D   R   D   R   D   R   D   R   F   L   Y   R   S   S
3632 TCG GCC CAA AAT CGA CAG AGG GGC GGG GCC AAC TTC GCC CTG GGA CTG GGA GCC AAC
     AGC CGG GTT TTA GCT GTC TCC CCG CCC CGG TTG AAG CGG GAC CCT GAC CCT CGG TTG

  62▶ S   A   Q   N   R   Q   R   G   G   A   N   F   A   L   G   L   G   A   N
3689 GGA GTC ACC ATT CCC ACC AGT CTG GAG GAT AAG AAC AAG AAC GAG TTC GTC AAG GGG
     CCT CAG TGG TAA GGG TGG TCA GAC CTC CTA TTC TTG TTC TTG CTC AAG CAG TTC CCC

  81▶ G   V   T   I   P   T   S   L   E   D   K   N   K   N   E   F   V   K   G
3746 AAA ATC TGC ATC GGC ACT AAA TCT CGG CTC TCC GTG CCC TCC AAC AAG GAA CAT CAT
     TTT TAG ACG TAG CCG TGA TTT AGA GCC GAG AGG CAC GGG AGG TTG TTC CTT GTA GTA

 100▶ K   I   C   I   G   T   K   S   R   L   S   V   P   S   N   K   E   H   H
3803 TAC CGA AAC CTC AGA GAT CGG TAC ACG AAC TGT ACG TAT GTG GAT GGC AAC TTG AAA
     ATG GCT TTG GAG TCT CTA GCC ATG TGC TTG ACA TGC ATA CAC CTA CCG TTG AAC TTT

 119▶ Y   R   N   L   R   D   R   Y   T   N   C   T   Y   V   D   G   N   L   K
3860 CTG ACC TGG CTA CCC AAC GAG AAT TTG GAC CTC AGC TTC CTA GAC AAC ATA CGG GAG
     GAC TGG ACC GAT GGG TTG CTC TTA AAC CTG GAG TCG AAG GAT CTG TTG TAT GCC CTC

 138▶ L   T   W   L   P   N   E   N   L   D   L   S   F   L   D   N   I   R   E
3917 GTC ACC GGC TAT ATT CTG ATC AGT CAT GTG GAC GTT AAG AAA GTG GTG TTT CCC AAA
     CAG TGG CCG ATA TAA GAC TAG TCA GTA CAC CTG CAA TTC TTT CAC CAC AAA GGG TTT

 157▶ V   T   G   Y   I   L   I   S   H   V   D   V   K   K   V   V   F   P   K
3974 CTA CAA ATC ATT CGC GGA CGC ACG CTG TTC AGC TTA TCC GTG GAG GAG GAG AAG TAT
     GAT GTT TAG TAA GCG CCT GCG TGC GAC AAG TCG AAT AGG CAC CTC CTC CTC TTC ATA

 176▶ L   Q   I   I   R   G   R   T   L   F   S   L   S   V   E   E   E   K   Y
4031 GCC TTG TTC GTC ACT TAT TCC AAA ATG TAC ACG CTG GAG ATT CCC GAT CTA CGC GAT
     CGG AAC AAG CAG TGA ATA AGG TTT TAC ATG TGC GAC CTC TAA GGG CTA GAT GCG CTA

 195▶ A   L   F   V   T   Y   S   K   M   Y   T   L   E   I   P   D   L   R   D
4088 GTC TTA AAT GGC CAA GTG GGC TTC CAC AAC AAC TAC AAT CTC TGC CAC ATG CGA ACG
     CAG AAT TTA CCG GTT CAC CCG AAG GTG TTG TTG ATG TTA GAG ACG GTG TAC GCT TGC

 214▶ V   L   N   G   Q   V   G   F   H   N   N   Y   N   L   C   H   M   R   T
```

```
4145 ATC CAG TGG TCG GAG ATT GTA TCC AAC GGC ACG GAT GCA TAC TAC AAC TAC GAC TTT
     TAG GTC ACC AGC CTC TAA CAT AGG TTG CCG TGC CTA CGT ATG ATG TTG ATG CTG AAA

 233▶ I   Q   W   S   E   I   V   S   N   G   T   D   A   Y   Y   N   Y   D   F

4202 ACT GCT CCG GAG CGC GAG TGT CCC AAG TGC CAC GAG AGC TGC ACG CAC GGA TGT TGG
     TGA CGA GGC CTC GCG CTC ACA GGG TTC ACG GTG CTC TCG ACG TGC GTG CCT ACA ACC

 252▶ T   A   P   E   R   E   C   P   K   C   H   E   S   C   T   H   G   C   W

4259 GGC GAG GGT CCC AAG AAT TGC CAG AAG TTC AGC AAG CTC ACC TGC TCG CCA CAG TGT
     CCG CTC CCA GGG TTC TTA ACG GTC TTC AAG TCG TTC GAG TGG ACG AGC GGT GTC ACA

 271▶ G   E   G   P   K   N   C   Q   K   F   S   K   L   T   C   S   P   Q   C

4316 GCC GGA GGT CGT TGC TAT GGA CCA AAG CCG CGG GAG TGT TGT CAC CTC TTC TGC GCC
     CGG CCT CCA GCA ACG ATA CCT GGT TTC GGC GCC CTC ACA ACA GTG GAG AAG ACG CGG

 290▶ A   G   G   R   C   Y   G   P   K   P   R   E   C   C   H   L   F   C   A

4373 GGA GGA TGC ACT GGT CCC ACG CAA AAG GAT TGC ATC GCC TGC AAG AAC TTC TTC GAC
     CCT CCT ACG TGA CCA GGG TGC GTT TTC CTA ACG TAG CGG ACG TTC TTG AAG AAG CTG

 309▶ G   G   C   T   G   P   T   Q   K   D   C   I   A   C   K   N   F   F   D

4430 GAG GCA GTA TCA AAG GAG GAA TGC CCG CCC ATG CGC AAG TAC AAT CCC ACC ACC TAT
     CTC CGT CAT AGT TTC CTC CTT ACG GGC GGG TAC GCG TTC ATG TTA GGG TGG TGG ATA

 328▶ E   A   V   S   K   E   E   C   P   P   M   R   K   Y   N   P   T   T   Y

4487 GTT CTT GAA ACG AAT CCT GAG GGA AAG TAT GCC TAT GGT GCC ACC TGC GTC AAG GAG
     CAA GAA CTT TGC TTA GGA CTC CCT TTC ATA CGG ATA CCA CGG TGG ACG CAG TTC CTC

 347▶ V   L   E   T   N   P   E   G   K   Y   A   Y   G   A   T   C   V   K   E

4544 TGT CCC GGT CAT CTG TTG CGG GAT AAT GGC GCC TGC GTG CGC AGC TGT CCC CAG GAC
     ACA GGG CCA GTA GAC AAC GCC CTA TTA CCG CGG ACG CAC GCG TCG ACA GGG GTC CTG

 366▶ C   P   G   H   L   L   R   D   N   G   A   C   V   R   S   C   P   Q   D

4601 AAG ATG GAC AAG GGG GGC GAG TGT GTG CCC TGC AAT GGA CCG TGC CCC AAA ACC TGC
     TTC TAC CTG TTC CCC CCG CTC ACA CAC GGG ACG TTA CCT GGC ACG GGG TTT TGG ACG

 385▶ K   M   D   K   G   G   E   C   V   P   C   N   G   P   C   P   K   T   C

4658 CCG GGC GTT ACT GTC CTG CAT GCC GGC AAC ATT GAC TCG TTC CGG AAT TGT ACG GTG
     GGC CCG CAA TGA CAG GAC GTA CGG CCG TTG TAA CTG AGC AAG GCC TTA ACA TGC CAC

 404▶ P   G   V   T   V   L   H   A   G   N   I   D   S   F   R   N   C   T   V

4715 ATC GAT GGC AAC ATT CGC ATT TTG GAT CAG ACC TTC TCG GGC TTC CAG GAT GTC TAT
     TAG CTA CCG TTG TAA GCG TAA AAC CTA GTC TGG AAG AGC CCG AAG GTC CTA CAG ATA

 423▶ I   D   G   N   I   R   I   L   D   Q   T   F   S   G   F   Q   D   V   Y

4772 GCC AAC TAC ACG ATG GGA CCA CGA TAC ATA CCG CTG GAT CCC GAG CGA CGG GAG GTG
     CGG TTG ATG TGC TAC CCT GGT GCT ATG TAT GGC GAC CTA GGG CTC GCT GCC CTC CAC

 442▶ A   N   Y   T   M   G   P   R   Y   I   P   L   D   P   E   R   R   E   V

4829 TTC TCC ACG GTG AAG GAG ATC ACC GGG TAT CTG AAT ATC GAG GGA ACC CAC CCG CAG
     AAG AGG TGC CAC TTC CTC TAG TGG CCC ATA GAC TTA TAG CTC CCT TGG GTG GGC GTC

 461▶ F   S   T   V   K   E   I   T   G   Y   L   N   I   E   G   T   H   P   Q
```

```
4886 TTC CGG AAT CTG TCG TAC TTT CGC AAT CTG GAA ACA ATT CAT GGC CGC CAG CTG ATG
     AAG GCC TTA GAC AGC ATG AAA GCG TTA GAC CTT TGT TAA GTA CCG GCG GTC GAC TAC

480▶  F   R   N   L   S   Y   F   R   N   L   E   T   I   H   G   R   Q   L   M
4943 GAG AGC ATG TTT GCC GCT TTG GCG ATC GTT AAG TCA TCC CTG TAC AGC CTG GAG ATG
     CTC TCG TAC AAA CGG CGA AAC CGC TAG CAA TTC AGT AGG GAC ATG TCG GAC CTC TAC

499▶  E   S   M   F   A   A   L   A   I   V   K   S   S   L   Y   S   L   E   M
5000 CGC AAT CTG AAG CAG ATT AGT TCC GGC AGT GTG GTC ATC CAG CAT AAT AGA GAC CTC
     GCG TTA GAC TTC GTC TAA TCA AGG CCG TCA CAC CAG TAG GTC GTA TTA TCT CTG GAG

518▶  R   N   L   K   Q   I   S   S   G   S   V   V   I   Q   H   N   R   D   L
5057 TGC TAC GTA AGC AAT ATC CGT TGG CCG GCC ATT CAG AAG GAG CCC GAA CAG AAG GTG
     ACG ATG CAT TCG TTA TAG GCA ACC GGC CGG TAA GTC TTC CTC GGG CTT GTC TTC CAC

537▶  C   Y   V   S   N   I   R   W   P   A   I   Q   K   E   P   E   Q   K   V
5114 TGG GTC AAC GAG AAT CTC AGG GCG GAT CTA TGC GAG AAA AAT GGA ACC ATT TGC TCG
     ACC CAG TTG CTC TTA GAG TCC CGC CTA GAT ACG CTC TTT TTA CCT TGG TAA ACG AGC

556▶  W   V   N   E   N   L   R   A   D   L   C   E   K   N   G   T   I   C   S
5171 GAT CAG TGC AAC GAG GAC GGC TGC TGG GGA GCT GGC ACG GAT CAG TGC CTT ACC TGC
     CTA GTC ACG TTG CTC CTG CCG ACG ACC CCT CGA CCG TGC CTA GTC ACG GAA TGG ACG

575▶  D   Q   C   N   E   D   G   C   W   G   A   G   T   D   Q   C   L   T   C
5228 AAG AAC TTC AAT TTC AAT GGC ACC TGC ATC GCC GAC TGT GGT TAT ATA TCC AAT GCC
     TTC TTG AAG TTA AAG TTA CCG TGG ACG TAG CGG CTG ACA CCA ATA TAT AGG TTA CGG

594▶  K   N   F   N   F   N   G   T   C   I   A   D   C   G   Y   I   S   N   A
5285 TAC AAG TTT GAC AAT AGA ACG TGC AAG ATA TGC CAT CCA GAG TGC CGG ACT TGC AAT
     ATG TTC AAA CTG TTA TCT TGC ACG TTC TAT ACG GTA GGT CTC ACG GCC TGA ACG TTA

613▶  Y   K   F   D   N   R   T   C   K   I   C   H   P   E   C   R   T   C   N
5342 GGA GCT GGA GCA GAT CAC TGC CAG GAG TGC GTC CAT GTG AGG GAC GGT CAG CAC TGT
     CCT CGA CCT CGT CTA GTG ACG GTC CTC ACG CAG GTA CAC TCC CTG CCA GTC GTG ACA

632▶  G   A   G   A   D   H   C   Q   E   C   V   H   V   R   D   G   Q   H   C
5399 GTG TCC GAG TGC CCG AAG AAC AAG TAC AAC GAT CGT GGT GTC TGC CGA GAG TGC CAC
     CAC AGG CTC ACG GGC TTC TTG TTC ATG TTG CTA GCA CCA CAG ACG GCT CTC ACG GTG

651▶  V   S   E   C   P   K   N   K   Y   N   D   R   G   V   C   R   E   C   H
5456 GCC ACC TGC GAT GGA TGC ACT GGG CCC AAG GAC ACC ATC GGC ATT GGA GCG TGT ACG
     CGG TGG ACG CTA CCT ACG TGA CCC GGG TTC CTG TGG TAG CCG TAA CCT CGC ACA TGC

670▶  A   T   C   D   G   C   T   G   P   K   D   T   I   G   I   G   A   C   T
5513 ACG TGC AAT TTG GCC ATT ATC AAC AAT GAC GCC ACA GTA AAA CGC TGC CTG CTG AAG
     TGC ACG TTA AAC CGG TAA TAG TTG TTA CTG CGG TGT CAT TTT GCG ACG GAC GAC TTC

689▶  T   C   N   L   A   I   I   N   N   D   A   T   V   K   R   C   L   L   K
5570 GAC GAC AAG TGC CCC GAT GGG TAT TTC TGG GAG TAT GTG CAT CCG CAA GAG CAG GGA
     CTG CTG TTC ACG GGG CTA CCC ATA AAG ACC CTC ATA CAC GTA GGC GTT CTC GTC CCT

708▶  D   D   K   C   P   D   G   Y   F   W   E   Y   V   H   P   Q   E   Q   G
```

```
5627 TCG CTG AAG CCA TTG GCC GGC AGA GCA GTT TGC CGA AAG TGC CAT CCC CTT TGC GAG
     AGC GAC TTC GGT AAC CGG CCG TCT CGT CAA ACG GCT TTC ACG GTA GGG GAA ACG CTC

 727  S   L   K   P   L   A   G   R   A   V   C   R   K   C   H   P   L   C   E
5684 CTG TGC ACG AAC TAC GGA TAC CAT GAA CAG GTG TGC TCC AAG TGC ACC CAC TAC AAG
     GAC ACG TGC TTG ATG CCT ATG GTA CTT GTC CAC ACG AGG TTC ACG TGG GTG ATG TTC

 746  L   C   T   N   Y   G   Y   H   E   Q   V   C   S   K   C   T   H   Y   K
5741 CGA CGG GAG CAG TGC GAG ACC GAG TGT CCG GCC GAT CAC TAC ACG GAT GAG GAG CAG
     GCT GCC CTC GTC ACG CTC TGG CTC ACA GGC CGG CTA GTG ATG TGC CTA CTC CTC GTC

 765  R   R   E   Q   C   E   T   E   C   P   A   D   H   Y   T   D   E   E   Q
5798 CGC GAG TGC TTC CAG CGC CAC CCG GAA TGC AAT GGT TGC ACG GGT CCG GGT GCC GAC
     GCG CTC ACG AAG GTC GCG GTG GGC CTT ACG TTA CCA ACG TGC CCA GGC CCA CGG CTG

 784  R   E   C   F   Q   R   H   P   E   C   N   G   C   T   G   P   G   A   D
5855 GAT TGC AAG TCT TGC CGC AAC TTT AAG TTG TTC GAC GCG AAT GAG ACG GGT CCC TAT
     CTA ACG TTC AGA ACG GCG TTG AAA TTC AAC AAG CTG CGC TTA CTC TGC CCA GGG ATA

 803  D   C   K   S   C   R   N   F   K   L   F   D   A   N   E   T   G   P   Y
5912 GTG AAC TCC ACG ATG TTC AAT TGC ACC TCG AAG TGT CCC TTG GAG ATG CGA CAT GTG
     CAC TTG AGG TGC TAC AAG TTA ACG TGG AGC TTC ACA GGG AAC CTC TAC GCT GTA CAC

 822  V   N   S   T   M   F   N   C   T   S   K   C   P   L   E   M   R   H   V
5969 AAC TAT CAG TAC ACG GCC ATT GGA CCC TAC TGC GCA GCT AGT CCG CCG AGG AGC AGC
     TTG ATA GTC ATG TGC CGG TAA CCT GGG ATG ACG CGT CGA TCA GGC GGC TCC TCG TCG

 841  N   Y   Q   Y   T   A   I   G   P   Y   C   A   A   S   P   P   R   S   S

                                                V5  epitope
                                     attB2  (hybrid)

6026 AAG CCA ACT TTC TTG TAC AAA GTG GTG GTA CCG  GGT AAG CCT ATC CCT AAC CCT CTC

     TTC GGT TGA AAG AAC ATG TTT CAC CAC CAT GGC  CCA TTC GGA TAG GGA TTG GGA GAG

 860  K   P   T   F   L   Y   K   V   V   V   P    G   K   P   I   P   N   P   L

                                        6xHis                          SV40  Poly
6083 CTC GGT CTC GAT TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA TCTAGAGGATCTT
     GAG CCA GAG CTA AGA TGC GCA TGG CCA GTA GTA GTG GTA GTG GTA ACT AGATCTCCTAGAA

 879  L   G   L   D   S   T   R   T   G   H   H   H   H   H   H   •

6144 TGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATA
     ACACTTCCTTGGAATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTAT

6221 TAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACT
     ATTTTAAAAATTCACATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGA

6298 GATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGA
     CTACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACT

6375 TGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTT
     ACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAA
```

```
6452 CAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAG
     GTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTC

6529 GAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAA
     CTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATT

6606 TCATAACATACTGTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTA
     AGTATTGTATGACAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACAT

6683 CCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAG
     GGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTC

6760 CCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGA
     GGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACT

6837 ATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACA
     TACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGT

6914 AATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGG
     TTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCC

         white gene
6991 ATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATT
     TAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAA

7068 TTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTAT
     AAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATA

7145 TTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTC
     AACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAG

7222 TCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTA
     AGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGAT

7299 GCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTAA
     CGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAATT

7376 ACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGGTT
     TGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACCAA

7453 TTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCGGAG
     AACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGCCTC

7530 GCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGATCTA
     CGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACTAGAT

7607 TAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAATCGCA
     ATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTTAGCGT

7684 CAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCATATCAT
     GTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGTATAGTA
```

```
7761 GATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTACAAAAAATATAACAACCAGATATTTTAAGCT
     CTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAAAATTCGA

7838 GATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGGTTAGATGA
     CTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCCAATCTACT

7915 GCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAGAGCTGCATT
     CGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTCTCGACGTAA

7992 AACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCGGTTCAGGGAG
     TTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGCCAAGTCCCTC

8069 CGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGGGCTCCG
     GCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGTCGGCCCGAGGC

8146 GATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCCAGGAAACATTTG
     CTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGGTCCTTTGTAAAC

8223 CTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTGA
     GAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTTAATGGTTAAACT

8300 AACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGGAAAGACGACCCT
     TTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCCTTTCTGCTGGGA

8377 GCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTCAATGGCCAAC
     CGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACGAGTTACCGGTTG

8454 CTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGCTCCCTAACGGCC
     GACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCGAGGGATTGCCGG

8531 AGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGTGGCCCGCGTGGA
     TCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCACCGGGCGCACCT

8608 TCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGGTGAAAGGTCTGT
     AGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCCACTTTCCAGACA

8685 CCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATCTGCGATGAGCCC
     GGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAGACGCTACTCGGG

8762 ACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAAGGGCAAGACCGT
     TGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTTCCCGTTCTGGCA

8839 CATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGGCCGAGGGCAGGG
     GTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTACCGGCTCCCGTCCC

8916 TAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTTATTAAGGGTATCTA
     ATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAATAATTCCCATAGAT

8993 GCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGGCGGACTTTTACG
     CGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAGGCCGCCTGAAAATGC
```

```
 9070 TACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGATATGCGACAATTTTGCT
      ATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTATACGCTGTTAAAACGA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9147 ATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAA
      TAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9224 TGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGG
      ACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9301 AACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTAC
      TTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9378 AATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGA
      TTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9455 TGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATG
      ACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9532 AGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCA
      TCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9609 AGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGA
      TCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9686 TAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGT
      ATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9763 GTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCA
      CAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9840 AAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTG
      TTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9917 CGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATA
      GCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTAT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 9994 TCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCT
      AGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10071 TTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTAC
      AACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10148 GCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCAC
      CGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10225 GTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTC
      CACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
10302 TGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCG
      ACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
10379 ACATATATCCGAAATAACTGCTTGTTTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTATTGCCCCCTCAAA
      TGTATATAGGCTTTATTGACGAACAAAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAATAACGGGGGAGTTT

10456 AAGCTAATGTAATTATATTTGTGCCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATC
      TTCGATTACATTAATATAAACACGGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAG

10533 CCCACAAGTAGACTTTGGATTTGTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTA
      GGGTGTTCATCTGAAACCTAAACAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAAT

10610 TCGCTACATAAAACACCGGGATATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCC
      AGCGATGTATTTTGTGGCCCTATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGG

10687 ACCACACCACGTTTCGTAGTTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGA
      TGGTGTGGTGCAAAGCATCAACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCT

10764 CCTTGGAGCGACTGTCGTTAGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAG
      GGAACCTCGCTGACAGCAATCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATC

10841 AAATTAGTAATAAATATTTGTATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTT
      TTTAATCATTATTTATAAACATACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAA

10918 ATTCTAATTTAATATTATGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACT
      TAAGATTAAATTATAATACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGA

10995 GAAAGTGACATCCAGTGTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATA
      CTTTCACTGTAGGTCACAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATAT

11072 TATCAAAAATAACAACTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAA
      ATAGTTTTTATTGTTGATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATT

11149 TTCGGCTGCTGCTCTAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAA
      AAGCCGACGACGAGATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTT

11226 CAATGAACAGGACCTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTT
      GTTACTTGTCCTGGATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAA

11303 TGTGTACTCCCACTGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGC
      ACACATGAGGGTGACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACG

11380 GTTGAGTCGCCTCCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCAC
      CAACTCAGCGGAGGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTG

11457 AAATTTGGGAGTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACG
      TTTAAACCCTCAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGC

11534 TGAATTAATTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTTATTCCACGTAAGGGTT
      ACTTAATTAAAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAAATAAGGTGCATTCCCAA

11611 AATGTTTTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATA
      TTACAAAAGTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATAT
```

```
                                                         5' P
11688 CTTCGGTAAGCTTCGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      GAAGCCATTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
```

pUAST-sKek1-GFP

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTGG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGACC

  78 GGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCTCG
     CCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGAGC

 155 TTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAAGGC
     AACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTTCCG

 232 GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
     CAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTC

 309 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
     CTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCT

 386 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
     ATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACAT

 463 GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
     CCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAG

 540 GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
     CAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCC

 617 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
     ATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGAC

 694 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
     GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGC

 771 GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
     CACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGC

 848 GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
     CCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGAT

 925 GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
     CTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTA

1002 GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
     CGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATC

1079 ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
     TATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

1156 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
     TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGG

1233 AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
     TCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGA

1310 ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
     TGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATG

1387 ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
     TACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTC

1464 TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
     ACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGA

1541 GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGA
     CCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCT

1618 TAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
     ATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCT

1695 TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
     AGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGG

1772 AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
     TCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTA
```

```
1849 ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
     TGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTA

1926 GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
     CATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAA

2003 ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
     TAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACT

2080 AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
     TTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAG

2157 AGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
     TCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCAC

2234 CACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGG
     GTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCC

2311 TGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGA
     ACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACT

2388 AGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGA
     TCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACT

2465 TGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGA
     ACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCT

2542 AATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATA
     TTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTAT

2619 AACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCG
     TTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGC

                                                                      3'P
2696 TCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATA
     AGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTAT

2773 ACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATT
     TGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAA

2850 CTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAAT
     GACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTTA

2927 TAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGC
     ATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACG

3004 ATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG
     TACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTC

                                                                   UAS sites
3081 TACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGT
     ATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCA

3158 ATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCT
     TATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGA

3235 AAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGT
     TTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCA

3312 AACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCT
     TTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGA
```

```
                  Polylinker              attB1  (hybrid)
3389 GAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGTACAAAAAAGTTGGAAA ATG CAT ATC AGG
     CTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAACATGTTTTTTCAACCTTT TAC GTA TAG TCC
                                                                1▸ M   H   I   R
--------------------------------------------------------------------------------
3461 GAA GCA GTT TTC CTG GTC CTC ACC CTG CTG CCT GGA ATG ATC CTG GGC ACT CGC TAC
     CTT CGT CAA AAG GAC CAG GAG TGG GAC GAC GGA CCT TAC TAG GAC CCG TGA GCG ATG
   5▸ E   A   V   F   L   V   L   T   L   L   P   G   M   I   L   G   T   R   Y
--------------------------------------------------------------------------------
3518 AAT CAG CTG CAT CTG TAT GCC AAT GGA GGA GCA TCG TCA TCG GGC CCT GGA GGC TAC
     TTA GTC GAC GTA GAC ATA CGG TTA CCT CCT CGT AGC AGT AGC CCG GGA CCT CCG ATG
  24▸ N   Q   L   H   L   Y   A   N   G   G   A   S   S   S   G   P   G   G   Y
--------------------------------------------------------------------------------
3575 AGG CCC GCC CCC TCG TCC CAG AAC GAG GTG TAC TCC ATA GCG GAC AGC CAG CCG ATG
     TCC GGG CGG GGG AGC AGG GTC TTG CTC CAC ATG AGG TAT CGC CTG TCG GTC GGC TAC
  43▸ R   P   A   P   S   S   Q   N   E   V   Y   S   I   A   D   S   Q   P   M
--------------------------------------------------------------------------------
3632 ACT GAG GAT GGC TAC ATG CCC CCC AGC CAG CAC TTT CCG CCC ACC CAC TCC GAC TTG
     TGA CTC CTA CCG ATG TAC GGG GGG TCG GTC GTG AAA GGC GGG TGG GTG AGG CTG AAC
  62▸ T   E   D   G   Y   M   P   P   S   Q   H   F   P   P   T   H   S   D   L
--------------------------------------------------------------------------------
3689 GAT CCC CCC GCC CAG CAG CAG AGC ACC TGC CAA ACG GTT TGC GCC TGC AAG TGG AAG
     CTA GGG GGG CGG GTC GTC GTC TCG TGG ACG GTT TGC CAA ACG CGG ACG TTC ACC TTC
  81▸ D   P   P   A   Q   Q   Q   S   T   C   Q   T   V   C   A   C   K   W   K
--------------------------------------------------------------------------------
3746 GGT GGC AAG CAG ACG GTG GAG TGC ATC GAT CGC CAC CTC ATC CAG ATA CCC GAG CAC
     CCA CCG TTC GTC TGC CAC CTC ACG TAG CTA GCG GTG GAG TAG GTC TAT GGG CTC GTG
 100▸ G   G   K   Q   T   V   E   C   I   D   R   H   L   I   Q   I   P   E   H
--------------------------------------------------------------------------------
3803 ATC GAT CCC AAT ACC CAG GTG CTG GAC ATG TCC GGT AAT AAG CTG CAG ACC CTC TCC
     TAG CTA GGG TTA TGG GTC CAC GAC CTG TAC AGG CCA TTA TTC GAC GTC TGG GAG AGG
 119▸ I   D   P   N   T   Q   V   L   D   M   S   G   N   K   L   Q   T   L   S
--------------------------------------------------------------------------------
3860 AAC GAG CAG TTC ATC CGT GCG AAT CTG CTA AAT CTG CAG AAG CTG TAT TTG AGG AAC
     TTG CTC GTC AAG TAG GCA CGC TTA GAC GAT TTA GAC GTC TTC GAC ATA AAC TCC TTG
 138▸ N   E   Q   F   I   R   A   N   L   L   N   L   Q   K   L   Y   L   R   N
--------------------------------------------------------------------------------
3917 TGC AAG ATC GGC GAA ATC GAG CGG GAG ACC TTC AAG GGA CTG ACC AAT CTG GTG GAG
     ACG TTC TAG CCG CTT TAG CTC GCC CTC TGG AAG TTC CCT GAC TGG TTA GAC CAC CTC
 157▸ C   K   I   G   E   I   E   R   E   T   F   K   G   L   T   N   L   V   E
--------------------------------------------------------------------------------
3974 TTG GAT CTG TCA CAT AAT CTG CTG GTT ACC GTG CCC AGT TTG GCC CTG GGC CAC ATA
     AAC CTA GAC AGT GTA TTA GAC GAC CAA TGG CAC GGG TCA AAC CGG GAC CCG GTG TAT
 176▸ L   D   L   S   H   N   L   L   V   T   V   P   S   L   A   L   G   H   I
--------------------------------------------------------------------------------
4031 CCC TCA CTG CGC GAA CTC ACC CTG GCC TCC AAT CAC ATA CAC AAA ATC GAG AGC CAG
     GGG AGT GAC GCG CTT GAG TGG GAC CGG AGG TTA GTG TAT GTG TTT TAG CTC TCG GTC
 195▸ P   S   L   R   E   L   T   L   A   S   N   H   I   H   K   I   E   S   Q
--------------------------------------------------------------------------------
4088 GCC TTC GGG AAC ACA CCA TCG CTG CAC AAA TTG GAT CTG TCG CAT TGC GAT ATT CAG
     CGG AAG CCC TTG TGT GGT AGC GAC GTG TTT AAC CTA GAC AGC GTA ACG CTA TAA GTC
 214▸ A   F   G   N   T   P   S   L   H   K   L   D   L   S   H   C   D   I   Q
--------------------------------------------------------------------------------
4145 ACC ATT TCC GCC CAG GCA TTT GGT GGC CTC CAA GGA TTG ACT TTG CTC CGA TTG AAT
     TGG TAA AGG CGG GTC CGT AAA CCA CCG GAG GTT CCT AAC TGA AAC GAG GCT AAC TTA
 233▸ T   I   S   A   Q   A   F   G   G   L   Q   G   L   T   L   L   R   L   N
--------------------------------------------------------------------------------
4202 GGC AAT AAA CTG AGC GAG CTT TTG CCC AAG ACA ATT GAG ACC CTG AGT CGA CTT CAT
     CCG TTA TTT GAC TCG CTC GAA AAC GGG TTC TGT TAA CTC TGG GAC TCA GCT GAA GTA
 252▸ G   N   K   L   S   E   L   L   P   K   T   I   E   T   L   S   R   L   H
--------------------------------------------------------------------------------
4259 GGC ATC GAA CTG CAC GAC AAT CCC TGG CTC TGT GAT TGT CGA TTG AGG GAC ACG AAG
     CCG TAG CTT GAC GTG CTG TTA GGG ACC GAG ACA CTA ACA GCT AAC TCC CTG TGC TTC
 271▸ G   I   E   L   H   D   N   P   W   L   C   D   C   R   L   R   D   T   K
--------------------------------------------------------------------------------
4316 CTC TGG CTG ATG AAG AGG AAC ATA CCC TAT CCG GTG GCT CCG GTT TGC TCG GGT GGC
     GAG ACC GAC TAC TTC TCC TTG TAT GGG ATA GGC CAC CGA GGC CAA ACG AGC CCA CCG
 290▸ L   W   L   M   K   R   N   I   P   Y   P   V   A   P   V   C   S   G   G
--------------------------------------------------------------------------------
```

```
4373 CCC GAA AGG ATT ATC GAT CGC AGC TTT GCG GAT CTG CAT GTG GAT GAG TTT GCC TGC
     GGG CTT TCC TAA TAG CTA GCG TCG AAA CGC CTA GAC GTA CAC CTA CTC AAA CGG ACG
 309▸ P   E   R   I   I   D   R   S   F   A   D   L   H   V   D   E   F   A   C
-------------------------------------------------------------------------------
4430 CGA CCG GAG ATG TTG CCC ATA TCG CAT TAT GTG GAG GCG GCC ATG GGC GAG AAT GCC
     GCT GGC CTC TAC AAC GGG TAT AGC GTA ATA CAC CTC CGC CGG TAC CCG CTC TTA CGG
 328▸ R   P   E   M   L   P   I   S   H   Y   V   E   A   A   M   G   E   N   A
-------------------------------------------------------------------------------
4487 TCG ATT ACA TGT CGA GCT CGA GCG GTT CCA GCT GCG AAT ATC AAC TGG TAC TGG AAC
     AGC TAA TGT ACA GCT CGA GCT CGC CAA GGT CGA CGC TTA TAG TTG ACC ATG ACC TTG
 347▸ S   I   T   C   R   A   R   A   V   P   A   A   N   I   N   W   Y   W   N
-------------------------------------------------------------------------------
4544 GGA CGG CTG CTG GCC AAC AAT TCC GCC TTC ACC GCG TAC CAG AGG ATA CAC ATG TTG
     CCT GCC GAC GAC CGG TTG TTA AGG CGG AAG TGG CGC ATG GTC TCC TAT GTG TAC AAC
 366▸ G   R   L   L   A   N   N   S   A   F   T   A   Y   Q   R   I   H   M   L
-------------------------------------------------------------------------------
4601 GAG CAG GTG GAA GGT GGA TTC GAA AAG CGA TCC AAA CTG GTG CTG ACC AAC GCA CAG
     CTC GTC CAC CTT CCA CCT AAG CTT TTC GCT AGG TTT GAC CAC GAC TGG TTG CGT GTC
 385▸ E   Q   V   E   G   G   F   E   K   R   S   K   L   V   L   T   N   A   Q
-------------------------------------------------------------------------------
4658 GAA ACG GAT TCC AGT GAG TTC TAC TGC GTG GCC GAG AAT CGA GCT GGG ATG GCC GAG
     CTT TGC CTA AGG TCA CTC AAG ATG ACG CAC CGG CTC TTA GCT CGA CCC TAC CGG CTC
 404▸ E   T   D   S   S   E   F   Y   C   V   A   E   N   R   A   G   M   A   E
-------------------------------------------------------------------------------
4715 GCC AAC TTC ACC CTG CAC GTG AGC ATG AGA GCT GCG GGC ATG GCC TCC CTG GGT AGT
     CGG TTG AAG TGG GAC GTG CAC TCG TAC TCT CGA CGC CCG TAC CGG AGG GAC CCA TCA
 423▸ A   N   F   T   L   H   V   S   M   R   A   A   G   M   A   S   L   G   S
-------------------------------------------------------------------------------
                                            attB2  (hybrid)
4772 GGC CAA CCA ACT TTC TTG TAC AAA GTG GTG GTA CCG CGG GCC CGG GAT CCA CCG GTC
     CCG GTT GGT TGA AAG AAC ATG TTT CAC CAC CAT GGC GCC CGG GCC CTA GGT GGC CAG
 442▸ G   Q   P   T   F   L   Y   K   V   V   V   P   R   A   R   D   P   P   V
-------------------------------------------------------------------------------
             EGFP
4829 GCC ACC ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC
     CGG TGG TAC CAC TCG TTC CCG CTC CTC GAC AAG TGG CCC CAC CAC GGG TAG GAC CAG
 461▸ A   T   M   V   S   K   G   E   E   L   F   T   G   V   V   P   I   L   V
-------------------------------------------------------------------------------
4886 GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC GAG GGC
     CTC GAC CTG CCG CTG CAT TTG CCG GTG TTC AAG TCG CAC AGG CCG CTC CCG CTC CCG
 480▸ E   L   D   G   D   V   N   G   H   K   F   S   V   S   G   E   G   E   G
-------------------------------------------------------------------------------
4943 GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC
     CTA CGG TGG ATG CCG TTC GAC TGG GAC TTC AAG TAG ACG TGG TGG CCG TTC GAC GGG
 499▸ D   A   T   Y   G   K   L   T   L   K   F   I   C   T   T   G   K   L   P
-------------------------------------------------------------------------------
5000 GTG CCC TGG CCC ACC CTC GTG ACC ACC CTG ACC TAC GGC GTG CAG TGC TTC AGC CGC
     CAC GGG ACC GGG TGG GAG CAC TGG TGG GAC TGG ATG CCG CAC GTC ACG AAG TCG GCG
 518▸ V   P   W   P   T   L   V   T   T   L   T   Y   G   V   Q   C   F   S   R
-------------------------------------------------------------------------------
5057 TAC CCC GAC CAC ATG AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC
     ATG GGG CTG GTG TAC TTC GTC GTG CTG AAG AAG TTC AGG CGG TAC GGG CTT CCG ATG
 537▸ Y   P   D   H   M   K   Q   H   D   F   F   K   S   A   M   P   E   G   Y
-------------------------------------------------------------------------------
5114 GTC CAG GAG CGC ACC ATC TTC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG
     CAG GTC CTC GCG TGG TAG AAG AAG TTC CTG CTG CCG TTG ATG TTC TGG GCG CGG CTC
 556▸ V   Q   E   R   T   I   F   F   K   D   D   G   N   Y   K   T   R   A   E
-------------------------------------------------------------------------------
5171 GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC
     CAC TTC AAG CTC CCG CTG TGG GAC CAC TTG GCG TAG CTC GAC TTC CCG TAG CTG AAG
 575▸ V   K   F   E   G   D   T   L   V   N   R   I   E   L   K   G   I   D   F
-------------------------------------------------------------------------------
5228 AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC
     TTC CTC CTG CCG TTG TAG GAC CCC GTG TTC GAC CTC ATG TTG ATG TTG TCG GTG TTG
 594▸ K   E   D   G   N   I   L   G   H   K   L   E   Y   N   Y   N   S   H   N
-------------------------------------------------------------------------------
```

```
5285 GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC
     CAG ATA TAG TAC CGG CTG TTC GTC TTC TTG CCG TAG TTC CAC TTG AAG TTC TAG GCG
 613▸ V   Y   I   M   A   D   K   Q   K   N   G   I   K   V   N   F   K   I   R

5342 CAC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC CCC
     GTG TTG TAG CTC CTG CCG TCG CAC GTC GAG CGG CTG GTG ATG GTC GTC TTG TGG GGG
 632▸ H   N   I   E   D   G   S   V   Q   L   A   D   H   Y   Q   Q   N   T   P

5399 ATC GGC GAC GGC CCC GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC CAG TCC GCC
     TAG CCG CTG CCG GGG CAC GAC GAC GGG CTG TTG GTG ATG GAC TCG TGG GTC AGG CGG
 651▸ I   G   D   G   P   V   L   L   P   D   N   H   Y   L   S   T   Q   S   A

5456 CTG AGC AAA GAC CCC AAC GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG ACC
     GAC TCG TTT CTG GGG TTG CTC TTC GCG CTA GTG TAC CAG GAC GAC CTC AAG CAC TGG
 670▸ L   S   K   D   P   N   E   K   R   D   H   M   V   L   L   E   F   V   T

5513 GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG TAA AGC GGC CGC GAC TCT
     CGG CGG CCC TAG TGA GAG CCG TAC CTG CTC GAC ATG TTC ATT TCG CCG GCG CTG AGA
 689▸ A   A   G   I   T   L   G   M   D   E   L   Y   K   •

       SV40 Poly A
5570 AGA G GATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCT
     TCT C CTAGAAACACTTCCTTGGAATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGA

5645 CTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCA
     GATTCCATTTATATTTTAAAAATTCACATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGT

5722 ACCTATGGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCC
     TGGATACCTTGACTACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGG

5799 ATCTAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCA
     TAGATCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGT

5876 AGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATT
     TCCTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAA

5953 TACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCA
     ATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGT

6030 TAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTC
     ATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAG

6107 AAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGA
     TTTTTAACACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCT

6184 GATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGA
     CTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACT

6261 AACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATC
     TTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAG

6338 ACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCA
     TGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGT
```

```
                            white gene
6415  TGTCTGGATCGGATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTA
      ACAGACCTAGCCTAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAAT

6492  AAAAAACATATTTTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCG
      TTTTTTGTATAAAAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGC

6569  ATCCAAGGGTATTTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTG
      TAGGTTCCCATAAACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGAC

6646  TTTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTG
      AAACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAAC

6723  TAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTAC
      ATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATG

6800  GCCATCAATTAAACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGG
      CGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCC

6877  TGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTAT
      ACCACTAACCAAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATA

6954  TAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGT
      ATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCA

7031  TCACTTGATCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAA
      AGTGAACTAGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTT

7108  ACCGCAATCGCACAAACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTA
      TGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAAT

7185  AGATCATATCATGATCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAG
      TCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTC

7262  ATATTTTAAGCTGATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTC
      TATAAAATTCGACTAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAG

7339  GGGGTTAGATGAGCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCG
      CCCCAATCTACTCGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGC

7416  CAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTC
      GTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAG

7493  CGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATC
      GCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAG

7570  AGCCGGGCTCCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCC
      TCGGCCCGAGGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGG

7647  AGGAAACATTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAA
      TCCTTTGTAAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTT
```

```
7724 TTACCAATTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGG
     AATGGTTAAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCC

7801 AAAGACGACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGC
     TTTCTGCTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACG

7878 TCAATGGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGC
     AGTTACCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCG

7955 TCCCTAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGT
     AGGGATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCA

8032 GGCCCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGG
     CCGGGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCC

8109 TGAAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATC
     ACTTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAG

8186 TGCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAA
     ACGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTT

8263 GGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGG
     CCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTACC

8340 CCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTTAT
     GGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAATA

8417 TAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGG
     ATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAGGCC

8494 CGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGATATGC
     GCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTATACG

8571 GACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGA
     CTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGTGACCT

8648 GCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCTGGCTGT
     CGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGACCGACA

8725 CGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGA
     GCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCTTTGTTTACT

8802 TATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCTTCTTAGTTT
     ATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGAAGAATCAAA

8879 TTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATGTTGAAGATA
     AAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTACAACTTCTAT

8956 TGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCTTTTTGGGCCA
     ACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGAAAAACCCGGT
```

```
 9033 ACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGACCTTTCAAAACG
      TGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAAGTTTTGC

 9110 TCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGA
      AGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAAGATTACTTAGCT

 9187 TTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCGCTGTGACACAT
      AAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGCGACACTGTGTA

 9264 ACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCG
      TGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCGCTAACGGATAGGC

 9341 ATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAAC
      TACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCACCGGTTACACAGTTG

 9418 GTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATAC
      CAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCAGGCGGCCAATAGTATG

 9495 CATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCA
      GTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATATGGAGTTTACCAACAGCATGGAGAGT

 9572 TGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATC
      ACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAG

 9649 GTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGG
      CAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACC

 9726 ACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGC
      TGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCG

 9803 AAGGAGTAGCCGACATATATCCGAAATAACTGCTTGTTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTAT
      TTCCTCATCGGCTGTATATAGGCTTTATTGACGAACAAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATA

 9880 TGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCC
      ACGGGGGAGTTTTTCGATTACATTAATATAAACACGGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGG

 9957 CCTTCGAACATCCCCACAAGTAGACTTTGGATTTGTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCA
      GGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGT

10034 AAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCA
      TTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGT

10111 TAATTCACCTCCACCACACCACGTTTCGTAGTTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCT
      ATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGA

10188 TTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATT
      AAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAA

10265 TCGTCTCAATAGAAATTAGTAATAAATATTTGTATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTC
      AGCAGAGTTATCTTTAATCATTATTTATAAACATACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAG
```

```
10342 ACAGCTATATTTATTCTAATTTAATATTATGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTT
      TGTCGATATAAATAAGATTAAATTATAATACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAA

10419 ACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGT
      TGTTAAACTTGACTTTCACTGTAGGTCACAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCA

10496 GTTGTTTATATATATCAAAAATAACAACTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGG
      CAACAAATATATATAGTTTTTATTGTTGATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACC

10573 AACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTT
      TTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAA

10650 ATTTTCATTAAACAATGAACAGGACCTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAA
      TAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTT

10727 ACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGG
      TGAATGAGACAAACACATGAGGGTGACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCC

10804 TACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAA
      ATGCCGTAGACGCAACTCAGCGGAGGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATT

10881 AGTGGGTATCACAAATTTGGGAGTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCA
      TCACCCATAGTGTTTAAACCCTCAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGT

10958 CACTTCGGCACGTGAATTAATTTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTC
      GTGAAGCCGTGCACTTAATTAAAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAG

11035 CACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAA
      GTGCATTCCCAATTACAAAAGTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTT

                                                                    5' P
11112 TTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      AAATTCACATATGAAGCCATTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
```

pUAST-sKek1-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGAC

  77 GGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCT
     CCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGA

 153 CGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAA
     GCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTT

 229 GGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCG
     CCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGC

 305 ACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTA
     TGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAAT

 381 CCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTC
     GGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAG

 457 GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGT
     CCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCA

 533 AACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
     TTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGT

 609 GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATT
     CTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAA

 685 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACC
     ACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGG

 761 GCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGA
     CGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACT

 837 TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAG
     AGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTC

 913 GATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT
     CTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGA

 989 GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
     CTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTG

1065 TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCC
     AGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGG

1141 ACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACT
     TGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGA

1217 TTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA
     AATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGT

1293 ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCA
     TGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGT

1369 ACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTC
     TGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAG

1445 AGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCG
     TCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGC

1521 TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTC
     ATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAG

1597 TTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
     AACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGA

1673 TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT
     AGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTA

1749 CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
     GAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTA
```

```
1825 AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
     TTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACA

1901 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
     GAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTC

1977 TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG
     ACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGC

2053 TCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAA
     AGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATT

2129 GCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG
     CGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATAC

2205 CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACG
     GCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGC

2281 ACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTA
     TGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGAT

2357 CGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCC
     GCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGG

2433 GGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAAT
     CCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTA

2509 CCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACA
     GGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGT

2585 AATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAG
     TTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTC

2661 CAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGA
     GTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCT

                            3 ' P
2737 AGCTCACGATGAGAATGGCCAGACCATGATGAAATAACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACG
     TCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGC

2813 TATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTCTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGA
     ATACGAACGTTATTCACGCTCACTTTCCTTATCATAAGACTCACAGCATAACTCAGACTCACTCTGTCGCTATACT

2889 TTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAATTAACTCATAATATTAATTAGACGAAATTATTTTTAAAG
     AACAACTAATTGGGAATCGTACAGGCACCCCAAACTTAATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTC

2965 TTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGG
     AAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCC

3041 AGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGG
     TCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCC

                               UAS sites
3117 AGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATT
     TCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAA

3193 CAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAA
     GTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTT

3269 CAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACT
     GTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGA
```

```
                                                                        Polylinker
3345 GAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACAGATC
     CTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGTCTAG
------------------------------------------------------------------------------------
       attB1  (hybrid)
3421 TGACAAGTTTGTACAAAAAAGTTGGAAA ATG CAT ATC AGG GAA GCA GTT TTC CTG GTC CTC ACC
     ACTGTTCAAACATGTTTTTTCAACCTTT TAC GTA TAG TCC CTT CGT CAA AAG GAC CAG GAG TGG

---------------------------1▸ M   H   I   R   E   A   V   F   L   V   L   T ---
3485 CTG CTG CCT GGA ATG ATC CTG GGC ACT CGC TAC AAT CAG CTG CAT CTG TAT GCC AAT
     GAC GAC GGA CCT TAC TAG GAC CCG TGA GCG ATG TTA GTC GAC GTA GAC ATA CGG TTA

---13▸ L   L   P   G   M   I   L   G   T   R   Y   N   Q   L   H   L   Y   A   N ---
3542 GGA GGA GCA TCG TCA TCG GGC CCT GGA GGC TAC AGG CCC GCC CCC TCG TCC CAG AAC
     CCT CCT CGT AGC AGT AGC CCG GGA CCT CCG ATG TCC GGG CGG GGG AGC AGG GTC TTG

---32▸ G   G   A   S   S   S   G   P   G   G   Y   R   P   A   P   S   S   Q   N ---
3599 GAG GTG TAC TCC ATA GCG GAC AGC CAG CCG ATG ACT GAG GAT GGC TAC ATG CCC CCC
     CTC CAC ATG AGG TAT CGC CTG TCG GTC GGC TAC TGA CTC CTA CCG ATG TAC GGG GGG

---51▸ E   V   Y   S   I   A   D   S   Q   P   M   T   E   D   G   Y   M   P   P ---
3656 AGC CAG CAC TTT CCG CCC ACC CAC TCC GAC TTG GAT CCC CCC GCC CAG CAG CAG AGC
     TCG GTC GTG AAA GGC GGG TGG GTG AGG CTG AAC CTA GGG GGG CGG GTC GTC GTC TCG

---70▸ S   Q   H   F   P   P   T   H   S   D   L   D   P   P   A   Q   Q   Q   S ---
3713 ACC TGC CAA ACG GTT TGC GCC TGC AAG TGG AAG GGT GGC AAG CAG ACG GTG GAG TGC
     TGG ACG GTT TGC CAA ACG CGG ACG TTC ACC TTC CCA CCG TTC GTC TGC CAC CTC ACG

---89▸ T   C   Q   T   V   C   A   C   K   W   K   G   G   K   Q   T   V   E   C ---
3770 ATC GAT CGC CAC CTC ATC CAG ATA CCC GAG CAC ATC GAT CCC AAT ACC CAG GTG CTG
     TAG CTA GCG GTG GAG TAG GTC TAT GGG CTC GTG TAG CTA GGG TTA TGG GTC CAC GAC

--108▸ I   D   R   H   L   I   Q   I   P   E   H   I   D   P   N   T   Q   V   L ---
3827 GAC ATG TCC GGT AAT AAG CTG CAG ACC CTC TCC AAC GAG CAG TTC ATC CGT GCG AAT
     CTG TAC AGG CCA TTA TTC GAC GTC TGG GAG AGG TTG CTC GTC AAG TAG GCA CGC TTA

--127▸ D   M   S   G   N   K   L   Q   T   L   S   N   E   Q   F   I   R   A   N ---
3884 CTG CTA AAT CTG CAG AAG CTG TAT TTG AGG AAC TGC AAG ATC GGC GAA ATC GAG CGG
     GAC GAT TTA GAC GTC TTC GAC ATA AAC TCC TTG ACG TTC TAG CCG CTT TAG CTC GCC

--146▸ L   L   N   L   Q   K   L   Y   L   R   N   C   K   I   G   E   I   E   R ---
3941 GAG ACC TTC AAG GGA CTG ACC AAT CTG GTG GAG TTG GAT CTG TCA CAT AAT CTG CTG
     CTC TGG AAG TTC CCT GAC TGG TTA GAC CAC CTC AAC CTA GAC AGT GTA TTA GAC GAC

--165▸ E   T   F   K   G   L   T   N   L   V   E   L   D   L   S   H   N   L   L ---
3998 GTT ACC GTG CCC AGT TTG GCC CTG GGC CAC ATA CCC TCA CTG CGC GAA CTC ACC CTG
     CAA TGG CAC GGG TCA AAC CGG GAC CCG GTG TAT GGG AGT GAC GCG CTT GAG TGG GAC

--184▸ V   T   V   P   S   L   A   L   G   H   I   P   S   L   R   E   L   T   L ---
4055 GCC TCC AAT CAC ATA CAC AAA ATC GAG AGC CAG GCC TTC GGG AAC ACA CCA TCG CTG
     CGG AGG TTA GTG TAT GTG TTT TAG CTC TCG GTC CGG AAG CCC TTG TGT GGT AGC GAC

  203▸ A   S   N   H   I   H   K   I   E   S   Q   A   F   G   N   T   P   S   L
```

```
4112 CAC AAA TTG GAT CTG TCG CAT TGC GAT ATT CAG ACC ATT TCC GCC CAG GCA TTT GGT
     GTG TTT AAC CTA GAC AGC GTA ACG CTA TAA GTC TGG TAA AGG CGG GTC CGT AAA CCA

 222▸ H   K   L   D   L   S   H   C   D   I   Q   T   I   S   A   Q   A   F   G
-------------------------------------------------------------------------------------
4169 GGC CTC CAA GGA TTG ACT TTG CTC CGA TTG AAT GGC AAT AAA CTG AGC GAG CTT TTG
     CCG GAG GTT CCT AAC TGA AAC GAG GCT AAC TTA CCG TTA TTT GAC TCG CTC GAA AAC

 241▸ G   L   Q   G   L   T   L   L   R   L   N   G   N   K   L   S   E   L   L
-------------------------------------------------------------------------------------
4226 CCC AAG ACA ATT GAG ACC CTG AGT CGA CTT CAT GGC ATC GAA CTG CAC GAC AAT CCC
     GGG TTC TGT TAA CTC TGG GAC TCA GCT GAA GTA CCG TAG CTT GAC GTG CTG TTA GGG

 260▸ P   K   T   I   E   T   L   S   R   L   H   G   I   E   L   H   D   N   P
-------------------------------------------------------------------------------------
4283 TGG CTC TGT GAT TGT CGA TTG AGG GAC ACG AAG CTC TGG CTG ATG AAG AGG AAC ATA
     ACC GAG ACA CTA ACA GCT AAC TCC CTG TGC TTC GAG ACC GAC TAC TTC TCC TTG TAT

 279▸ W   L   C   D   C   R   L   R   D   T   K   L   W   L   M   K   R   N   I
-------------------------------------------------------------------------------------
4340 CCC TAT CCG GTG GCT CCG GTT TGC TCG GGT GGC CCC GAA AGG ATT ATC GAT CGC AGC
     GGG ATA GGC CAC CGA GGC CAA ACG AGC CCA CCG GGG CTT TCC TAA TAG CTA GCG TCG

 298▸ P   Y   P   V   A   P   V   C   S   G   G   P   E   R   I   I   D   R   S
-------------------------------------------------------------------------------------
4397 TTT GCG GAT CTG CAT GTG GAT GAG TTT GCC TGC CGA CCG GAG ATG TTG CCC ATA TCG
     AAA CGC CTA GAC GTA CAC CTA CTC AAA CGG ACG GCT GGC CTC TAC AAC GGG TAT AGC

 317▸ F   A   D   L   H   V   D   E   F   A   C   R   P   E   M   L   P   I   S
-------------------------------------------------------------------------------------
4454 CAT TAT GTG GAG GCG GCC ATG GGC GAG AAT GCC TCG ATT ACA TGT CGA GCT CGA GCG
     GTA ATA CAC CTC CGC CGG TAC CCG CTC TTA CGG AGC TAA TGT ACA GCT CGA GCT CGC

 336▸ H   Y   V   E   A   A   M   G   E   N   A   S   I   T   C   R   A   R   A
-------------------------------------------------------------------------------------
4511 GTT CCA GCT GCG AAT ATC AAC TGG TAC TGG AAC GGA CGG CTG CTG GCC AAC AAT TCC
     CAA GGT CGA CGC TTA TAG TTG ACC ATG ACC TTG CCT GCC GAC GAC CGG TTG TTA AGG

 355▸ V   P   A   A   N   I   N   W   Y   W   N   G   R   L   L   A   N   N   S
-------------------------------------------------------------------------------------
4568 GCC TTC ACC GCG TAC CAG AGG ATA CAC ATG TTG GAG CAG GTG GAA GGT GGA TTC GAA
     CGG AAG TGG CGC ATG GTC TCC TAT GTG TAC AAC CTC GTC CAC CTT CCA CCT AAG CTT

 374▸ A   F   T   A   Y   Q   R   I   H   M   L   E   Q   V   E   G   G   F   E
-------------------------------------------------------------------------------------
4625 AAG CGA TCC AAA CTG GTG CTG ACC AAC GCA CAG GAA ACG GAT TCC AGT GAG TTC TAC
     TTC GCT AGG TTT GAC CAC GAC TGG TTG CGT GTC CTT TGC CTA AGG TCA CTC AAG ATG

 393▸ K   R   S   K   L   V   L   T   N   A   Q   E   T   D   S   S   E   F   Y
-------------------------------------------------------------------------------------
4682 TGC GTG GCC GAG AAT CGA GCT GGG ATG GCC GAG GCC AAC TTC ACC CTG CAC GTG AGC
     ACG CAC CGG CTC TTA GCT CGA CCC TAC CGG CTC CGG TTG AAG TGG GAC GTG CAC TCG

 412▸ C   V   A   E   N   R   A   G   M   A   E   A   N   F   T   L   H   V   S
-------------------------------------------------------------------------------------
4739 ATG AGA GCT GCG GGC ATG GCC TCC CTG GGT AGT GGC CAA CCA ACT TTC TTG TAC AAA
     TAC TCT CGA CGC CCG TAC CGG AGG GAC CCA TCA CCG GTT GGT TGA AAG AAC ATG TTT

 431▸ M   R   A   A   G   M   A   S   L   G   S   G   Q   P   T   F   L   Y   K
```

```
                               V5  epitope
                 attB2  (hybrid)
4796  GTG GTG GTA CCG  GGT AAG CCT ATC CCT AAC CCT CTC CTC GGT CTC GAT TCT ACG
      CAC CAC CAT GGC  CCA TTC GGA TAG GGA TTG GGA GAG GAG CCA GAG CTA AGA TGC

 450  V   V   V   P    G   K   P   I   P   N   P   L   L   G   L   D   S   T
----------------------------------------------------------------------------
                6xHis                           SV40  Poly  A
4850  CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA TCTAGAGGATCTTTGTGAAGGAACCTTACTTCTGTG
      GCA TGG CCA GTA GTA GTG GTA GTG GTA ACT AGATCTCCTAGAAACACTTCCTTGGAATGAAGACAC

 468  R   T   G   H   H   H   H   H   H   •
----------------------------------------------------------------------------
4916  GTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATG
      CACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTCACATATTAC
----------------------------------------------------------------------------
4992  TGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGCAGTGGTGG
      ACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCGTCACCACC
----------------------------------------------------------------------------
5068  AATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCTGACTCTC
      TTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGACTGAGAG
----------------------------------------------------------------------------
5144  AACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTT
      TTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTCAAAAAA
----------------------------------------------------------------------------
5220  GAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCACTGCTA
      CTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTGACGAT
----------------------------------------------------------------------------
5296  TACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACATACTGTTTT
      ATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTATGACAAAA
----------------------------------------------------------------------------
5372  TTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAAT
      AAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACATGGAAATCGAAAAATTA
----------------------------------------------------------------------------
5448  TTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACCACATTTGT
      AACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGGTGTAAACA
----------------------------------------------------------------------------
5524  AGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTT
      TCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAA
----------------------------------------------------------------------------
5600  GTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
      CAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAA
----------------------------------------------------------------------------
                                                                 white gene
5676  TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGGATCCACTAGAAG
      AAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCCTAGGTGATCTTC
----------------------------------------------------------------------------
5752  GCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATTTTTTTTTTATT
      CGGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAAAAAAAAAATAA
----------------------------------------------------------------------------
5828  TTTTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTATTTGAAGTACC
      AAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATAAACTTCATGG
```

```
5904 AGGTTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTCTCTGTCCAC
     TCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAGAGACAGGTG

5980 AGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTAGCGTCAAT
     TCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGATCGCAGTTA

6056 GTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTAAACACAAA
     CAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAATTTGTGTTT

6132 GTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGGTTTTGGGT
     CACGACACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACCAAAACCCA

6208 GGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCGGAGGCAGC
     CCCATTCGTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGCCTCCGTCG

6284 AAACACCCATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGATCTATAGG
     TTTGTGGGTAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACTAGATATCC

6360 AACTGCGATTGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAATCGCACAA
     TTGACGCTAACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTTAGCGTGTT

6436 ACAAATAGTGACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCATATCATGA
     TGTTTATCACTGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGTATAGTACT

6512 TCAAGACATCTAAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAGATATTTTAAGCTG
     AGTTCTGTAGATTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAAAATTCGAC

6588 ATCCTAGATGCACAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGGTTAGATGA
     TAGGATCTACGTGTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCCAATCTACT

6664 GCATAACGCTTGTAGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAGAGCTGCAT
     CGTATTGCGAACATCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTCTCGACGTA

6740 TAACCAGGGCTTCGGGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCGGTTCAGGG
     ATTGGTCCCGAAGCCCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGCCAAGTCCC

6816 AGCGGCCAACTAGCCGAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGGGCT
     TCGCCGGTTGATCGGCTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGTCGGCCCGA

6892 CCGGATGGCGGCAGCTGGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCCAGGAAACA
     GGCCTACCGCCGTCGACCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGGTCCTTTGT

6968 TTTGCTCAAGAACGGTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAA
     AAACGAGTTCTTGCCACTCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTTAATGGTT

7044 TTTGAAACTCAGTTTGCGGCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGGAAAGAC
     AAACTTTGAGTCAAACGCCGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCCTTTCTG

7120 GACCCTGCTGAATGCCCTTGCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTCAAT
     CTGGGACGACTTACGGGAACGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACGAGTTA
```

```
7196 GGCCAACCTGTGGACGCCAAGGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGCTCCC
     CCGGTTGGACACCTGCGGTTCCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCGAGGG

7272 TAACGGCCAGGGAACACCTGATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGTGGC
     ATTGCCGGTCCCTTGTGGACTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCACCG

7348 CCGCGTGGATCAGGTGATCCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGGTG
     GGCGCACCTAGTCCACTAGGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCCAC

7424 AAAGGTCTGTCCGGCGGAGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATCT
     TTTCCAGACAGGCCGCCTCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAGA

7500 GCGATGAGCCCACCTCCGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAA
     CGCTACTCGGGTGGAGGCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTT

7576 GGGCAAGACCGTCATCCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATG
     CCCGTTCTGGCAGTAGGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTAC

7652 GCCGAGGGCAGGGTAGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTT
     CGGCTCCCGTCCCATCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAA

7728 ATTAAGGGTATCTAGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATC
     TAATTCCCATAGATCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAG

7804 CGGCGGACTTTTACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGAT
     GCCGCCTGAAAATGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTA

7880 ATGCGACAATTTTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCA
     TACGCTGTTAAAACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGT

7956 CTGGAGCAGCCGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCT
     GACCTCGTCGGCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGA

8032 GGCTGTCGGTGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAA
     CCGACAGCCACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCTTT

8108 CAAATGATATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCTT
     GTTTACTATATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGAA

8184 CTTAGTTTTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATG
     GAATCAAAAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTAC

8260 TTGAAGATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCT
     AACTTCTATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGA

8336 TTTTGGGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGAC
     AAAACCCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTG

8412 CTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTT
     GAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAA
```

```
8488 CTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTAT
     GATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8564 CGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGG
     GCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8640 CGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGT
     GCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8716 GGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGT
     CCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8792 CCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAAAT
     GGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATATGGAGTTTA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8868 GGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTGGAGCCGGG
     CCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCACCTCGGCCC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
8944 CGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTTCTCC
     GCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAATTGAAGAGG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9020 GCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATATCTGGCTC
     CGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTATAGACCGAG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9096 TAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTGTTTTTTTTTTTACCATTATT
     ATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAACAAAAAAAAAAATGGTAATAA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9172 ACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAATAAAAACAAGATATGA
     TGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACACGGTTATTTTTGTTCTATACT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9248 CCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTTGTCTTCTAACCAAAAGACTT
     GGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACAGAAGATTGGTTTTCTGAA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9324 ACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTATATACATAC
     TGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATATAAAAAATATATGTATG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9400 TTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGTTGCTCTTTCGCTGTCTCCCA
     AAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGCGACAGAGGGT
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9476 CCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCGCGATTCGGTTC
     GGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGCGCTAAGCCAAG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9552 GCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGTATGTACAATTTATTTGC
     CGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACATACATGTTAAATAAACG
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9628 TCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACTTTTTAAGGTAATTTT
     AGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGAAAAATTCCATTAAAA
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
9704 TTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGTTCCTTGTGTAGATG
     AACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACAAGGAACACATCTAC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
 9780 CATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAACTATAATAATAAGAATACAT
      GTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTTGATATTATTATTCTTATGTA

 9856 TTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAACGACGCATTTCGTACT
      AATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTTGCTGCGTAAAGCATGA

 9932 CCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGACCTAACGCACAGTCACGTTAT
      GGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGATTGCGTGTCAGTGCAATA

10008 TGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCACTGGTATAGCCTTCTTTTAT
      ACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTGACCATATCGGAAGAAAATA

10084 CTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTCCTTTTAAATGTCTGACCT
      GAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGAGGAAAATTTACAGACTGGA

10160 TTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTTTTCACCAAGGCTGCACC
      AAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAAAAGTGGTTCCGACGTGG

10236 CAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTACTCCAGTCACAGCTTT
      GTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAATGAGGTCAGTGTCGAAA

10312 GCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAAAAAAATTCGTCCGCA
      CGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTTTTTTTAAGCAGGCGT

10388 CACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTTCGGCTATCGACGGG
      GTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAAGCCGATAGCTGCCC

                                 5' P
10464 ACCACCTTATGTTATTTCATCATG
      TGGTGGAATACAATAAAGTAGTAC
```

pUAST-sKek2-GFP

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACTGG
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGACC

  78 GGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACCCTCG
     CCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGGGAGC

 155 TTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAAGAAGGC
     AACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTTCTTCCG

 232 GTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
     CAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGTC

 309 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
     CTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCT

 386 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTA
     ATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACAT

 463 GGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATC
     CCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAATAGGCCATTGATAG

 540 GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
     CAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTCCTAATCGTCTCGCTCC

 617 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG
     ATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTCCTGTCATAAACCATAGAC

 694 CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
     GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTTTGTTTGGTGGCGACCATCGC

 771 GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
     CACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGATGC

 848 GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTA
     CCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTCTAATAGTTTTTCCTAGAAGTGGAT

 925 GATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
     CTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTA

1002 GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
     CGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATC

1079 ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCC
     TATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTACTATGGCGCTCTGGGTGCGAGTGGCCGAGG

1156 AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
     TCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGG

1233 AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
     TCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGA

1310 ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTAC
     TGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATG

1387 ATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAG
     TACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTC

1464 TGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACT
     ACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGA

1541 GGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGA
     CCACTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCT

1618 TAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
     ATTATGGCGCGGTGTATCGTCTTGAAATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCT

1695 TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
     AGAATGGCGACAACTCTAGGTCAAGCTACATTGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGG

1772 AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAAT
     TCGCAAAGACCCACTCGTTTTTGTCCTTCCGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTA
```

```
1849 ACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
     TGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTA

1926 GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
     CATAAATCTTTTTATTTGTTTATCCCCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAA

2003 ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGA
     TAATAGTACTGTAATTGGATATTTTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACT

2080 AAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTC
     TTTGGAGACTGTGTACGTCGAGGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAG

2157 AGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTG
     TCCCGCGCAGTCGCCCACAACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCAC

2234 CACCATATGCGGTGTGAAATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGG
     GTGGTATACGCCACACTTTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCC

2311 TGAATGTGTTGCGGAGAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGA
     ACTTACACAACGCCTCTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACT

2388 AGAGGGCGCGCGATTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGA
     TCTCCCGCGCGCTAAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACT

2465 TGTAAGGTATGCAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGA
     ACATTCCATACGTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCT

2542 AATACTTGTATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATA
     TTATGAACATAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTAT

2619 AACAGTAAACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCG
     TTGTCATTTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGC

                                                                     3'P
2696 TCGGGCCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATA
     AGCCCGGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTAT

2773 ACATAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATT
     TGTATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAA

2850 CTGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAAT
     GACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTTA

2927 TAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTCTGC
     ATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAAATAAAAATTATTAAACGCTCATGCGTTTCGAAGACG

3004 ATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAG
     TACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTC

                                                                  UAS sites
3081 TACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCGCCGGAGT
     ATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGCGGCCTCA

3158 ATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAAGCGAAAGCT
     TATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATTCGCTTTCGA

3235 AAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAATCAATTAAAAGT
     TTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTAGTTAATTTTCA

3312 AACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACAAGAAGAGAACTCT
     TTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGTTCTTCTCTTGAGA
```

```
                           Polylinker            attB1  (hybrid)
3389 GAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGTACAAAAAAGTTGGAAA ATG AGT GGT CTG
     CTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAAACATGTTTTTTCAACCTTT TAC TCA CCA GAC
                                                               1▶ M   S   G   L
------------------------------------------------------------------------------------
3461 CCA ATC TGG ATA CCG CTC CTT GCA CTT CTG GCC ATA ACT GCC GCC TGT CCG CCG GAG
     GGT TAG ACC TAT GGC GAG GAA CGT GAA GAC CGG TAT TGA CGG CGG ACA GGC GGC CTC
   5▶ P   I   W   I   P   L   L   A   L   L   A   I   T   A   A   C   P   P   E
------------------------------------------------------------------------------------
3518 GTG TGT GTA TGC AAA TGG AAG GGG GGC AAG CAG ACG GTG GAG TGC GGC GGC CAG CAG
     CAC ACA CAT ACG TTT ACC TTC CCC CCG TTC GTC TGC CAC CTC ACG CCG CCG GTC GTC
  24▶ V   C   V   C   K   W   K   G   G   K   Q   T   V   E   C   G   G   Q   Q
------------------------------------------------------------------------------------
3575 CTC TCC AAT CTA CCG GAG GGC ATG GAT CCG GGC ACC CAG GTC CTC AAC TTT AGC GGC
     GAG AGG TTA GAT GGC CTC CCG TAC CTA GGC CCG TGG GTC CAG GAG TTG AAA TCG CCG
  43▶ L   S   N   L   P   E   G   M   D   P   G   T   Q   V   L   N   F   S   G
------------------------------------------------------------------------------------
3632 AAT GCG CTG CAG GTA CTG CAA TCG GAG CGG TTT CTA CGT ATG GAT CTG CTA AAC CTG
     TTA CGC GAC GTC CAT GAC GTT AGC CTC GCC AAA GAT GCA TAC CTA GAC GAT TTG GAC
  62▶ N   A   L   Q   V   L   Q   S   E   R   F   L   R   M   D   L   L   N   L
------------------------------------------------------------------------------------
3689 CAG AAG ATT TAT CTG TCA CGG AAT CAG TTG ATC CGG ATA CAC GAG AAG GCC TTC AGG
     GTC TTC TAA ATA GAC AGT GCC TTA GTC AAC TAG GCC TAT GTG CTC TTC CGG AAG TCC
  81▶ Q   K   I   Y   L   S   R   N   Q   L   I   R   I   H   E   K   A   F   R
------------------------------------------------------------------------------------
3746 GGG CTG ACG AAT CTG GTC GAG CTG GAT CTC AGC GAG AAT GCG CTG CAG AAT GTG CCA
     CCC GAC TGC TTA GAC CAG CTC GAC CTA GAG TCG CTC TTA CGC GAC GTC TTA CAC GGT
 100▶ G   L   T   N   L   V   E   L   D   L   S   E   N   A   L   Q   N   V   P
------------------------------------------------------------------------------------
3803 AGC GAA ACG TTT CAG GAC TAC AGC TCT CTA ATG CGC CTT TCG TTA AGT GGA AAT CCT
     TCG CTT TGC AAA GTC CTG ATG TCG AGA GAT TAC GCG GAA AGC AAT TCA CCT TTA GGA
 119▶ S   E   T   F   Q   D   Y   S   S   L   M   R   L   S   L   S   G   N   P
------------------------------------------------------------------------------------
3860 ATC AGG GAG TTA AAG ACA TCC GCC TTT CGG CAC TTG TCT TTT CTC ACG ACA CTA GAG
     TAG TCC CTC AAT TTC TGT AGG CGG AAA GCC GTG AAC AGA AAA GAG TGC TGT GAT CTC
 138▶ I   R   E   L   K   T   S   A   F   R   H   L   S   F   L   T   T   L   E
------------------------------------------------------------------------------------
3917 CTG TCC AAC TGC CAG GTG GAG CGG ATC GAG AAT GAG GCC TTC GTG GGC ATG GAC AAC
     GAC AGG TTG ACG GTC CAC CTC GCC TAG CTC TTA CTC CGG AAG CAC CCG TAC CTG TTG
 157▶ L   S   N   C   Q   V   E   R   I   E   N   E   A   F   V   G   M   D   N
------------------------------------------------------------------------------------
3974 CTG GAG TGG CTG CGA CTG GAC GGC AAT CGG ATT GGG TTC ATC CAG GGC ACC CAC ATC
     GAC CTC ACC GAC GCT GAC CTG CCG TTA GCC TAA CCC AAG TAG GTC CCG TGG GTG TAG
 176▶ L   E   W   L   R   L   D   G   N   R   I   G   F   I   Q   G   T   H   I
------------------------------------------------------------------------------------
4031 CTG CCC AAG TCG CTG CAC GGC ATC AGC CTG CAC AGC AAT CGG TGG AAC TGC GAC TGC
     GAC GGG TTC AGC GAC GTG CCG TAG TCG GAC GTG TCG TTA GCC ACC TTG ACG CTG ACG
 195▶ L   P   K   S   L   H   G   I   S   L   H   S   N   R   W   N   C   D   C
------------------------------------------------------------------------------------
4088 CGC CTT CTA GAC ATC CAC TTC TGG CTG GTC AAC TAT AAC ACG CCT CTG GCG GAG GAA
     GCG GAA GAT CTG TAG GTG AAG ACC GAC CAG TTG ATA TTG TGC GGA GAC CGC CTC CTT
 214▶ R   L   L   D   I   H   F   W   L   V   N   Y   N   T   P   L   A   E   E
------------------------------------------------------------------------------------
4145 CCC AAA TGT ATG GAA CCG GCG AGG CTG AAA GGT CAG GTG ATC AAG AGC CTG CAG CGG
     GGG TTT ACA TAC CTT GGC CGC TCC GAC TTT CCA GTC CAC TAG TTC TCG GAC GTC GCC
 233▶ P   K   C   M   E   P   A   R   L   K   G   Q   V   I   K   S   L   Q   R
------------------------------------------------------------------------------------
4202 GAG CAG CTG GCC TGT CTG CCG GAG GTT AGT CCC CAG TCG AGT TAT ACG GAG GTG AGT
     CTC GTC GAC CGG ACA GAC GGC CTC CAA TCA GGG GTC AGC TCA ATA TGC CTC CAC TCA
 252▶ E   Q   L   A   C   L   P   E   V   S   P   Q   S   S   Y   T   E   V   S
------------------------------------------------------------------------------------
4259 GAG GGC AGG AAC ATG TCC ATC ACC TGC CTG GTC AGG GCC ATC CCG GAG CCG AAG GTC
     CTC CCG TCC TTG TAC AGG TAG TGG ACG GAC CAG TCC CGG TAG GGC CTC GGC TTC CAG
 271▶ E   G   R   N   M   S   I   T   C   L   V   R   A   I   P   E   P   K   V
------------------------------------------------------------------------------------
4316 CTT TGG CTG TTC AAT GGC CAG GTG ATG AGC AAC GAC AGC CTG ATG GAC AAC CTG CAC
     GAA ACC GAC AAG TTA CCG GTC CAC TAC TCG TTG CTG TCG GAC TAC CTG TTG GAC GTG
 290▶ L   W   L   F   N   G   Q   V   M   S   N   D   S   L   M   D   N   L   H
------------------------------------------------------------------------------------
```

```
4373 ATG TAC TAC TAT ATC GAC GAG ACG ATC GGA GTA AGC GGC GCC GAG GAG AAG CGC AGC
     TAC ATG ATG ATA TAG CTG CTC TGC TAG CCT CAT TCG CCG CGG CTC CTC TTC GCG TCG
 309▸ M   Y   Y   Y   I   D   E   T   I   G   V   S   G   A   E   E   K   R   S

4430 GAG ATC TTC ATC TAC AAC GTT GGT GCC GAG GAT AAT GGC ACC TTC TCC TGT GTG GGC
     CTC TAG AAG TAG ATG TTG CAA CCA CGG CTC CTA TTA CCG TGG AAG AGG ACA CAC CCG
 328▸ E   I   F   I   Y   N   V   G   A   E   D   N   G   T   F   S   C   V   G

4487 CAG AAC ATA GCT GGC ACC ACC TTC AGT AAC TAC ACC CTG AGA GTC ATA ATC AAG GAG
     GTC TTG TAT CGA CCG TGG TGG AAG TCA TTG ATG TGG GAC TCT CAG TAT TAG TTC CTC
 347▸ Q   N   I   A   G   T   T   F   S   N   Y   T   L   R   V   I   I   K   E

4544 CCG CCG GTG GTG AAT GAG GTC TCC TTC CCC AGG GAT TAC ATG AAC CCA ACT TTC TTG
     GGC GGC CAC CAC TTA CTC CAG AGG AAG GGG TCC CTA ATG TAC TTG GGT TGA AAG AAC
 366▸ P   P   V   V   N   E   V   S   F   P   R   D   Y   M   N   P   T   F   L

                     attB2  (hybrid)                            EGFP
4601 TAC AAA GTG GTG GTA CCG CGG GCC CGG GAT CCA CCG GTC GCC ACC ATG GTG AGC AAG
     ATG TTT CAC CAC CAT GGC GCC CGG GCC CTA GGT GGC CAG CGG TGG TAC CAC TCG TTC
 385▸ Y   K   V   V   V   P   R   A   R   D   P   P   V   A   T   M   V   S   K

4658 GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC GTA
     CCG CTC CTC GAC AAG TGG CCC CAC CAC GGG TAG GAC CAG CTC GAC CTG CCG CTG CAT
 404▸ G   E   E   L   F   T   G   V   V   P   I   L   V   E   L   D   G   D   V

4715 AAC GGC CAC AAG TTC AGC GTG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG
     TTG CCG GTG TTC AAG TCG CAC AGG CCG CTC CCG CTC CCG CTA CGG TGG ATG CCG TTC
 423▸ N   G   H   K   F   S   V   S   G   E   G   E   G   D   A   T   Y   G   K

4772 CTG ACC CTG AAG TTC ATC TGC ACC ACC GGC AAG CTG CCC GTG CCC TGG CCC ACC CTC
     GAC TGG GAC TTC AAG TAG ACG TGG TGG CCG TTC GAC GGG CAC GGG ACC GGG TGG GAG
 442▸ L   T   L   K   F   I   C   T   T   G   K   L   P   V   P   W   P   T   L

4829 GTG ACC ACC CTG ACC TAC GGC GTG CAG TGC TTC AGC CGC TAC CCC GAC CAC ATG AAG
     CAC TGG TGG GAC TGG ATG CCG CAC GTC ACG AAG TCG GCG ATG GGG CTG GTG TAC TTC
 461▸ V   T   T   L   T   Y   G   V   Q   C   F   S   R   Y   P   D   H   M   K

4886 CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC
     GTC GTG CTG AAG AAG TTC AGG CGG TAC GGG CTT CCG ATG CAG GTC CTC GCG TGG TAG
 480▸ Q   H   D   F   F   K   S   A   M   P   E   G   Y   V   Q   E   R   T   I

4943 TTC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC
     AAG AAG TTC CTG CTG CCG TTG ATG TTC TGG GCG CGG CTC CAC TTC AAG CTC CCG CTG
 499▸ F   F   K   D   D   G   N   Y   K   T   R   A   E   V   K   F   E   G   D

5000 ACC CTG GTG AAC CGC ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC
     TGG GAC CAC TTG GCG TAG CTC GAC TTC CCG TAG CTG AAG TTC CTC CTG CCG TTG TAG
 518▸ T   L   V   N   R   I   E   L   K   G   I   D   F   K   E   D   G   N   I

5057 CTG GGG CAC AAG CTG GAG TAC AAC TAC AAC AGC CAC AAC GTC TAT ATC ATG GCC GAC
     GAC CCC GTG TTC GAC CTC ATG TTG ATG TTG TCG GTG TTG CAG ATA TAG TAC CGG CTG
 537▸ L   G   H   K   L   E   Y   N   Y   N   S   H   N   V   Y   I   M   A   D

5114 AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC CAC AAC ATC GAG GAC GGC
     TTC GTC TTC TTG CCG TAG TTC CAC TTG AAG TTC TAG GCG GTG TTG TAG CTC CTG CCG
 556▸ K   Q   K   N   G   I   K   V   N   F   K   I   R   H   N   I   E   D   G

5171 AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CCC GTG
     TCG CAC GTC GAG CGG CTG GTG ATG GTC GTC TTG TGG GGG TAG CCG CTG CCG GGG CAC
 575▸ S   V   Q   L   A   D   H   Y   Q   Q   N   T   P   I   G   D   G   P   V

5228 CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC CAG TCC GCC CTG AGC AAA GAC CCC AAC
     GAC GAC GGG CTG TTG GTG ATG GAC TCG TGG GTC AGG CGG GAC TCG TTT CTG GGG TTG
 594▸ L   L   P   D   N   H   Y   L   S   T   Q   S   A   L   S   K   D   P   N

5285 GAG AAG CGC GAT CAC ATG GTC CTG CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC
     CTC TTC GCG CTA GTG TAC CAG GAC GAC CTC AAG CAC TGG CGG CGG CCC TAG TGA GAG
 613▸ E   K   R   D   H   M   V   L   L   E   F   V   T   A   A   G   I   T   L
```

```
                                                                    SV40  Poly  A
5342  GGC ATG GAC GAG CTG TAC AAG TAA AGC GGC CGC GAC TCT AGA G GATCTTTGTGAAGGAACCT
      CCG TAC CTG CTC GAC ATG TTC ATT TCG CCG GCG CTG AGA TCT C CTAGAAACACTTCCTTGGA
 632▸ G   M   D   E   L   Y   K   •
-------------------------------------------------------------------------------------
5404  TACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAG
      ATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCATTTATATTTTAAAAATTC
-------------------------------------------------------------------------------------
5481  TGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGAGC
      ACATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATACCTTGACTACTTACCCTCG
-------------------------------------------------------------------------------------
5558  AGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGTGATGATGAGGCTACTGCT
      TCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGATCACTACTACTCCGATGACGA
-------------------------------------------------------------------------------------
5635  GACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAG
      CTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTCCTGAAAGGAAGTCTTAACGATTC
-------------------------------------------------------------------------------------
5712  TTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGCTGCAC
      AAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAAATGTGGTGTTTCCTTTTTCGACGTG
-------------------------------------------------------------------------------------
5789  TGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCATAACAGTTATAATCATAACATACTG
      ACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCCGTATTGTCAATATTAGTATTGTATGAC
-------------------------------------------------------------------------------------
5866  TTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTT
      AAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATACGAGTTTTTAACACATGGAAATCGAAAAA
-------------------------------------------------------------------------------------
5943  AATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACCACATTT
      TTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAACTGATCTCTAGTATTAGTCGGTATGGTGTAAA
-------------------------------------------------------------------------------------
6020  GTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGT
      CATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACA
-------------------------------------------------------------------------------------
6097  TGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
      ACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAA
-------------------------------------------------------------------------------------
                                                                       white gene
6174  TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCGGATCCACTAGAAGG
      AAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACCTAGCCTAGGTGATCTTCC
-------------------------------------------------------------------------------------
6251  CCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGGAGAATGTAGATTTAAAAAAACATATTTTTTTTTTATTTT
      GGAATCATACATACATTCAATTATTTTGGGAAAAAACCTCTTACATCTAAATTTTTTTGTATAAAAAAAAAATAAAA
-------------------------------------------------------------------------------------
6328  TTACTGCACTGGACATCATTGAACTTATCTGATCAGTTTTAAATTTACTTCGATCCAAGGGTATTTGAAGTACCAGG
      AATGACGTGACCTGTAGTAACTTGAATAGACTAGTCAAAATTTAAATGAAGCTAGGTTCCCATAAACTTCATGGTCC
-------------------------------------------------------------------------------------
6405  TTCTTTCGATTACCTCTCACTCAAAATGACATTCCACTCAAAGTCAGCGCTGTTTGCCTCCTTCTCTGTCCACAGAA
      AAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAGGTGAGTTTCAGTCGCGACAAACGGAGGAAGAGACAGGTGTCTT
-------------------------------------------------------------------------------------
6482  ATATCGCCGTCTCTTTCGCCGCTGCGTCCGCTATCTCTTTCGCCACCGTTTGTAGCGTTACCTAGCGTCAATGTCCG
      TATAGCGGCAGAGAAAGCGGCGACGCAGGCGATAGAGAAAGCGGTGGCAAACATCGCAATGGATCGCAGTTACAGGC
```

```
6559 CCTTCAGTTGCACTTTGTCAGCGGTTTCGTGACGAAGCTCCAAGCGGTTTACGCCATCAATTAAACACAAAGTGCTG
     GGAAGTCAACGTGAAACAGTCGCCAAAGCACTGCTTCGAGGTTCGCCAAATGCGGTAGTTAATTTGTGTTTCACGAC

6636 TGCCAAAACTCCTCTCGCTTCTTATTTTTGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGGTTTTGGGTGGGTAAG
     ACGGTTTTGAGGAGAGCGAAGAATAAAAACAAACAAAAAACTCACTAACCCCACCACTAACCAAAACCCACCCATTC

6713 CAGGGGAAAGTGTGAAAAATCCCGGCAATGGGCCAAGAGGATCAGGAGCTATTAATTCGCGGAGGCAGCAAACACCC
     GTCCCCTTTCACACTTTTTAGGGCCGTTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGCCTCCGTCGTTTGTGGG

6790 ATCTGCCGAGCATCTGAACAATGTGAGTAGTACATGTGCATACATCTTAAGTTCACTTGATCTATAGGAACTGCGAT
     TAGACGGCTCGTAGACTTGTTACACTCATCATGTACACGTATGTAGAATTCAAGTGAACTAGATATCCTTGACGCTA

6867 TGCAACATCAAATTGTCTGCGGCGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAATCGCACAAACAAATAGTG
     ACGTTGTAGTTTAACAGACGCCGCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTTAGCGTGTTTGTTTATCAC

6944 ACACGAAACAGATTATTCTGGTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCATATCATGATCAAGACATCT
     TGTGCTTTGTCTAATAAGACCATCGACACGAGCGATATATTCTGTTAAAAATTCTAGTATAGTACTAGTTCTGTAGA

7021 AAAGGCATTCATTTTCGACTACATTCTTTTTTACAAAAAATATAACAACCAGATATTTTAAGCTGATCCTAGATGCA
     TTTCCGTAAGTAAAAGCTGATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAAAATTCGACTAGGATCTACGT

7098 CAAAAAATAAATAAAAGTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGGTTAGATGAGCATAACGCTTGT
     GTTTTTTATTTATTTTCATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCCAATCTACTCGTATTGCGAACA

7175 AGTTGATATTTGAGATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAGAGCTGCATTAACCAGGGCTTCG
     TCAACTATAAACTCTAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTCTCGACGTAATTGGTCCCGAAGC

7252 GGCAGGCCAAAAACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCC
     CCGTCCGGTTTTTGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGCCAAGTCCCTCGCCGGTTGATCGG

7329 GAGAACCTCACCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGGGCTCCGGATGGCGGCAGCT
     CTCTTGGAGTGGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGTCGGCCCGAGGCCTACCGCCGTCGA

7406 GGTCAACCGGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTG
     CCAGTTGGCCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGGTCCTTTGTAAACGAGTTCTTGCCAC

7483 AGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTGAAACTCAGTTTGCG
     TCAAAGATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTTAATGGTTAAACTTTGAGTCAAACGC

7560 GCGTGGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCCCTT
     CGCACCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCCTTTCTGCTGGGACGACTTACGGGAA

7637 GCCTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTCAATGGCCAACCTGTGGACGCCAA
     CGGAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACGAGTTACCGGTTGGACACCTGCGGTT

7714 GGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGCTCCCTAACGGCCAGGGAACACCTGA
     CCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCGAGGGATTGCCGGTCCCTTGTGGACT

7791 TTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGTGGCCCGCGTGGATCAGGTGATCCAG
     AAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCACCGGGCGCACCTAGTCCACTAGGTC
```

```
7868 GAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGGAGAAAG
     CTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCCACTTTCCAGACAGGCCGCCTCTTTC

7945 GAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATCTGCGATGAGCCCACCTCCGGACTGG
     CTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAGACGCTACTCGGGTGGAGGCCTGACC

8022 ACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCATCCTGACCATT
     TGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTTCCCGTTCTGGCAGTAGGACTGGTAA

8099 CATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGGCCGAGGGCAGGGTAGCTTTCTTGGG
     GTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTACCGGCTCCCGTCCCATCGAAAGAACCC

8176 CACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTTATTAAGGGTATCTAGCATTACATTACA
     GTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAATAATTCCCATAGATCGTAATGTAATGT

8253 TCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGGCGGACTTTTACGTACAGGTGTTGGC
     AGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAGGCCGCCTGAAAATGCATGTCCACAACCG

8330 CGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGATATGCGACAATTTTGCTATTAGCAAAGTAG
     GCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTATACGCTGTTAAAACGATAATCGTTTCATC

8407 CCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGCCGGAGAATGGGTACACCTAC
     GGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGTGACCTCGTCGGCCTCTTACCCATGTGGATG

8484 AAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCTGGCTGTCGGTGCTCAAGGAACCACTCCTCGT
     TTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGACCGACAGCCACGAGTTCCTTGGTGAGGAGCA

8561 AAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGATATAACGCTTACAATTCTTGGAAAC
     TTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCTTTGTTTACTATATTGCGAATGTTAAGAACCTTTG

8638 AAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCTTCTTAGTTTTTTTCAATGAGATGTATAGTTTATA
     TTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGAAGAATCAAAAAAAGTTACTCTACATATCAAATAT

8715 GTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATGTTGAAGATATGAATATTAATGAGATGCGAGTAAC
     CAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTACAACTTCTATACTTATAATTACTCTACGCTCATTG

8792 ATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCTTTTTGGGCCAACAACTCACGCAAGTGGGCGTGATG
     TAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGAAAAACCCGGTTGTTGAGTGCGTTCACCCGCACTAC

8869 AATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGACCTTTCAAAACGTCTTTGCCACGATAAATGTAAGTCT
     TTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAAGTTTTGCAGAAACGGTGCTATTTACATTCAGA

8946 TGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTAATGAATCGATTCGATTTAGGTGTTCACCTCAGAG
     ACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAAGATTACTTAGCTAAGCTAAATCCACAAGTGGAGTCTC

9023 CTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCGCTGTGACACATACTTTCTGGGCAAAACGATTGCCGA
     GACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGCGACACTGTGTATGAAAGACCCGTTTTGCTAACGGCT

9100 ATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGCGATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGC
     TAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCGCTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACG
```

```
 9177 TGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGTGGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGC
      ACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCACCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACG

 9254 GCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGGTCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTT
      CGGTCGAGGAGCTGGAGCTACCGCGACAGACACCCAGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAA

 9331 CTTGAACTCGGGCTCGGTGCCAGTATACCTCAAATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTC
      GAACTTGAGCCCGAGCCACGGTCATATGGAGTTTACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAG

 9408 TGCTGATTAACCAATGGGCGGACGTGGAGCCGGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCG
      ACGACTAATTGGTTACCCGCCTGCACCTCGGCCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGC

 9485 GGCAAGGTCATCCTGGAGACGCTTAACTTCTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCAT
      CCGTTCCAGTAGGACCTCTGCGAATTGAAGAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTA

 9562 CGTGAGCTTCCGGGTGCTCGCATATCTGGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAA
      GCACTCGAAGGCCCACGAGCGTATAGACCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTT

 9639 ATAACTGCTTGTTTTTTTTTTTACCATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAAT
      TATTGACGAACAAAAAAAAAAATGGTAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTA

 9716 TATATTTGTGCCAATAAAAACAAGATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGAC
      ATATAAACACGGTTATTTTTGTTCTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTG

 9793 TTTGGATTTGTCTTCTAACCAAAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAA
      AAACCTAAACAGAAGATTGGTTTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTT

 9870 CACCGGGATATATTTTTTATATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTT
      GTGGCCCTATATAAAAAATATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAA

 9947 TCGTAGTTGCTCTTTCGCTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACT
      AGCATCAACGAGAAAGCGACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGA

10024 GTCGTTAGTTCCGCGCGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAA
      CAGCAATCAAGGCGCGCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATT

10101 ATATTTGTATGTACAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAAT
      TATAAACATACATGTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTA

10178 ATTATGACTTTTTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCC
      TAATACTGAAAAATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGG

10255 AGTGTTTGTTCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAAC
      TCACAAACAAGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTG

10332 AACTATAATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCT
      TTGATATTATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGA

10409 CTAAACGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGAC
      GATTTGCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTG
```

```
10486 CTAACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCAC
      GATTGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTG
-------------------------------------------------------------------------------------
10563 TGGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCTC
      ACCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGAG
-------------------------------------------------------------------------------------
10640 CTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGAGTT
      GAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCTCAA
-------------------------------------------------------------------------------------
10717 TTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAATTTTA
      AAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATTAAAAT
-------------------------------------------------------------------------------------
10794 CTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTTTTCAAAA
      GAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAAAAGTTTT
-------------------------------------------------------------------------------------
10871 AAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTCGGTAAGCTT
      TTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAGCCATTCGAA
-------------------------------------------------------------------------------------
                                              5' P
10948 CGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      GCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
-------------------------------------------------------------------------------------
```

pUAST-sKek2-V5/6XHis

```
   1 GGCCAGACCCACGTAGTCCAGCGGCAGATCGGCGGCGGAGAAGTTAAGCGTCTCCAGGATGACCTTGCCCGAACT
     CCGGTCTGGGTGCATCAGGTCGCCGTCTAGCCGCCGCCTCTTCAATTCGCAGAGGTCCTACTGGAACGGGCTTGA

  76 GGGGCACGTGGTGTTCGACGATGTGCAGCTAATTTCGCCCGGCTCCACGTCCGCCCATTGGTTAATCAGCAGACC
     CCCCGTGCACCACAAGCTGCTACACGTCGATTAAAGCGGGCCGAGGTGCAGGCGGGTAACCAATTAGTCGTCTGG

 151 CTCGTTGGCGTAACGGAACCATGAGAGGTACGACAACCATTTGAGGTATACTGGCACCGAGCCCGAGTTCAAGAA
     GAGCAACCGCATTGCCTTGGTACTCTCCATGCTGTTGGTAAACTCCATATGACCGTGGCTCGGGCTCAAGTTCTT

 226 GAAGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
     CTTCCGCAAAAAGGTATCCGAGGCGGGGGGACTGCTCGTAGTGTTTTTAGCTGCGAGTTCAGTCTCCACCGCTTT

 301 CCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC
     GGGCTGTCCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGG

 376 GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCT
     CGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGCACCGCGAAAGAGTTACGAGTGCGACATCCATAGA

 451 CAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
     GTCAAGCCACATCCAGCAAGCGAGGTTCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACGCGGAA

 526 ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
     TAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCGTCGGTGACCATTGTC

 601 GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
     CTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGATTGATGCCGATGTGATCTTC

 676 GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAA
     CTGTCATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAACTAGGCCGTT

 751 ACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGA
     TGTTTGGTGGCGACCATCGCCACCAAAAAAACAAACGTTCGTCGTCTAATGCGCGTCTTTTTTTCCTAGAGTTCT

 826 AGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
     TCTAGGAAACTAGAAAAGATGCCCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATTCCCTAAAACCAGTACTC

 901 ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGA
     TAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATTTTTACTTCAAAATTTAGTTAGATTTCATATATACT

 976 GTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATC
     CATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAG

1051 CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAAT
     GTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGGTCACGACGTTA

1126 GATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
     CTATGGCGCTCTGGGTGCGAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTCGGTCGGCCTTCCCGGCTCGCGTC

1201 AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCC
     TTCACCAGGACGTTGAAATAGGCGGAGGTAGGTCAGATAATTAACAACGGCCCTTCGATCTCATTCATCAAGCGG

1276 AGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTC
     TCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACCACAGTGCGAGCAGCAAACCATACCGAAG

1351 ATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT
     TAAGTCGAGGCCAAGGGTTGCTAGTTCCGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAA

1426 CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTC
     GCCAGGAGGCTAGCAACAGTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAG

1501 TCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTG
     AGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCACTCATGAGTTGGTTCAGTAAGACTCTTATCAC

1576 TATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAACACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
     ATACGCCGCTGGCTCAACGAGAACGGGCCGCAGTTGTGCCCTATTATGGCGCGGTGTATCGTCTTGAAATTTTCA

1651 GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTA
     CGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCCTAGAATGGCGACAACTCTAGGTCAAGCTACAT

1726 ACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
     TGGGTGAGCACGTGGGTTGACTAGAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTCCTTC
```

```
1801 GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTA
     CGTTTTACGGCGTTTTTTCCCTTATTCCCGCTGTGCCTTTACAACTTATGAGTATGAGAAGGAAAAAGTTATAAT

1876 TTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGG
     AACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTTATTTGTTTATCC

1951 GGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAA
     CCAAGGCGCGTGTAAAGGGGCTTTTCACGGTGGACTGCAGATTCTTTGGTAATAATAGTACTGTAATTGGATATT

2026 AAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCT
     TTTATCCGCATAGTGCTCCGGGAAAGCAGAGCGCGCAAAGCCACTACTGCCACTTTTGGAGACTGTGTACGTCGA

2101 CCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGT
     GGGCCTCTGCCAGTGTCGAACAGACATTCGCCTACGGCCCTCGTCTGTTCGGGCAGTCCCGCGCAGTCGCCCACA

2176 TGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGA
     ACCGCCCACAGCCCCGACCGAATTGATACGCCGTAGTCTCGTCTAACATGACTCTCACGTGGTATACGCCACACT

2251 AATACCGCACCGAATCGCGCGGAACTAACGACAGTCGCTCCAAGGTCGTCGAACAAAAGGTGAATGTGTTGCGGA
     TTATGGCGTGGCTTAGCGCGCCTTGATTGCTGTCAGCGAGGTTCCAGCAGCTTGTTTTCCACTTACACAACGCCT

2326 GAGCGGGTGGGAGACAGCGAAAGAGCAACTACGAAACGTGGTGTGGTGGAGGTGAATTATGAAGAGGGCGCGCGA
     CTCGCCCACCCTCTGTCGCTTTCTCGTTGATGCTTTGCACCACACCACCTCCACTTAATACTTCTCCCGCGCGCT

2401 TTTGAAAAGTATGTATATAAAAAATATATCCCGGTGTTTTATGTAGCGATAAACGAGTTTTTGATGTAAGGTATG
     AAACTTTTCATACATATATTTTTTATATAGGGCCACAAAATACATCGCTATTTGCTCAAAAACTACATTCCATAC

2476 CAGGTGTGTAAGTCTTTTGGTTAGAAGACAAATCCAAAGTCTACTTGTGGGGATGTTCGAAGGGGAAATACTTGT
     GTCCACACATTCAGAAAACCAATCTTCTGTTTAGGTTTCAGATGAACACCCCTACAAGCTTCCCCTTTATGAACA

2551 ATTCTATAGGTCATATCTTGTTTTTATTGGCACAAATATAATTACATTAGCTTTTTGAGGGGGCAATAAACAGTA
     TAAGATATCCAGTATAGAACAAAAATAACCGTGTTTATATTAATGTAATCGAAAAACTCCCCCGTTATTTGTCAT

2626 AACACGATGGTAATAATGGTAAAAAAAAAAACAAGCAGTTATTTCGGATATATGTCGGCTACTCCTTGCGTCGGG
     TTGTGCTACCATTATTACCATTTTTTTTTTTGTTCGTCAATAAAGCCTATATACAGCCGATGAGGAACGCAGCCC

                                                                3'P
2701 CCCGAAGTCTTAGAGCCAGATATGCGAGCACCCGGAAGCTCACGATGAGAATGGCCAGACCATGATGAAATAACA
     GGGCTTCAGAATCTCGGTCTATACGCTCGTGGGCCTTCGAGTGCTACTCTTACCGGTCTGGTACTACTTTATTGT

2776 TAAGGTGGTCCCGTCGGCAAGAGACATCCACTTAACGTATGCTTGCAATAAGTGCGAGTGAAAGGAATAGTATTC
     ATTCCACCAGGGCAGCCGTTCTCTGTAGGTGAATTGCATACGAACGTTATTCACGCTCACTTTCCTTATCATAAG

2851 TGAGTGTCGTATTGAGTCTGAGTGAGACAGCGATATGATTGTTGATTAACCCTTAGCATGTCCGTGGGGTTTGAA
     ACTCACAGCATAACTCAGACTCACTCTGTCGCTATACTAACAACTAATTGGGAATCGTACAGGCACCCCAAACTT

2926 TTAACTCATAATATTAATTAGACGAAATTATTTTTAAAGTTTATTTTTAATAATTTGCGAGTACGCAAAGCTTC
     AATTGAGTATTATAATTAATCTGCTTTAATAAAAATTTCAAATAAAAATTATTAAACGCTCATGCGTTTCGAAG

3001 TGCATGAGCTCGGATCCAAGCTTGCATGCCTGCAGGTCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAG
     ACGTACTCGAGCCTAGGTTCGAACGTACGGACGTCCAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTC

                                                                        UAS sites
3076 CGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGTACTGTCCTCCGAGCGGAGACTCTAGCGAGCG
     GCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTCTGAGATCGCTCGC

3151 CCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACACGTCGCTAA
     GGCCTCATATTTATCTCCGCGAAGCAGATGCCTCGCTGTTAAGTTAAGTTTGTTCGTTTCACTTGTGCAGCGATT

3226 GCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGTAAAGTGCAAGTTAAAGTGAAT
     CGCTTTCGATTCGTTTATTTGTTCGCGTCGACTTGTTCGATTTGTTAGACGTCATTTCACGTTCAATTTCACTTA
```

```
3301 CAATTAAAAGTAACCAGCAACCAAGTAAATCAACTGCAACTACTGAAATCTGCCAAGAAGTAATTATTGAATACA
     GTTAATTTTCATTGGTCGTTGGTTCATTTAGTTGACGTTGATGACTTTAGACGGTTCTTCATTAATAACTTATGT
---------------------------------------------------------------------------------
                                 Polylinker           attB1  (hybrid)
3376 AGAAGAGAACTCTGAATAGGGAATTGGGGAATTCGTTAACAGATCTGACAAGTTTGTACAAAAAAGTTGGAAA
     TCTTCTCTTGAGACTTATCCCTTAACCCCTTAAGCAATTGTCTAGACTGTTCAAACATGTTTTTTCAACCTTT
---------------------------------------------------------------------------------
3449 ATG AGT GGT CTG CCA ATC TGG ATA CCG CTC CTT GCA CTT CTG GCC ATA ACT GCC GCC
     TAC TCA CCA GAC GGT TAG ACC TAT GGC GAG GAA CGT GAA GAC CGG TAT TGA CGG CGG

   1▸ M   S   G   L   P   I   W   I   P   L   L   A   L   L   A   I   T   A   A
---------------------------------------------------------------------------------
3506 TGT CCG CCG GAG GTG TGT GTA TGC AAA TGG AAG GGG GGC AAG CAG ACG GTG GAG TGC
     ACA GGC GGC CTC CAC ACA CAT ACG TTT ACC TTC CCC CCG TTC GTC TGC CAC CTC ACG

  20▸ C   P   P   E   V   C   V   C   K   W   K   G   G   K   Q   T   V   E   C
---------------------------------------------------------------------------------
3563 GGC GGC CAG CAG CTC TCC AAT CTA CCG GAG GGC ATG GAT CCG GGC ACC CAG GTC CTC
     CCG CCG GTC GTC GAG AGG TTA GAT GGC CTC CCG TAC CTA GGC CCG TGG GTC CAG GAG

  39▸ G   G   Q   Q   L   S   N   L   P   E   G   M   D   P   G   T   Q   V   L
---------------------------------------------------------------------------------
3620 AAC TTT AGC GGC AAT GCG CTG CAG GTA CTG CAA TCG GAG CGG TTT CTA CGT ATG GAT
     TTG AAA TCG CCG TTA CGC GAC GTC CAT GAC GTT AGC CTC GCC AAA GAT GCA TAC CTA

  58▸ N   F   S   G   N   A   L   Q   V   L   Q   S   E   R   F   L   R   M   D
---------------------------------------------------------------------------------
3677 CTG CTA AAC CTG CAG AAG ATT TAT CTG TCA CGG AAT CAG TTG ATC CGG ATA CAC GAG
     GAC GAT TTG GAC GTC TTC TAA ATA GAC AGT GCC TTA GTC AAC TAG GCC TAT GTG CTC

  77▸ L   L   N   L   Q   K   I   Y   L   S   R   N   Q   L   I   R   I   H   E
---------------------------------------------------------------------------------
3734 AAG GCC TTC AGG GGG CTG ACG AAT CTG GTC GAG CTG GAT CTC AGC GAG AAT GCG CTG
     TTC CGG AAG TCC CCC GAC TGC TTA GAC CAG CTC GAC CTA GAG TCG CTC TTA CGC GAC

  96▸ K   A   F   R   G   L   T   N   L   V   E   L   D   L   S   E   N   A   L
---------------------------------------------------------------------------------
3791 CAG AAT GTG CCA AGC GAA ACG TTT CAG GAC TAC AGC TCT CTA ATG CGC CTT TCG TTA
     GTC TTA CAC GGT TCG CTT TGC AAA GTC CTG ATG TCG AGA GAT TAC GCG GAA AGC AAT

 115▸ Q   N   V   P   S   E   T   F   Q   D   Y   S   S   L   M   R   L   S   L
---------------------------------------------------------------------------------
3848 AGT GGA AAT CCT ATC AGG GAG TTA AAG ACA TCC GCC TTT CGG CAC TTG TCT TTT CTC
     TCA CCT TTA GGA TAG TCC CTC AAT TTC TGT AGG CGG AAA GCC GTG AAC AGA AAA GAG

 134▸ S   G   N   P   I   R   E   L   K   T   S   A   F   R   H   L   S   F   L
---------------------------------------------------------------------------------
3905 ACG ACA CTA GAG CTG TCC AAC TGC CAG GTG GAG CGG ATC GAG AAT GAG GCC TTC GTG
     TGC TGT GAT CTC GAC AGG TTG ACG GTC CAC CTC GCC TAG CTC TTA CTC CGG AAG CAC

 153▸ T   T   L   E   L   S   N   C   Q   V   E   R   I   E   N   E   A   F   V
---------------------------------------------------------------------------------
3962 GGC ATG GAC AAC CTG GAG TGG CTG CGA CTG GAC GGC AAT CGG ATT GGG TTC ATC CAG
     CCG TAC CTG TTG GAC CTC ACC GAC GCT GAC CTG CCG TTA GCC TAA CCC AAG TAG GTC

 172▸ G   M   D   N   L   E   W   L   R   L   D   G   N   R   I   G   F   I   Q
---------------------------------------------------------------------------------
4019 GGC ACC CAC ATC CTG CCC AAG TCG CTG CAC GGC ATC AGC CTG CAC AGC AAT CGG TGG
     CCG TGG GTG TAG GAC GGG TTC AGC GAC GTG CCG TAG TCG GAC GTG TCG TTA GCC ACC

 191▸ G   T   H   I   L   P   K   S   L   H   G   I   S   L   H   S   N   R   W
```

```
4076 AAC TGC GAC TGC CGC CTT CTA GAC ATC CAC TTC TGG CTG GTC AAC TAT AAC ACG CCT
     TTG ACG CTG ACG GCG GAA GAT CTG TAG GTG AAG ACC GAC CAG TTG ATA TTG TGC GGA

 210▸ N   C   D   C   R   L   L   D   I   H   F   W   L   V   N   Y   N   T   P

4133 CTG GCG GAG GAA CCC AAA TGT ATG GAA CCG GCG AGG CTG AAA GGT CAG GTG ATC AAG
     GAC CGC CTC CTT GGG TTT ACA TAC CTT GGC CGC TCC GAC TTT CCA GTC CAC TAG TTC

 229▸ L   A   E   E   P   K   C   M   E   P   A   R   L   K   G   Q   V   I   K

4190 AGC CTG CAG CGG GAG CAG CTG GCC TGT CTG CCG GAG GTT AGT CCC CAG TCG AGT TAT
     TCG GAC GTC GCC CTC GTC GAC CGG ACA GAC GGC CTC CAA TCA GGG GTC AGC TCA ATA

 248▸ S   L   Q   R   E   Q   L   A   C   L   P   E   V   S   P   Q   S   S   Y

4247 ACG GAG GTG AGT GAG GGC AGG AAC ATG TCC ATC ACC TGC CTG GTC AGG GCC ATC CCG
     TGC CTC CAC TCA CTC CCG TCC TTG TAC AGG TAG TGG ACG GAC CAG TCC CGG TAG GGC

 267▸ T   E   V   S   E   G   R   N   M   S   I   T   C   L   V   R   A   I   P

4304 GAG CCG AAG GTC CTT TGG CTG TTC AAT GGC CAG GTG ATG AGC AAC GAC AGC CTG ATG
     CTC GGC TTC CAG GAA ACC GAC AAG TTA CCG GTC CAC TAC TCG TTG CTG TCG GAC TAC

 286▸ E   P   K   V   L   W   L   F   N   G   Q   V   M   S   N   D   S   L   M

4361 GAC AAC CTG CAC ATG TAC TAC TAT ATC GAC GAG ACG ATC GGA GTA AGC GGC GCC GAG
     CTG TTG GAC GTG TAC ATG ATG ATA TAG CTG CTC TGC TAG CCT CAT TCG CCG CGG CTC

 305▸ D   N   L   H   M   Y   Y   Y   I   D   E   T   I   G   V   S   G   A   E

4418 GAG AAG CGC AGC GAG ATC TTC ATC TAC AAC GTT GGT GCC GAG GAT AAT GGC ACC TTC
     CTC TTC GCG TCG CTC TAG AAG TAG ATG TTG CAA CCA CGG CTC CTA TTA CCG TGG AAG

 324▸ E   K   R   S   E   I   F   I   Y   N   V   G   A   E   D   N   G   T   F

4475 TCC TGT GTG GGC CAG AAC ATA GCT GGC ACC ACC TTC AGT AAC TAC ACC CTG AGA GTC
     AGG ACA CAC CCG GTC TTG TAT CGA CCG TGG TGG AAG TCA TTG ATG TGG GAC TCT CAG

 343▸ S   C   V   G   Q   N   I   A   G   T   T   F   S   N   Y   T   L   R   V

4532 ATA ATC AAG GAG CCG CCG GTG GTG AAT GAG GTC TCC TTC CCC AGG GAT TAC ATG AAC
     TAT TAG TTC CTC GGC GGC CAC CAC TTA CTC CAG AGG AAG GGG TCC CTA ATG TAC TTG

 362▸ I   I   K   E   P   P   V   V   N   E   V   S   F   P   R   D   Y   M   N

                                           V5  epitope
                               attB2  (hybrid)
4589 CCA ACT TTC TTG TAC AAA GTG GTG GTA CCG  GGT AAG CCT ATC CCT AAC CCT CTC

     GGT TGA AAG AAC ATG TTT CAC CAC CAT GGC  CCA TTC GGA TAG GGA TTG GGA GAG

 381▸ P   T   F   L   Y   K   V   V   V   P    G   K   P   I   P   N   P   L

                                         6xHis                   SV40  Poly
4643 CTC GGT CTC GAT TCT ACG CGT ACC GGT CAT CAT CAC CAT CAC CAT TGA TCTAGAGGATC
     GAG CCA GAG CTA AGA TGC GCA TGG CCA GTA GTA GTG GTA GTG GTA ACT AGATCTCCTAG

 399▸ L   G   L   D   S   T   R   T   G   H   H   H   H   H   H   •

4702 TTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAAAGCTCTAAGGTA
     AAACACTTCCTTGGAATGAAGACACCACACTGTATTAACCTGTTTGATGGATGTCTCTAAATTTCGAGATTCCAT
```

```
4777 AATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTAT
     TTATATTTTAAAAATTCACATATTACACAATTTGATGACTAAGATTAACAAACACATAAAATCTAAGGTTGGATA

4852 GGAACTGATGAATGGGAGCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCT
     CCTTGACTACTTACCCTCGTCACCACCTTACGGAAATTACTCCTTTTGGACAAAACGAGTCTTCTTTACGGTAGA

4927 AGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGAAAGGTAGAAGACCCCAAG
     TCACTACTACTCCGATGACGACTGAGAGTTGTAAGATGAGGAGGTTTTTTCTTCTCTTTCCATCTTCTGGGGTTC

5002 GACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGCTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATT
     CTGAAAGGAAGTCTTAACGATTCAAAAAACTCAGTACGACACAAATCATTATCTTGAGAACGAACGAAACGATAA

5077 TACACCACAAAGGAAAAAGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGG
     ATGTGGTGTTTCCTTTTTCGACGTGACGATATGTTCTTTTAATACCTTTTTATAAGACATTGGAAATATTCATCC

5152 CATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCTGCTATTAATAACTAT
     GTATTGTCAATATTAGTATTGTATGACAAAAAAGAATGAGGTGTGTCCGTATCTCACAGACGATAATTATTGATA

5227 GCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTG
     CGAGTTTTTAACACATGGAAATCGAAAAATTAAACATTTCCCCAATTATTCCTTATAAACTACATATCACGGAAC

5302 ACTAGAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCT
     TGATCTCTAGTATTAGTCGGTATGGTGTAAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTGTGGAGGGGGA

5377 GAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAG
     CTTGGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTC

5452 CAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAA
     GTTATCGTAGTGTTTAAAGTGTTTATTTCGTAAAAAAAGTGACGTAAGATCAACACCAAACAGGTTTGAGTAGTT

                                 white gene
5527 TGTATCTTATCATGTCTGGATCGGATCCACTAGAAGGCCTTAGTATGTATGTAAGTTAATAAAACCCTTTTTTGG
     ACATAGAATAGTACAGACCTAGCCTAGGTGATCTTCCGGAATCATACATACATTCAATTATTTTGGGAAAAAACC

5602 AGAATGTAGATTTAAAAAAACATATTTTTTTTTTATTTTTTACTGCACTGGACATCATTGAACTTATCTGATCAG
     TCTTACATCTAAATTTTTTTGTATAAAAAAAAAATAAAAAATGACGTGACCTGTAGTAACTTGAATAGACTAGTC

5677 TTTTAAATTTACTTCGATCCAAGGGTATTTGAAGTACCAGGTTCTTTCGATTACCTCTCACTCAAAATGACATTC
     AAAATTTAAATGAAGCTAGGTTCCCATAAACTTCATGGTCCAAGAAAGCTAATGGAGAGTGAGTTTTACTGTAAG

5752 CACTCAAAGTCAGCGCTGTTTGCCTCCTTCTCTGTCCACAGAAATATCGCCGTCTCTTTCGCCGCTGCGTCCGCT
     GTGAGTTTCAGTCGCGACAAACGGAGGAAGAGACAGGTGTCTTTATAGCGGCAGAGAAAGCGGCGACGCAGGCGA

5827 ATCTCTTTCGCCACCGTTTGTAGCGTTACCTAGCGTCAATGTCCGCCTTCAGTTGCACTTTGTCAGCGGTTTCGT
     TAGAGAAAGCGGTGGCAAACATCGCAATGGATCGCAGTTACAGGCGGAAGTCAACGTGAAACAGTCGCCAAAGCA

5902 GACGAAGCTCCAAGCGGTTTACGCCATCAATTAAACACAAAGTGCTGTGCCAAAACTCCTCTCGCTTCTTATTTT
     CTGCTTCGAGGTTCGCCAAATGCGGTAGTTAATTTGTGTTTCACGACACGGTTTTGAGGAGAGCGAAGAATAAAA

5977 TGTTTGTTTTTTGAGTGATTGGGGTGGTGATTGGTTTTGGGTGGGTAAGCAGGGGAAAGTGTGAAAAATCCCGGC
     ACAAACAAAAAACTCACTAACCCCACCACTAACCAAAACCCACCCATTCGTCCCCTTTCACACTTTTTAGGGCCG
```

```
6052 AATGGGCCAAGAGGATCAGGAGCTATTAATTCGCGGAGGCAGCAAACACCCATCTGCCGAGCATCTGAACAATGT
     TTACCCGGTTCTCCTAGTCCTCGATAATTAAGCGCCTCCGTCGTTTGTGGGTAGACGGCTCGTAGACTTGTTACA

6127 GAGTAGTACATGTGCATACATCTTAAGTTCACTTGATCTATAGGAACTGCGATTGCAACATCAAATTGTCTGCGG
     CTCATCATGTACACGTATGTAGAATTCAAGTGAACTAGATATCCTTGACGCTAACGTTGTAGTTTAACAGACGCC

6202 CGTGAGAACTGCGACCCACAAAAATCCCAAACCGCAATCGCACAAACAAATAGTGACACGAAACAGATTATTCTG
     GCACTCTTGACGCTGGGTGTTTTTAGGGTTTGGCGTTAGCGTGTTTGTTTATCACTGTGCTTTGTCTAATAAGAC

6277 GTAGCTGTGCTCGCTATATAAGACAATTTTTAAGATCATATCATGATCAAGACATCTAAAGGCATTCATTTTCGA
     CATCGACACGAGCGATATATTCTGTTAAAAATTCTAGTATAGTACTAGTTCTGTAGATTTCCGTAAGTAAAAGCT

6352 CTACATTCTTTTTTACAAAAAATATAACAACCAGATATTTTAAGCTGATCCTAGATGCACAAAAAATAAATAAAA
     GATGTAAGAAAAAATGTTTTTTATATTGTTGGTCTATAAAATTCGACTAGGATCTACGTGTTTTTTATTTATTTT

6427 GTATAAACCTACTTCGTAGGATACTTCGTTTTGTTCGGGGTTAGATGAGCATAACGCTTGTAGTTGATATTTGAG
     CATATTTGGATGAAGCATCCTATGAAGCAAAACAAGCCCCAATCTACTCGTATTGCGAACATCAACTATAAACTC

6502 ATCCCCTATCATTGCAGGGTGACAGCGGAGCGGCTTCGCAGAGCTGCATTAACCAGGGCTTCGGGCAGGCCAAAA
     TAGGGGATAGTAACGTCCCACTGTCGCCTCGCCGAAGCGTCTCGACGTAATTGGTCCCGAAGCCCGTCCGGTTTT

6577 ACTACGGCACGCTCCTGCCACCCAGTCCGCCGGAGGACTCCGGTTCAGGGAGCGGCCAACTAGCCGAGAACCTCA
     TGATGCCGTGCGAGGACGGTGGGTCAGGCGGCCTCCTGAGGCCAAGTCCCTCGCCGGTTGATCGGCTCTTGGAGT

6652 CCTATGCCTGGCACAATATGGACATCTTTGGGGCGGTCAATCAGCCGGGCTCCGGATGGCGGCAGCTGGTCAACC
     GGATACGGACCGTGTTATACCTGTAGAAACCCCGCCAGTTAGTCGGCCCGAGGCCTACCGCCGTCGACCAGTTGG

6727 GGACACGCGGACTATTCTGCAACGAGCGACACATACCGGCGCCCAGGAAACATTTGCTCAAGAACGGTGAGTTTC
     CCTGTGCGCCTGATAAGACGTTGCTCGCTGTGTATGGCCGCGGGTCCTTTGTAAACGAGTTCTTGCCACTCAAAG

6802 TATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTGAAACTCAGTTTGCGGCGT
     ATAAGCGTCAGCCGACTAGACACACTTTAGAATTATTTCCCAGGTTAATGGTTAAACTTTGAGTCAAACGCCGCA

6877 GGCCTATCCGGGCGAACTTTTGGCCGTGATGGGCAGTTCCGGTGCCGGAAAGACGACCCTGCTGAATGCCCTTGC
     CCGGATAGGCCCGCTTGAAAACCGGCACTACCCGTCAAGGCCACGGCCTTTCTGCTGGGACGACTTACGGGAACG

6952 CTTTCGATCGCCGCAGGGCATCCAAGTATCGCCATCCGGGATGCGACTGCTCAATGGCCAACCTGTGGACGCCAA
     GAAAGCTAGCGGCGTCCCGTAGGTTCATAGCGGTAGGCCCTACGCTGACGAGTTACCGGTTGGACACCTGCGGTT

7027 GGAGATGCAGGCCAGGTGCGCCTATGTCCAGCAGGATGACCTCTTTATCGGCTCCCTAACGGCCAGGGAACACCT
     CCTCTACGTCCGGTCCACGCGGATACAGGTCGTCCTACTGGAGAAATAGCCGAGGGATTGCCGGTCCCTTGTGGA

7102 GATTTTCCAGGCCATGGTGCGGATGCCACGACATCTGACCTATCGGCAGCGAGTGGCCCGCGTGGATCAGGTGAT
     CTAAAAGGTCCGGTACCACGCCTACGGTGCTGTAGACTGGATAGCCGTCGCTCACCGGGCGCACCTAGTCCACTA

7177 CCAGGAGCTTTCGCTCAGCAAATGTCAGCACACGATCATCGGTGTGCCCGGCAGGGTGAAAGGTCTGTCCGGCGG
     GGTCCTCGAAAGCGAGTCGTTTACAGTCGTGTGCTAGTAGCCACACGGGCCGTCCCACTTTCCAGACAGGCCGCC

7252 AGAAAGGAAGCGTCTGGCATTCGCCTCCGAGGCACTAACCGATCCGCCGCTTCTGATCTGCGATGAGCCCACCTC
     TCTTTCCTTCGCAGACCGTAAGCGGAGGCTCCGTGATTGGCTAGGCGGCGAAGACTAGACGCTACTCGGGTGGAG
```

```
7327 CGGACTGGACTCATTTACCGCCCACAGCGTCGTCCAGGTGCTGAAGAAGCTGTCGCAGAAGGGCAAGACCGTCAT
     GCCTGACCTGAGTAAATGGCGGGTGTCGCAGCAGGTCCACGACTTCTTCGACAGCGTCTTCCCGTTCTGGCAGTA

7402 CCTGACCATTCATCAGCCGTCTTCCGAGCTGTTTGAGCTCTTTGACAAGATCCTTCTGATGGCCGAGGGCAGGGT
     GGACTGGTAAGTAGTCGGCAGAAGGCTCGACAAACTCGAGAAACTGTTCTAGGAAGACTACCGGCTCCCGTCCCA

7477 AGCTTTCTTGGGCACTCCCAGCGAAGCCGTCGACTTCTTTTCCTAGTGAGTTCGATGTGTTTATTAAGGGTATCT
     TCGAAAGAACCCGTGAGGGTCGCTTCGGCAGCTGAAGAAAAGGATCACTCAAGCTACACAAATAATTCCCATAGA

7552 AGCATTACATTACATCTCAACTCCTATCCAGCGTGGGTGCCCAGTGTCCTACCAACTACAATCCGGCGGACTTTT
     TCGTAATGTAATGTAGAGTTGAGGATAGGTCGCACCCACGGGTCACAGGATGGTTGATGTTAGGCCGCCTGAAAA

7627 ACGTACAGGTGTTGGCCGTTGTGCCCGGACGGGAGATCGAGTCCCGTGATCGGATCGCCAAGATATGCGACAATT
     TGCATGTCCACAACCGGCAACACGGGCCTGCCCTCTAGCTCAGGGCACTAGCCTAGCGGTTCTATACGCTGTTAA

7702 TTGCTATTAGCAAAGTAGCCCGGGATATGGAGCAGTTGTTGGCCACCAAAAATTTGGAGAAGCCACTGGAGCAGC
     AACGATAATCGTTTCATCGGGCCCTATACCTCGTCAACAACCGGTGGTTTTTAAACCTCTTCGGTGACCTCGTCG

7777 CGGAGAATGGGTACACCTACAAGGCCACCTGGTTCATGCAGTTCCGGGCGGTCCTGTGGCGATCCTGGCTGTCGG
     GCCTCTTACCCATGTGGATGTTCCGGTGGACCAAGTACGTCAAGGCCCGCCAGGACACCGCTAGGACCGACAGCC

7852 TGCTCAAGGAACCACTCCTCGTAAAAGTGCGACTTATTCAGACAACGGTGAGTGGTTCCAGTGGAAACAAATGAT
     ACGAGTTCCTTGGTGAGGAGCATTTTCACGCTGAATAAGTCTGTTGCCACTCACCAAGGTCACCTTTGTTTACTA

7927 ATAACGCTTACAATTCTTGGAAACAAATTCGCTAGATTTTAGTTAGAATTGCCTGATTCCACACCCTTCTTAGTT
     TATTGCGAATGTTAAGAACCTTTGTTTAAGCGATCTAAAATCAATCTTAACGGACTAAGGTGTGGGAAGAATCAA

8002 TTTTTCAATGAGATGTATAGTTTATAGTTTTGCAGAAAATAAATAAATTTCATTTAACTCGCGAACATGTTGAAG
     AAAAAGTTACTCTACATATCAAATATCAAAACGTCTTTTATTTATTTAAAGTAAATTGAGCGCTTGTACAACTTC

8077 ATATGAATATTAATGAGATGCGAGTAACATTTTAATTTGCAGATGGTTGCCATCTTGATTGGCCTCATCTTTTTG
     TATACTTATAATTACTCTACGCTCATTGTAAAATTAAACGTCTACCAACGGTAGAACTAACCGGAGTAGAAAAAC

8152 GGCCAACAACTCACGCAAGTGGGCGTGATGAATATCAACGGAGCCATCTTCCTCTTCCTGACCAACATGACCTTT
     CCGGTTGTTGAGTGCGTTCACCCGCACTACTTATAGTTGCCTCGGTAGAAGGAGAAGGACTGGTTGTACTGGAAA

8227 CAAAACGTCTTTGCCACGATAAATGTAAGTCTTGTTTAGAATACATTTGCATATTAATAATTTACTAACTTTCTA
     GTTTTGCAGAAACGGTGCTATTTACATTCAGAACAAATCTTATGTAAACGTATAATTATTAAATGATTGAAAGAT

8302 ATGAATCGATTCGATTTAGGTGTTCACCTCAGAGCTGCCAGTTTTTATGAGGGAGGCCCGAAGTCGACTTTATCG
     TACTTAGCTAAGCTAAATCCACAAGTGGAGTCTCGACGGTCAAAAATACTCCCTCCGGGCTTCAGCTGAAATAGC

8377 CTGTGACACATACTTTCTGGGCAAAACGATTGCCGAATTACCGCTTTTTCTCACAGTGCCACTGGTCTTCACGGC
     GACACTGTGTATGAAAGACCCGTTTTGCTAACGGCTTAATGGCGAAAAAGAGTGTCACGGTGACCAGAAGTGCCG

8452 GATTGCCTATCCGATGATCGGACTGCGGGCCGGAGTGCTGCACTTCTTCAACTGCCTGGCGCTGGTCACTCTGGT
     CTAACGGATAGGCTACTAGCCTGACGCCCGGCCTCACGACGTGAAGAAGTTGACGGACCGCGACCAGTGAGACCA

8527 GGCCAATGTGTCAACGTCCTTCGGATATCTAATATCCTGCGCCAGCTCCTCGACCTCGATGGCGCTGTCTGTGGG
     CCGGTTACACAGTTGCAGGAAGCCTATAGATTATAGGACGCGGTCGAGGAGCTGGAGCTACCGCGACAGACACCC
```

```
8602 TCCGCCGGTTATCATACCATTCCTGCTCTTTGGCGGCTTCTTCTTGAACTCGGGCTCGGTGCCAGTATACCTCAA
     AGGCGGCCAATAGTATGGTAAGGACGAGAAACCGCCGAAGAAGAACTTGAGCCCGAGCCACGGTCATATGGAGTT

8677 ATGGTTGTCGTACCTCTCATGGTTCCGTTACGCCAACGAGGGTCTGCTGATTAACCAATGGGCGGACGTGGAGCC
     TACCAACAGCATGGAGAGTACCAAGGCAATGCGGTTGCTCCCAGACGACTAATTGGTTACCCGCCTGCACCTCGG

8752 GGGCGAAATTAGCTGCACATCGTCGAACACCACGTGCCCCAGTTCGGGCAAGGTCATCCTGGAGACGCTTAACTT
     CCCGCTTTAATCGACGTGTAGCAGCTTGTGGTGCACGGGGTCAAGCCCGTTCCAGTAGGACCTCTGCGAATTGAA

8827 CTCCGCCGCCGATCTGCCGCTGGACTACGTGGGTCTGGCCATTCTCATCGTGAGCTTCCGGGTGCTCGCATATCT
     GAGGCGGCGGCTAGACGGCGACCTGATGCACCCAGACCGGTAAGAGTAGCACTCGAAGGCCCACGAGCGTATAGA

8902 GGCTCTAAGACTTCGGGCCCGACGCAAGGAGTAGCCGACATATATCCGAAATAACTGCTTGTTTTTTTTTTTACC
     CCGAGATTCTGAAGCCCGGGCTGCGTTCCTCATCGGCTGTATATAGGCTTTATTGACGAACAAAAAAAAAAATGG

8977 ATTATTACCATCGTGTTTACTGTTTATTGCCCCCTCAAAAAGCTAATGTAATTATATTTGTGCCAATAAAAACAA
     TAATAATGGTAGCACAAATGACAAATAACGGGGGAGTTTTTCGATTACATTAATATAAACACGGTTATTTTTGTT

9052 GATATGACCTATAGAATACAAGTATTTCCCCTTCGAACATCCCCACAAGTAGACTTTGGATTTGTCTTCTAACCA
     CTATACTGGATATCTTATGTTCATAAAGGGGAAGCTTGTAGGGGTGTTCATCTGAAACCTAAACAGAAGATTGGT

9127 AAAGACTTACACACCTGCATACCTTACATCAAAAACTCGTTTATCGCTACATAAAACACCGGGATATATTTTTTA
     TTTCTGAATGTGTGGACGTATGGAATGTAGTTTTTGAGCAAATAGCGATGTATTTTGTGGCCCTATATAAAAAAT

9202 TATACATACTTTTCAAATCGCGCGCCCTCTTCATAATTCACCTCCACCACACCACGTTTCGTAGTTGCTCTTTCG
     ATATGTATGAAAAGTTTAGCGCGCGGGAGAAGTATTAAGTGGAGGTGGTGTGGTGCAAAGCATCAACGAGAAAGC

9277 CTGTCTCCCACCCGCTCTCCGCAACACATTCACCTTTTGTTCGACGACCTTGGAGCGACTGTCGTTAGTTCCGCG
     GACAGAGGGTGGGCGAGAGGCGTTGTGTAAGTGGAAAACAAGCTGCTGGAACCTCGCTGACAGCAATCAAGGCGC

9352 CGATTCGGTTCGCTCAAATGGTTCCGAGTGGTTCATTTCGTCTCAATAGAAATTAGTAATAAATATTTGTATGTA
     GCTAAGCCAAGCGAGTTTACCAAGGCTCACCAAGTAAAGCAGAGTTATCTTTAATCATTATTTATAAACATACAT

9427 CAATTTATTTGCTCCAATATATTTGTATATATTTCCCTCACAGCTATATTTATTCTAATTTAATATTATGACTTT
     GTTAAATAAACGAGGTTATATAAACATATATAAAGGGAGTGTCGATATAAATAAGATTAAATTATAATACTGAAA

9502 TTAAGGTAATTTTTTGTGACCTGTTCGGAGTGATTAGCGTTACAATTTGAACTGAAAGTGACATCCAGTGTTTGT
     AATTCCATTAAAAAACACTGGACAAGCCTCACTAATCGCAATGTTAAACTTGACTTTCACTGTAGGTCACAAACA

9577 TCCTTGTGTAGATGCATCTCAAAAAAATGGTGGGCATAATAGTGTTGTTTATATATATCAAAAATAACAACTATA
     AGGAACACATCTACGTAGAGTTTTTTTACCACCCGTATTATCACAACAAATATATATAGTTTTTATTGTTGATAT

9652 ATAATAAGAATACATTTAATTTAGAAAATGCTTGGATTTCACTGGAACTAGAATTAATTCGGCTGCTGCTCTAAA
     TATTATTCTTATGTAAATTAAATCTTTTACGAACCTAAAGTGACCTTGATCTTAATTAAGCCGACGACGAGATTT

9727 CGACGCATTTCGTACTCCAAAGTACGAATTTTTTCCCTCAAGCTCTTATTTTCATTAAACAATGAACAGGACCTA
     GCTGCGTAAAGCATGAGGTTTCATGCTTAAAAAAGGGAGTTCGAGAATAAAAGTAATTTGTTACTTGTCCTGGAT

9802 ACGCACAGTCACGTTATTGTTTACATAAATGATTTTTTTTACTATTCAAACTTACTCTGTTTGTGTACTCCCACT
     TGCGTGTCAGTGCAATAACAAATGTATTTACTAAAAAAAATGATAAGTTTGAATGAGACAAACACATGAGGGTGA
```

```
 9877 GGTATAGCCTTCTTTTATCTTTTCTGGTTCAGGCTCTATCACTTTACTAGGTACGGCATCTGCGTTGAGTCGCCT
      CCATATCGGAAGAAAATAGAAAAGACCAAGTCCGAGATAGTGAAATGATCCATGCCGTAGACGCAACTCAGCGGA
-----------------------------------------------------------------------------------
 9952 CCTTTTAAATGTCTGACCTTTTGCAGGTGCAGCCTTCCACTGCGAATCATTAAAGTGGGTATCACAAATTTGGGA
      GGAAAATTTACAGACTGGAAAACGTCCACGTCGGAAGGTGACGCTTAGTAATTTCACCCATAGTGTTTAAACCCT
-----------------------------------------------------------------------------------
10027 GTTTTCACCAAGGCTGCACCCAAGGCTCTGCTCCCACAATTTTCTCTTAATAGCACACTTCGGCACGTGAATTAA
      CAAAAGTGGTTCCGACGTGGGTTCCGAGACGAGGGTGTTAAAAGAGAATTATCGTGTGAAGCCGTGCACTTAATT
-----------------------------------------------------------------------------------
10102 TTTTACTCCAGTCACAGCTTTGCAGCAAAATTTGCAATATTTCATTTTTTTTTATTCCACGTAAGGGTTAATGTT
      AAAATGAGGTCAGTGTCGAAACGTCGTTTTAAACGTTATAAAGTAAAAAAAAATAAGGTGCATTCCCAATTACAA
-----------------------------------------------------------------------------------
10177 TTCAAAAAAAAATTCGTCCGCACACAACCTTTCCTCTCAACAAGCAAACGTGCACTGAATTTAAGTGTATACTTC
      AAGTTTTTTTTTAAGCAGGCGTGTGTTGGAAAGGAGAGTTGTTCGTTTGCACGTGACTTAAATTCACATATGAAG
-----------------------------------------------------------------------------------
                                                    5' P
10252 GGTAAGCTTCGGCTATCGACGGGACCACCTTATGTTATTTCATCATG
      CCATTCGAAGCCGATAGCTGCCCTGGTGGAATACAATAAAGTAGTAC
-----------------------------------------------------------------------------------
```

Appendix E:

Drosophila PDZ Domain-Containing Proteins

FB2010_03, released March 19th, 2010

Genes

Home Tools Files Species Documents Resources News Help Archives Jump to Gene Gc

72 matches | Show related Genes | Results Analysis/Refinement | HitList Conversion Tools

☑	#	↓Symbol↑	↓Name↑	↓Annotation ID↑	↓Cytology↑	↓Alleles #↑	↓Stocks #↑	↓Clones #↑
☑	1	a	arc	CG6741	58C1-58C5	35	21	19
☑	2	baz	bazooka	CG5055	15F1-15F1	54	16	18
☑	3	bbg	big bang	CG42230	70E1-70D7	39	29	132
☑	4	Caki	Calcium/calmodulin-dependent protein kinase	CG6703	93F10-93F12	12	8	105
☑	5	CG3402	-	CG3402	61C8-61C8	1	2	39
☑	6	CG5921	-	CG5921	5D3-5D3	1	2	-
☑	7	CG6498	-	CG6498	71E1-71E1	3	4	29
☑	8	CG6509	-	CG6509	32F2-32F2	7	6	48
☑	9	CG6619	-	CG6619	65A5-65A5	3	2	9
☑	10	CG6688	-	CG6688	94D9-94D9	3	4	1
☑	11	CG9588	-	CG9588	87F7-87F7	4	4	32
☑	12	CG10362	-	CG10362	10F7-10F7	1	1	1
☑	13	CG12187	-	CG12187	62E9-62F1	3	3	11
☑	14	CG14168	-	CG14168	67C8-67C8	4	4	5
☑	15	CG15617	-	CG15617	53D5-53D5	2	2	2
☑	16	CG15803	-	CG15803	91A3-91A3	1	2	6
☑	17	CG31158	-	CG31158	94B5-94B6	8	7	61
☑	18	CG31304	-	CG31304	88E1-88E1	2	2	22
☑	19	CG31342	-	CG31342	87D7-87D7	1	2	44
☑	20	CG31772	-	CG31772	24C3-24C3	2	2	9
☑	21	CG32758	-	CG32758	5B2-5B3	7	5	92
☑	22	CG34375	-	CG34375	94D8-94D9	4	3	9
☑	23	CG34400	-	CG34400	70C5-70C5	10	10	34
☑	24	CG34404	-	CG34404	88F1-88F1	5	5	20
☑	25	CG42319	-	CG42319	49F14-49F15	4	4	44
☑	26	cnk	connector enhancer of ksr	CG6556	54B7-54B7	104	5	26
☑	27	cno	canoe	CG42312	82F4-82F6	49	21	65
☑	28	dlg1	discs large 1	CG1725	10B6-10B10	81	18	26
☑	29	dsh	dishevelled	CG18361	10B4-10B5	107	19	1
☑	30	Gef26	Gef26	CG9491	26C2-26C3	22	6	48
☑	31	Grasp65	Grasp65	CG7809	76D5-76D5	2	1	17
☑	32	Grip	Glutamate receptor binding protein	CG14447	5D1-5D1	32	4	3
☑	33	HtrA2	HtrA2	CG8464	88C3-88C3	9	5	5
☑	34	in	inturned	CG16993	77B6-77B6	35	76	2
☑	35	inaD	inactivation no afterpotential D	CG3504	59B3-59B3	19	1	62
☑	36	kermit	kermit	CG11546	44B5-44B5	10	7	56
☑	37	Lap1	Lap1	CG10255	51D4-51D4	6	5	38

☑	38	LIMK1	LIM-kinase1	CG1848	11B2-11B2	16	10	3
☑	39	loco	locomotion defects	CG5248	94B6-94B8	34	10	63
☑	40	Magi	Magi	CG30388	57C2-57C3	6	4	30
☑	41	Mhcl	Myosin heavy chain-like	CG31045	89B2-89B3	6	4	102
☑	42	par-6	par-6	CG5884	16C1-16C1	12	5	172
☑	43	Patj	Patj	CG12021	62B4-62B4	15	2	87
☑	44	PDZ-GEF	PDZ-GEF	-	-	-	-	-
☑	45	Pdz	Pdz	-	-	-	-	-
☑	46	PICK1	PICK1	CG6167	33F3-33F3	3	2	126
☑	47	Prosap	Prosap	CG30483	50D3-50E1	17	16	358
☑	48	PsGEF	Protostome-specific GEF	CG14045	3A2-3A2	6	4	1
☑	49	Ptpmeg	Ptpmeg	CG1228	61C1-61C1	40	25	137
☑	50	pyd	polychaetoid	CG31349	85B2-85B7	51	26	152
☑	51	rab3-GEF	rab3-GEF	CG5627	13A11-13A12	9	8	3
☑	52	RhoGAP19D	RhoGAP19D	CG1412	19D1-19D2	11	8	6
☑	53	RhoGAP100F	RhoGAP100F	CG1976	100D2-100D3	9	7	48
☑	54	RhoGEF2	RhoGEF2	CG9635	53E4-53F1	33	15	63
☑	55	Rhp	Rhophilin	CG8497	13E18-13E18	2	1	1
☑	56	Rim	Rim	CG33547	90C9-90C10	9	9	37
☑	57	scrib	scribbled	CG42614	97B9-97C2	48	20	167
☑	58	sdt	stardust	CG32717	7D17-7D22	53	19	20
☑	59	sif	still life	CG34418	64E1-64E5	18	16	92
☑	60	Sip1	SRY interacting protein 1	CG10939	54B6-54B7	10	9	58
☑	61	skf	skiff	CG30021	47E1-47E1	4	5	34
☑	62	Slip1	Slip1	CG1783	102D4-102D4	3	3	2
☑	63	Spn	Spinophilin	CG16757	62E4-62E5	16	10	31
☑	64	sprt	sprite	CG30023	47F8-47F8	8	7	24
☑	65	Syn1	Syntrophin-like 1	CG7152	78E6-78F1	8	5	29
☑	66	Syn2	Syntrophin-like 2	CG4905	53C7-53C7	4	3	22
☑	67	vari	varicose	CG9326	39D3-40B1	29	13	123
☑	68	veli	veli	CG7662	96B19-96B19	5	4	83
☑	69	X11L	X11L	CG5675	16B7-16B7	8	4	60
☑	70	X11Lβ	X11Lβ	CG32677	9D4-9E1	9	7	56
☑	71	Zasp52	Z band alternatively spliced PDZ-motif protein 52	CG30084	52C4-52C7	17	14	113
☑	72	Zasp66	Z band alternatively spliced PDZ-motif protein 66	CG6416	66D9-66D9	6	4	307

Contact FlyBase **version FB2010_03, released March 19th, 2010** Site Map