

**Characterization of chromosomal mutations in human embryonic stem cells  
that fail to differentiate**

A Major Qualifying Project Report

Submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

in

Biology and Biotechnology

Submitted by:

---

Arba Karcini

April 26, 2017

APPROVED:

---

Tanja Dominko, Ph.D.

WPI Project Advisor

# Abstract

Chromosomal mutations lead to abnormal karyotypes and multiple genetic disorders in a variety of mammalian species. During extended cell culture and proliferation, human stem cells have demonstrated a predisposition to aneuploidy of chromosomes 12, 17, and X (Baker et al., 2007) and deletions or duplications of regions on Ch18 and Ch20 (Spits et al., 2008). Here we describe two additional chromosomal abnormalities (deletions on Ch10 and loss of heterozygosity on Ch5) associated with a female human embryonic stem cell H9 line that failed to undergo induced differentiation. Cytogenetic studies detected amplifications, deletions and translocations in multiple chromosomes. In an attempt to characterize these mutations, remapping, sequencing, and gene ontology was performed. Here we identify and analyze protein coding genes localizing to mutated regions.

# Table of contents:

Abstract.....	2
Table of Contents.....	3
Acknowledgements.....	4
1.0 Introduction.....	5
2.0 Background.....	7
3.0 Methodology.....	12
4.0 Results.....	20
5.0 Discussion.....	36
6.0 Future Studies.....	38
7.0 References.....	40
8.0 Appendices.....	45

# Acknowledgments

I would like to thank Professor Dominko for allowing me to complete this project in her lab and for supporting me throughout the completion of this project. She has been a positive and encouraging advisor whose assistance and recommendations have been invaluable. I would also like to thank David Dolivo for his interest and technical support during the completion of this project.

# 1.0 Introduction

Human embryonic stem cells (hESCs) have the ability to differentiate into all three germ layers promising novel approaches to clinical researches and genetic diseases hoping to develop a genetic system for human gene functions as well as come one step closer to personalized medicine (Abyeta et al., 2004). As promising and exciting as such approaches seem, working and maintaining stem cell lines is complex. hESCs that are used in most research are approved by the National Institute of Health (NIH) through a process that tests among others for contamination, gene expression during self-renewal and differentiation, ability to adapt in uniform cultures, growth efficiency, and karyotype stability (Ware, Nelson, Blau, 2006). There were only 22 NIH approved cell lines in August 2006; however there are 384 NIH approved cell lines in April 2017 (NIH Human Embryonic Stem Cell Registry, 2017), meaning that in ten years the amount of research dedicated to the stem cells has massively increased, and so has the derivation of new cell lines. But, more cell lines means more chances of mutations and other random aberrations making the maintenance of *in vitro* stem cell lines challenging. Even though cell lines are tested for stability and uniformity in cell culture before NIH approval, extensive passages lead to mutations. Several studies have reported random acquisitions of chromosomal abnormalities in cells after prolonged culture (Baker et al., 2007) or cell dissociation technique during passaging methods (Tosca et. al., 2015). These karyotypically abnormal cells that exhibit increased growth are also found to be associated with reduced apoptosis and differentiation capacity leading to tumorigenesis *in vivo* (Baker et al., 2007). Hence, to support novel cell therapies and safe use of hESCs, research on characterizing common mutations among cell lines and identifying genes

that may be linked to their transformation, as well as research on optimal conditions of *in vitro* culture of hESCs are of special importance to the researchers wishing to advance cell therapy.

This project summarizes and presents novel mutations of the WA09/H9 stem cell line cultured and passaged *in vitro*. Even though being one of the most popular stem cell lines, H9 line possesses a tendency to develop mutations even under expert handling and culturing (Ware, Nelson, and Blau, 2006). The objective was to characterize the karyotype of this cell line that failed to differentiate and to map key genes altered by chromosomal mutations. To achieve the goal, cytogenetic analysis was performed and bioinformatics approaches were used to remap, sequence and classify genes included in mutated regions based on their molecular function. Further, protein identification and interaction analysis were performed to investigate possible mechanistic consequences. The use of this cell line can provide more insight in the developmental biology field, as well as help in the advancement of drug discovery or transplantation medicine practices.

Identifying and analyzing genes localized to mutated chromosomal regions of hESCs will contribute to the scientific knowledge of their role in cell transformation and will allow for the advancement of anti-cancer therapies.

## 2.0 Background

### 2.1 WA09/H9 embryonic stem cell line

H9 embryonic stem cell line is the most used and cited (more than 550 publications) in scientific research and is referred to as the “gold standard” due to its stability (WiCell, 2012).

H9 cells are pluripotent and are able to differentiate in to all three germ layers (mesoderm, endoderm and ectoderm). Several publications have documented the pluripotency of this cell line. This line was derived from a normal female embryo with a normal karyotype 46XX by Thomson Lab in Madison, Wisconsin (Thomson et. al., 1998). Distribution of the cell line is now managed by WiCell Research Institute, Madison, WI (International Stem Cell Registry, 2017). These human blastocyst-derived cells proliferated *in vitro* for four to five months and maintained the ability to form trophoblast and all the three embryonic layers, and were the first approved stem cell line (Thomson et. al., 1998).

Prolonged cell culture of H9 cells drives selection of variant cells that cause genetic and epigenetic changes to the stem cell line that prevent their differentiation and do not drive apoptosis (Desmarais et. al., 2012), indicating their transformation.

### 2.2 Commonly reported mutations

Embryonic stem cell lines are prone to mutations under extensive culturing (Baker et. al., 2007). Common mutations have been characterized across different cell lines in chromosome 12, 17, 18, 20 and X (Baker et. al., 2007; Spits et. al., 2008). Among the most common mutations is the

submicroscopic amplification of q arm in Ch20 that is considered a hot spot (Spits et. al., 2008). This genomic mutation was found consistent among several lines and 23 common genes were identified that were linked with oncogenic transformation (Lefort et. al., 2008). Gain of 20q11.21-q11.22 region is a highly reported mutation among hESC lines (Avery et. al., 2013; Nguyen et al., 2013) linked with malignant transformation. Overexpression of BCL2L1 gene is identified to be a driver mutation due to growth advantage it provides to the cell (Avery et. al., 2013). Other candidate genes for the mutated region of Ch20 were identified to be ID1 and HM13 (The International Stem Cell Initiative, 2011).

## **2.3 Mutations on chromosome 5**

A few studies report a loss of heterozygosity on chromosome 5 (Ch 5) in human cells which has been related to several carcinomas. Studies conducted on somatic cells have observed that 5q loss of heterozygosity is linked with the FAP (familial polyposis coli) gene location and has been associated with colorectal cancers (Solomon et. al., 1987).

Later studies in tumor cells (lung carcinoma) have identified several regions of chromosome 5q deletion or loss of heterozygosity for tumor suppressor genes in q11.2-q12.2, q22.3-q23.1, q31.1 and q35.3 regions. Chromosome 5 is also a frequent target of genomic mutations in other cancers such as ovarian, gastric, or myeloid diseases (Mendes-da-Silva et. al., 2000).

In patients with osteosarcoma, CHG analysis revealed several gains and losses, among which significant gains of chromosome 17 and loss of chromosome 5q were detected. Patients carrying a loss of 5q had a better prognosis. Data also indicated that the loss of 5q region may have an opposite effect to the overexpression of some oncogenes (Ozaki et. al., 2002). In malignant



glioma cell lines, cyclin B1 (CCNB1) and cyclin H (CCNH) have been identified as candidate genes due to exhibited increased growth and their location on 5q region (Weber et. al., 2000).

## **2.4 Mutations on chromosome 10**

Mutations on chromosome 10p15.3 (duplication) have been associated with development delay and intellectual disability due to greater copy number variants (CNVs) and several candidate genes associated with such mutations and disorders (Cooper et. al., 2011).

Chromosomal microarray assays (CMA) performed in individuals with deletions on the 10p15.3 region identified two genes (ZMYND11 and DIP2C) that localize to the deleted region. The 10p15.3 deletion was found to be mostly of a *de novo* nature and to be associated with several cognitive and development disorders such as speech delay, motor delay, brain anomalies and seizures (DeScipio et. al., 2012).

Further studies depict ZMYND11 gene as a tumor suppressor gene responsible for transcription elongation of RNA polymerase II. This activity is specifically regulated through recognition of H3K36me3 (histone variant H3.3 lysine 36 trimethylation) by ZMYND11 on H3.3. Such recognition is critical in regulation of chromatin, hence ZMYND11 plays a key role in transcriptional repression. Reduced levels of ZMYND11 lead to worse prognosis in breast cancer patients; and increased levels of ZMYND11 lead to suppression of cancer cells *in vitro* (Wen et. al., 2014). The interplay between ZMYND11 and H3K36me3 has also been described in regulation of pre-mRNA processing, specifically splicing events (Guo et. al., 2014).

## **2.5 Mutations on chromosome 17**

17q amplification mutation have been reported on previous studies (Abel et al., 1999; Varis et al., 2002). GTP binding proteins located on chromosome 17 (q25 region) have been identified as candidate genes in cases of chromosome 17 mutations related to ovarian and breast tumors.

These protein coding genes belong to the septin family and are involved in cytokinesis (McIlhatton et. al., 2001). The gains of 17q25.3 region have been identified through CGH analysis in human tumor cells and septin genes have been proposed as candidate gene.

Specifically, Sept9 overexpression was shown to play a role in breast tumorigenesis through transfection with small interference RNA (Montagna et. al., 2003).

In breast cancer studies, genomic alterations of chromosome regions in 17q have been reported (Shadeo & Wan, 2006). Amplifications of chromosome 17 regions reported by several studies usually corresponded with gains or mutation in other chromosomes. Since these gains are associated with neuroblastomas, in addition to breast cancer cases, BIRC5 located in 17q25 region is suspected as a potential candidate gene (Baker et. al., 2007).

Amplifications of 17q region and their relation with neuroblastomas has been established. For example, candidate genes were identified for 17q23 region, but not for the distal 17q region (Saito-Ohara et. al., 2003). Among other candidates on the amplified region of q arm of chromosome 17, FISH (fluorescence *in situ* hybridization) analysis revealed BIRC5 as a strong candidate gene in malignant peripheral nerve sheath tumor (Storlazzi et. al., 2006).

## **2.6 Non-coding genes and miRNAs mutations**

MicroRNAs (miRNAs) regulate gene expression at post-transcriptional level, hence being a target for treating neurological disorders (Sun, Liang & Pei, 2015). They are part of mRNA degradation process and responsible for inhibition of translation; they are linked with several biological processes such as cell proliferation, apoptosis, tumorigenesis and development. Micro RNAs were identified in chromosomes 7, 8, 11, and 17 and were associated with the mutations in these regions in breast cancer cell lines (Riaz et. al., 2013). Non-coding specific micro RNAs have been linked with corresponding mutated genes in thyroid carcinoma (Puppin et. al., 2014) and colon cancer (Nosho et. al., 2013), suggesting that miRNAs can be diagnostic biomarkers of cancer and key molecules in stem cell transformations.

## 3.0 Methods

### 3.1 Embryonic Stem Cell Culture

WA09/H9 stem cell line was obtained WiCell® and the WiCell Feeder Based (MEF) Pluripotent Stem Cell Protocol (SOP number: SOP-SH\_001) was followed for maintaining this cell line in culture.

hESCs were cultured on in-house, primary derived mouse embryonic fibroblasts (MEF) which was plated at a density of  $2.25 \times 10^4$  cells/cm<sup>2</sup> on plates that were coated for 60 minutes with 1% gelatin solution (MP Biologicals). MEF were inactivated with mitomycin-C (Sigma-Aldrich) for 3 hours before plating (Kole et. al., 2017).

Cells were grown in Stem Cell Culture Medium containing DMEM (Life Technologies), 15% Knockout serum (Life Technologies), 2 mM Glutamax (Life Technologies), 50 mM 2-mercaptoethanol (Sigma-Aldrich), 1× MEM non-essential amino acids (Mediatech) and 8 ng/mL FGF2 (Peprotech) (Kole et. al., 2017). After plating, cultures were incubated at 37 °C, 5% O<sub>2</sub>, 5% CO<sub>2</sub> and high humidity. hESCs were allowed to attach for a day and then daily feeding was performed through media replacement. Passaging procedures were performed after five days of growth. To dissociate the cells 0.01% trypsin (Mediatech) was used, then hESCs were split 1:12 into new MEF feeder layers that were prepared as described above (Kole et. al., 2017).

### 3.2 G-banded karyotyping

G-banded karyotyping is an optimized chromosome analysis for mouse and human pluripotent stem cells. A G-band karyotyping was used to assess and monitor the genomic stability of H9 stem cell line after 28 passages. A karyotyping assay detects microscopic genomic abnormalities (5-10Mb) such as inversions, duplications, deletions, aneuploidies, and balanced and unbalanced translocations (Cytogenetic Services WiCell, 2012). To prepare the cells for G-banded karyotyping, cells in fixative were diluted and 20 well-spread metaphases were produced. During evaporation of the fixative, the surface of the slide becomes grainy and the cells become visible. The slide was facing a hot water bath (75°C) for up to three seconds and then was dried in a hot metal plate, whose temperature regulated chromosome spreading (Henegariu et. al., 2001). After overnight incubation at 65°C, G-banding assay was performed at 425-525 resolution.

### **3.3 CGH+SNP microarray**

Comparative Genomic Hybridization (CGH) and Single Nucleotide Polymorphism (SNP) microarray was utilized (Agilent® dual color array with 60-mer oligonucleotide probes) to detect submicroscopic abnormalities smaller than 5Mb such as duplications, deletions, and unbalanced translocations. It can also distinguish loss of heterozygosity (LOH) and absence of heterozygosity (AOH), but it cannot detect balanced translocations or inversions, or the chromosomal positions of the genomic gains. When paired with a G-banded karyotype, it can define the translocation breakpoints (Cytogenetic Services WiCell, 2012). CGH utilizes *in situ* hybridization of metaphase spreads with two different genomic DNAs, one retrieved from the stem cell line (target) and one retrieved from a karyotypically normal reference. The two DNAs are labeled with different fluorochromes: green and red and are mixed 1:1 with excess of Cot 1

DNA (Bryndorf et. al., 1995). First, high-molecular-weight genomic DNA (both target and reference) was isolated from cell lysates and metaphase spreads were prepared. Before hybridization, target DNA was denatured and then mixed with the reference DNA, unlabeled Cot 1 DNA. Metaphase spreads dry slides were hybridized at 37°C for 3 days (Bryndorf et. al., 1995). Under the microscope, metaphase images were captured in two colors and ratios were determined. Regions that appear outside the -1 and +1 rate represented deletions and amplifications, respectively.

### **3.4 Remapping**

NCBI Remap is an online tool used to project data from one coordinate system to another through the use of genomic alignments. A base-by-base analysis was performed to the source sequence (the sequence retrieved for this study) to present the same features in the new sequence (NCBI, 2017). NCBI offers three variations of Remap, but in this study only the Assembly-Assembly remap feature was used to remap the features of hg19 assembly to the hg38 assembly. First, the organism of interest was selected, then the source assembly (the assembly the features are on) was chosen from the given list (hg19) and finally the target assembly (the assembly where the features will be projected) was selected (hg38). During remapping, the minimum ratio of bases was at default 0.5 meaning that at least 50% of the interval was able to be remapped. The maximum ratio for the difference between source and target length was at default 2 meaning that the algorithm allowed for insertions and deletions. This ratio was calculated as the ratio of interval lengths between target and source assembly.

Maximum ratio = target assembly (stop-start +1)/ source assembly (stop-start +1)

An insertion or deletion in the target assembly would affect the ratio.

Also, two other options of the NCBI Remap (the merge option and multiple location return option) were enabled to improve coordinate projections in case of duplications or large insertions in the sequence (NCBI, 2017).

### **3.5 Gene Ontology and Gene Analytics online database**

#### **Gene Ontology**

Gene Ontology (GO) emerged as a bioinformatics initiative to represent and annotate the gene functions at a molecular, cellular and tissue level. This computational approach describes gene products through a fixed vocabulary system which offers consistency genomic descriptions as well as serves as a tool to process data (Gene Ontology Consortium, 2017). Gene ontology is species-independent ontology performed in three structural aspects regarding biological processes, cellular components and molecular functions. It includes the majority of genomes for plants, animals and microbial species. Despite the throughout analysis of genomes, Gene Ontology does not include several areas such as protein domains, protein-protein interactions, or environment, evolution and expression. It also does not contain specific terms such as oncogenesis to describe functions or components that are unique to mutants because cancer is not the normal function of a gene. However, an ontology offers insight on the relations between a set of terms in the perspective of a biological function, molecular function or cellular component. In a simple representation, a gene ontology can be given as a graph where each node represents a GO term and edges represent the relationship between nodes. In a GO graph the more general terms will be up and act as “parent” terms for the more specialized “child” terms that can be

found toward the bottom part of the graph (Gene Ontology Consortium, 2017). The three ontologies of GO are briefly explained below:

### Cellular Component

This ontology describes the component of a cell and location in a subcellular and macromolecular complex level. It includes protein complexes and multi-subunit enzymes, but not individual proteins (Gene Ontology Consortium, 2017).

### Biological Process

This ontology describes series of biological events/processes as a result of one or more of the assemblies of molecular functions, however they do not present a pathway. During the collection process there is a definite beginning and end for each biological process which can be either part of or an entire process of a parent node (Gene Ontology Consortium, 2017).

### Molecular Function

This ontology describes activities of usually individual genes at the molecular level such as binding or transporting activity. The function described by the ontology is unambiguous independent of species (Gene Ontology Consortium, 2017).

### **GeneAnalytics**

GeneAnalytics is a gene set analysis tool provided by LifeMap Discovery®, a database for stem cell research supported by extensive molecular, cellular and medical data from scientific literature (LifeMap Discovery, 2017). Gene Analytics provides a GO analysis feature whose information is in integration with GeneCards, the human gene database. In addition to the three domains of GO, as described in the above section, GeneAnalytics provides additional



information on diseases and pathways, as well as simpler interpretation of results in regard to the provided gene set. The user-friendly features of this database make searching and data input and output easy to navigate.

The data (gene set) was pasted on the search box and the organism of choice was selected (homo sapiens); in a few minutes a complete analysis with external links and explanations was available. Since a free trial was used during this study, not all of the features of the program were available. However, GeneAnalytics is a broad tool that provided information in several fields such as diseases, pathways, ontologies and phenotypes that were involved in the gene set of interest.

### **3.6 Protein databases and resources**

Since the tools utilized so far in this study did not provide much information on the protein structure and interaction between the proteins retrieved from the protein coding genes, two more tools were utilized: Chimera USCF and STRING.

#### **UCSF Chimera**

UCSF Chimera is a modeling system used for interactive visualization and analysis of molecular structures. It is a free tool for download and it is associated with several tutorial videos to describe the functions and potentials of this tool. Several types of data can be processed in UCSF Chimera such as 3D density maps, light microscopy data, and protein or nucleic acid sequences and annotations (RBVI, 2016). In this study protein annotations were the primary data type used. The main goal of this software is to visually represent and analyze molecular assemblies, protein structure and function, binding interactions and protein network. Researchers use this program to annotate proteins of unknown function, identify targets for drug development as well as be able

to engineer and design new proteins (RBVI, 2016). In this study PDB annotations of proteins were used to retrieve a high quality image of the protein structure as well as to identify the binding domains and the protein interaction sites.

First, Fetch structure by ID option was chosen from the File command on the top bar in order to insert a structural image of the desired protein. Through the Action command, colors, strands and other features of the protein were changed. Through the Tools command, selecting Structure Analysis and Surface/Binding Analysis provided more information on the binding sites of the selected protein. The final image was saved as a picture file.

## **STRING**

STRING is an online database created by several research European institutions that provides known and predicted information on protein-protein interactions including physical and functional interactions. The five sources used by STRING database to retrieve this information are genomic predictions, high-throughput lab studies, co-expression data, automated text mining, and previous knowledge of databases. There are currently more than nine million proteins whose information is available on STRING (STRING, 2015). At the start page, the protein annotation and organism was selected. Different format for data input is available such as protein name or amino acid sequence. Then, a network view was shown which summarized basic information of the protein interactions in a user friendly way. Each node of the network contained detailed information on the respective protein. The confidence of each interaction was shown in a separate table under the Legend tab. Data setting tab allowed for change of the parameters such as the type of evidence used for the prediction. In this study, all five sources were used to retrieve data. The user can also choose the number of interactions that will be displayed, but a default of 10 best scoring hit interactions was chosen in this study.

There are many options available through the STRING website based on the analysis a researcher aims to perform. In this study basic features were used to give a representation of the network in which proteins interact with each other.

## 4.0 Results

Karyotyping and microarray tests provide the starting data, which are further analyzed through a series of online databases in order to spot individual candidate genes that may play a key role in the stem cell differentiation process. A karyotype containing random gained mutations was the starting point of this study. The karyotype represented the cell of a human embryonic stem cell line (WA09-H9-DL4) of a female line on its 28<sup>th</sup> passage. After these many passages the hESCs failed to differentiate, therefore a chromosome analysis report was utilized to reveal the genomic background of such non-differentiation. 20 metaphase chromosome spreads were examined. Below are presented the G-banded karyotyping results analysis and the CGH+SNP microarray results which depict the region of the mutation and the size of the mutation.

### 4.1 G-banded karyotyping and CGH+SNP microarray

Four different abnormalities were detected and presented below, where the first number shows the number of chromosomes (46), the sex (XX), type of abnormality, chromosome, and region of the chromosome:

A) 46, XX, add(20)(q11.2)

Five cells had an unbalanced 20q arm that shows gain (add for addition) of 20q11.21-q11.22 region of approximately 4.2Mb

B) 46, XX, der(10)t(10; 17)(p13; q11.2)

Two cells had translocations (der for derivation) between short arm (p) of Ch10 and long arm (q) of Ch17 resulting in loss in 10p13-p15.3 region of 14.76 Mb and gain in 17q21.31-q25.3 region of 38.59 Mb.

C) 46, XX, der(10)t(10;17)(p13;q11.2), add(20)(q11.2)

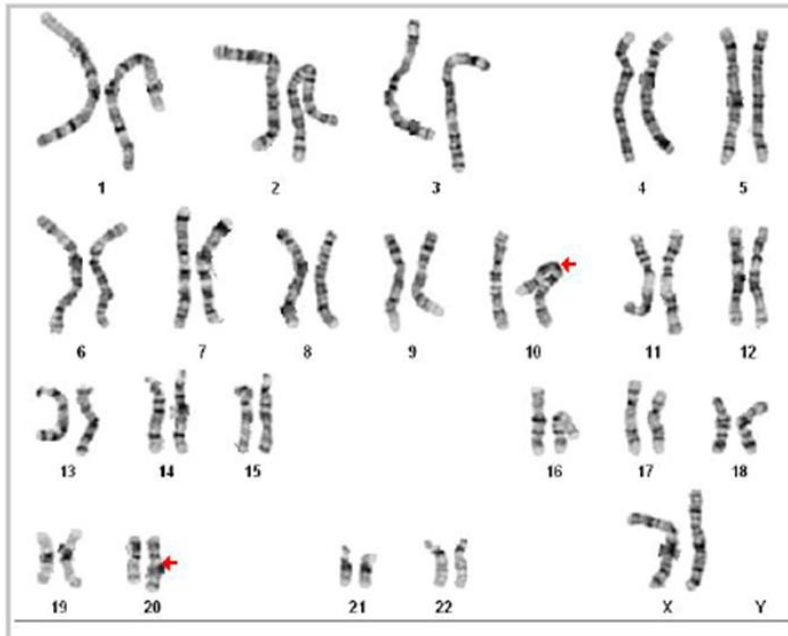
Eight cells displayed both of the abnormalities mentioned above, translocation between Ch10 and Ch17 and amplification of Ch20.

D) 46, XX, del(6)(p23), add(20)(q11.2)

Three cells displayed deletion (del) of p23 region in Ch6 accompanied with the amplification of Ch20.

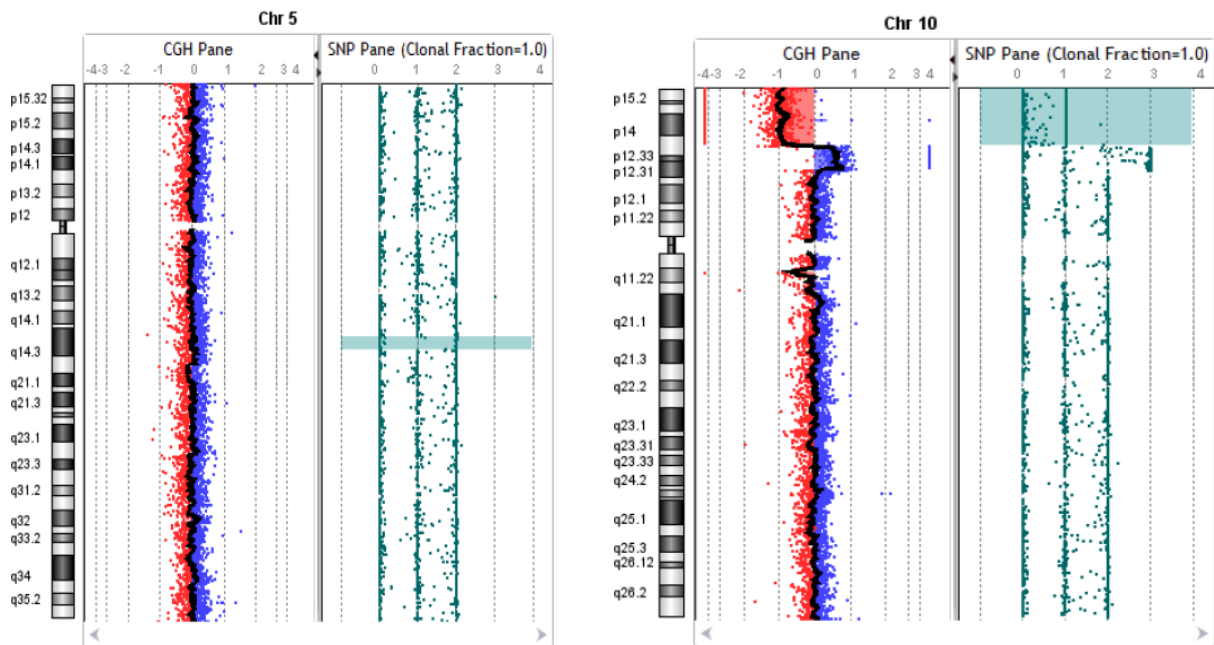
Two cells out of 20 did not have any detectable mutations. The above presented results were supported by the CGH+SNP microarray, which reported another amplification in the 10p12.31-p13 region of 6.06 Mb and a loss of heterozygosity greater than 3Mb in Ch5.

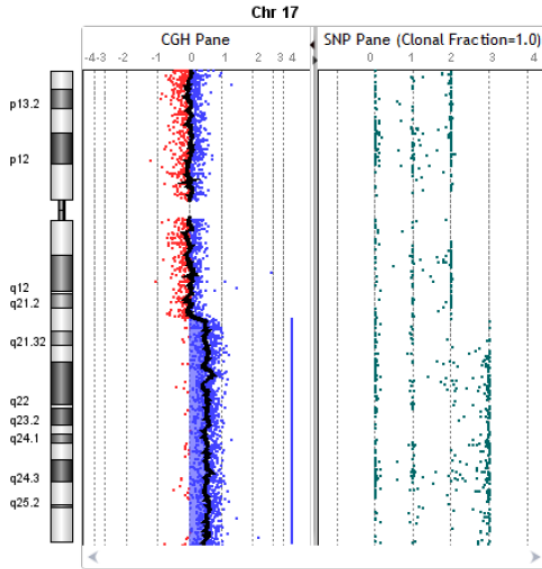
A summary of the G-banded karyotyping assay is presented in Figure 1, below.



**Figure 1:** Abnormal karyotype of WA09/H9 cell line of a human embryonic stem cell. The red arrows show two of the four aberrant clones: the loss in the short (p) arm of chromosome 10, and the unbalanced structural aberration gain in the long (q) arm of chromosome 20.

CGH+SNP microarray used to detect submicroscopic abnormalities presented all chromosomes in the dual color system as described in Section 3.3 and Figure 2 shows three of the chromosomes: Ch5, Ch10, Ch17.





**Figure 2:** CGH+SNP microarray result depicting the loss of heterozygosity on chromosome 5 at the region q14.3 as highlighted by the blue bar, the deletion on chromosome 10 at the region p15.3 and the amplification on chromosome 17 at the region q21.31.

For a full view of the CGH+SNP microarray results refer to Appendix 1. All the mutations identified on each affected chromosome of H9 cell line sample are included in Table 1.

**Table 1:** Chromosomes (Ch) with abnormalities and their specific region of mutation, size of the region in base pairs (bp), start and stop positions, and the type of mutation where AMP indicates amplification, DEL indicates deletion, and LOH indicates loss of heterozygosity. The data is presented in UCSC hg19 reference system.

<i>Ch</i>	<i>Region</i>	<i>Size(bp)</i>	<i>Start</i>	<i>Stop</i>	<i>Type</i>
1	q31.3	54582	196,744,721	196,799,302	AMP
2	p11.2	115913	89,185,302	89,301,214	AMP
4	p15.1	22133	34,787,128	34,809,260	DEL
4	q13.2	90702	69,392,576	69,483,277	DEL
5	q14.3	4463694	85,134,352	89,598,045	LOH

6	q14.1	44168	78,979,161	79,023,328	DEL
7	q11.21	617325	62,073,606	62,690,930	AMP
8	p11.22	122621	39,258,894	39,381,514	DEL
10	p15.3 - p13	14760322	136,361	14,896,682	DEL
10	p13 - p12.31	6064374	14,951,987	21,016,360	AMP
10	q11.22	742031	46,949,255	47,691,285	DEL
12	p13.31	76103	9,637,323	9,713,425	DEL
14	q23.2	295009	63,502,514	63,797,522	AMP
14	q23.3	398346	66,006,698	66,405,043	AMP
14	q32.33	397605	106,405,703	106,803,307	AMP
14	q32.33	362780	106,852,114	107,214,893	AMP
16	p11.2	1628866	31,934,834	33,563,699	DEL
17	q21.31 - q25.3	38592567	42,443,540	81,036,106	AMP
20	q11.21 - q11.22	4206245	29,842,786	34,049,030	AMP
X	p22.33	2624974	60,701	2,685,674	AMP
X	q28	258241	154,974,667	155,232,907	AMP

In addition to previously reported mutations, new or less common mutations were found in stem cells such as the deletion of 10p15.3 region or amplification of 10p12 region, as well as the loss of heterozygosity (LOH) in Ch5.

Next we examined chromosome 5, 10 and 17 as the largest mutations observed in this stem cell line.



## 4.2 Remapping

Since the results of the above tests are presented in hg19 assembly, an up to date presentation of data is retrieved through remapping into hg38 assembly. Below are presented the summary tables built from the information retrieved by NCBI Remap program.

### Chromosome 5

	<i>Source</i>	<i>Mapped</i>
<i>Id</i>	chr5	chr5
<i>Length</i>	4463694	4463695
<i>Start</i>	85134352	85838534
<i>Stop</i>	89598045	90302228
<i>Strand</i>	+	+
<i>Coverage</i>		1

### Chromosome 10

	<i>Source</i>	<i>Mapped</i>
<i>Id</i>	chr10	chr10
<i>Length</i>	147660322	14764263
<i>Start</i>	136361	90421
<i>Stop</i>	14896682	14854683
<i>Strand</i>	+	+
<i>Coverage</i>		1.00027

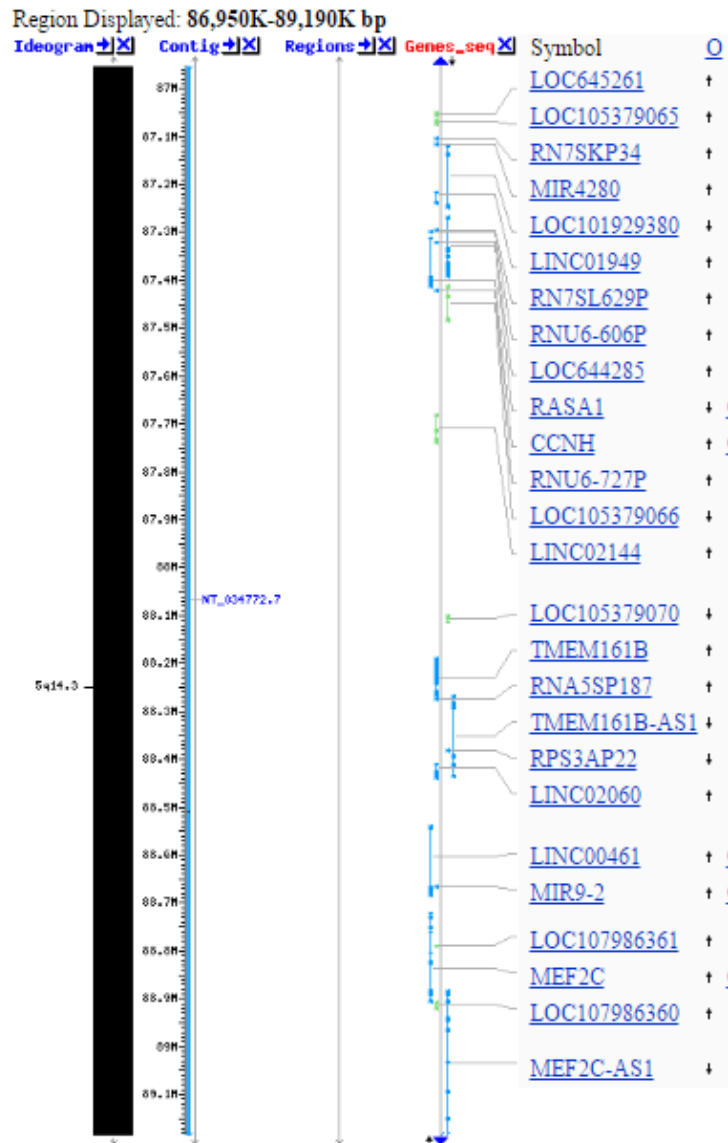
## Chromosome 17

	<i>Source</i>	<i>Mapped</i>
<i>Id</i>	chr17	chr17
<i>Length</i>	38592567	43083888
<i>Start</i>	42443540	36449217
<i>Stop</i>	81036106	79533104
<i>Strand</i>	+	+
<i>Coverage</i>	1.11638	

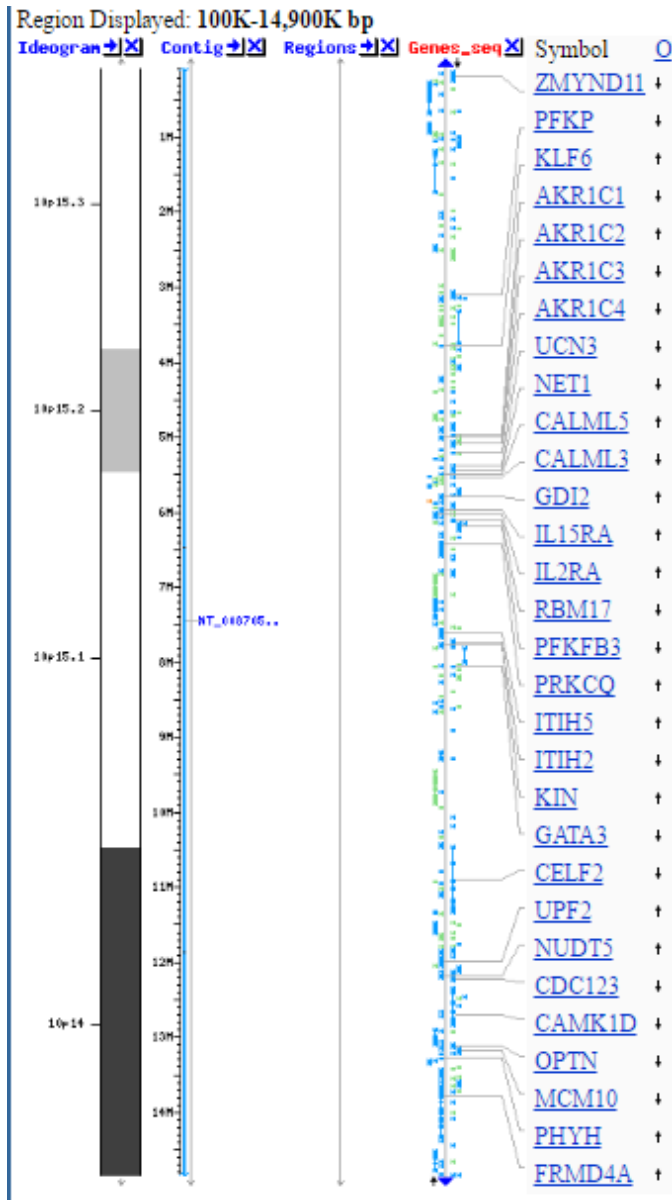
Since chr 17 had an amplified region a second pass was performed, hence the mapped region contained the duplication and had a greater length.

### **4.3 Gene Mapping and Gene Ontology**

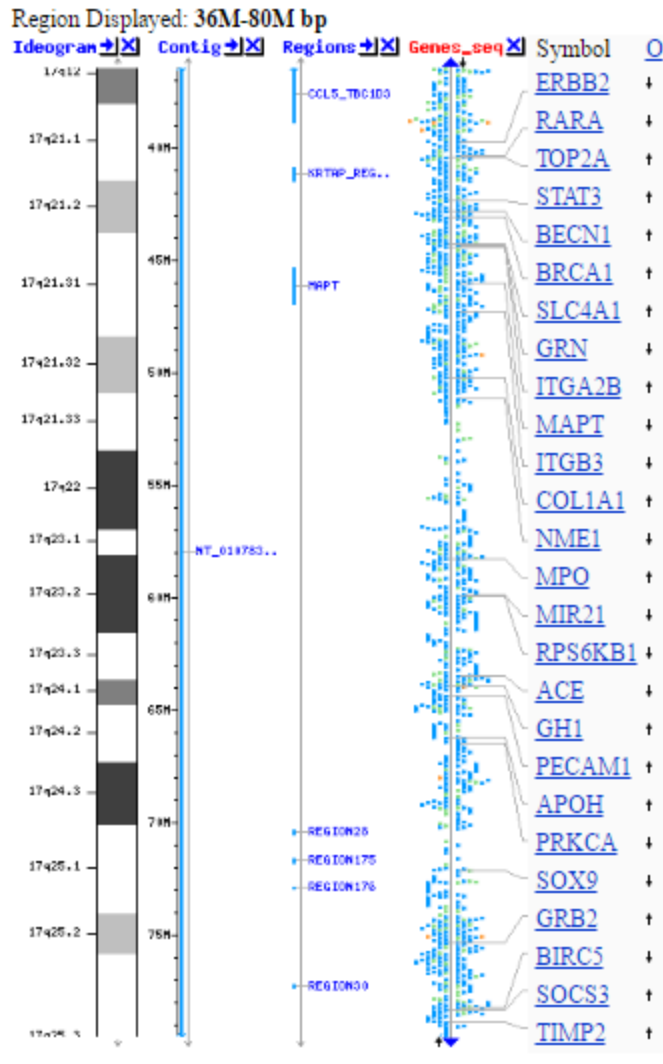
Using the new coordinates presented in the above section, a map of the mutated regions is retrieved together with the genes localized in these regions. Figures 3, 4 and 5 present the map of chromosome 5, 10 and 17, respectively.



**Figure 3:** 5q14.3 mutated region, the respective contig, genes and their symbols (names) and orientation strand (O) where the up arrow represents negative strand and the down arrow represents positive strand.



**Figure 4:** 10p15.3 mutated region, the respective contig, genes and their symbols (names) and orientation strand (O) where the up arrow represents negative strand and the down arrow represents positive strand.



**Figure 5:** 17q25.3 mutated region, the respective contig, genes and their symbols (names) and orientation strand (O) where the up arrow represents negative strand and the down arrow represents positive strand.

GeneAnalytics database is used as the primary source to receive the ontology data due to its user friendly options. Genes of the mutated regions of chromosomes 5, 10 and 17 are classified based on biological process ontology, which describes the biological function under name column and indicates the number of the matched genes of the sample out of the total number of genes sharing

the same function. If clicked, the number will show the specific gene names, as indicated in the samples below. The full link for each chromosome presented can be found in Appendix 2.

## Chromosome 5

Name	# Matched Genes (Total Genes)
Neural Crest Cell Differentiation	1 (6)
Cell Morphogenesis Involved in Neuron Differentiation	1 (6)
Positive Regulation of Alkaline Phosphatase Activity	1 (6)
Regulation of Dendritic Spine Development	1 (7)
Cartilage Morphogenesis	1 (7)
Positive Regulation of Cardiac Muscle Cell Differentiation	1 (8)
Regulation of Synapse Assembly	1 (8)
Positive Regulation of Protein Homodimerization Activity	1 (8)
Positive Regulation of Phosphorylation of RNA Polymerase II C-terminal Domain	1 (8)
Secondary Heart Field Specification	1 (8)
Geminal Center Formation	1 (9)
Cellular Response to Parathyroid Hormone Stimulus	1 (9)
Transdifferentiation	1 (10)
Positive Regulation of Skeletal Muscle Tissue Development	1 (10)
Regulation of Actin Filament Polymerization	1 (11)
Embryonic Viscerocranium Morphogenesis	1 (11)
Cellular Response to Fluid Shear Stress	1 (13)
Regulation of N-methyl-D-aspartate Selective Glutamate Receptor Activity	1 (14)
Smooth Muscle Cell Differentiation	1 (14)
MAPK Cascade	2 (258)

**Positive Regulation Of Phosphorylation Of RNA Polymerase II C-Terminal Domain**

Matched Genes: 1

**CCNH**  
Learn more about gene CCNH in GeneCards, the Human Gene Compendium

[Run as a New Query](#) | [Send to VarElect](#)

## Chromosome 10

▶ Name	# Matched Genes (Total Genes)
Cellular Response to Jasmonic Acid Stimulus	4 (4)
Doxorubicin Metabolic Process	4 (8)
Daunorubicin Metabolic Process	4 (8)
Progesterone Metabolic Process	3 (8)
Dimethylallyl Diphosphate Biosynthetic Process	2 (2)
Cellular Response to Prostaglandin D Stimulus	2 (6)
Steroid Metabolic Process	3 (39)
Retinal Metabolic Process	2 (12)
Isoprenoid Biosynthetic Process	2 (13)
Hyaluronan Metabolic Process	2 (13)
Retinoid Metabolic Process	3 (59)
Prostaglandin Metabolic Process	2 (14)
Digestion	3 (61)
Positive Regulation of T Cell Differentiation	2 (16)
MRNA Processing	4 (166)
Positive Regulation of Interleukin-4 Production	2 (20)
Positive Regulation of Protein Kinase B Signaling	3 (82)
Canonical Glycolysis	2 (26)
Regulation of Transcription Elongation From RNA Polymerase II Promoter	1 (1)
Negative Regulation of Collagen Binding	1 (1)

**Regulation Of Transcription Elongation From RNA Polymerase II Promoter**

Matched Genes: 1

**ZMYND11**  
[Learn more about gene ZMYND11 in GeneCards, the Human Gene Compendium](#)

## **Chromosome 17**

▶ Name	# Matched ▶ Genes (Total Genes)
Embryonic Skeletal System Morphogenesis	8 (50)
Anterior/posterior Pattern Specification	11 (101)
Embryonic Skeletal System Development	7 (42)
Rhombomere 4 Development	2 (2)
Regulation of Sodium:proton Antiporter Activity	2 (2)
Renal Phosphate Ion Absorption	2 (2)
Rhombomere Development	2 (2)
Renal Sodium Ion Transport	2 (2)
Glutathione Transport	2 (2)
Metanephric Tubule Development	2 (2)
Anatomical Structure Formation Involved in Morphogenesis	4 (21)
DNA Metabolic Process	4 (22)
Lipid Transport	7 (75)
Chondrocyte Differentiation Involved in Endochondral Bone Morphogenesis	2 (3)
Negative Regulation of Myeloid Leukocyte Differentiation	2 (3)
Regulation of Excretion	2 (3)
Negative Regulation of Sodium:proton Antiporter Activity	2 (3)
Positive Regulation of Signal Transduction By P53 Class Mediator	2 (3)
Negative Regulation of Catenin Import Into Nucleus	3 (11)
Monocarboxylic Acid Transport	3 (11)

## 4.4 Identification of cell proliferation-related genes

### Chromosome 5

**Candidate gene:** CCNH (Cyclin H) – protein coding

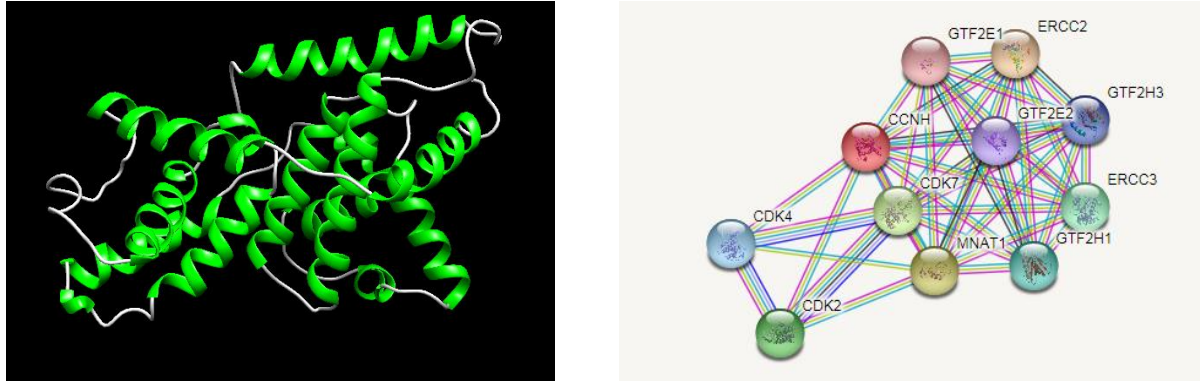
**Location:** q14.3

**GO Molecular Function:** contributes to RNA polymerase II carboxy-terminal domain kinase activity

The full map of the gene ontology analysis for this function can be found in Appendix 3.



CCNH gene, a protein-coding gene with PDB ID: 1KXU was fetched in the UCSF Chimera and the protein structure was retrieved as can be seen on the left image. On the right image a predicted and known protein interaction map retrieved by the STRING database is presented.



**Figure 6:** An UCSF Chimera representation of the CCNH protein structure showing the alpha helices (green) and an interaction map between CCNH protein and other proteins.

For more detailed data on specific interactions and the confidence score of each interaction, a link is provided in Appendix 2.

## **Chromosome 10**

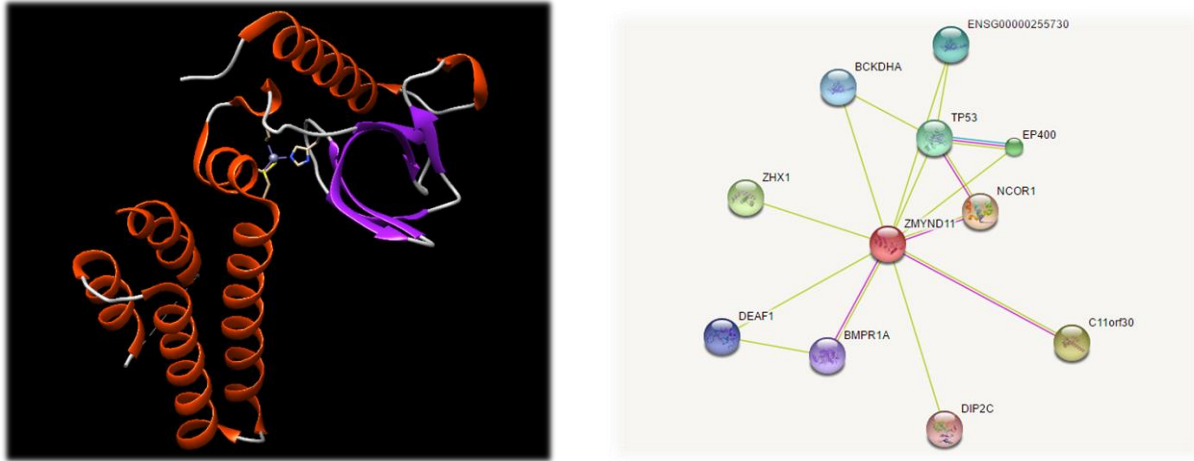
**Candidate gene:** ZMYND11 (Zinc Finger MYND-Type Containing 11) – protein coding

**Location:** p15.3

**GO Molecular Function:** RNA polymerase II regulatory region sequence-specific DNA binding

The full map of the gene ontology analysis for this function can be found in Appendix 4.

ZMYND11 gene, a protein-coding gene with PDB ID: 5HDA was fetched in the UCSF Chimera and the protein structure was retrieved as can be seen on the left image. On the right image a predicted and known protein interaction map retrieved by the STRING database is presented.



**Figure 7:** An UCSF Chimera representation of the ZMYND11 protein structure showing the alpha helices (red) and beta sheets (purple) and an interaction map between ZMYND11 protein and other proteins.

For more detailed data on specific interactions and the confidence score of each interaction, a link is provided in Appendix 2.

## **Chromosome 17**

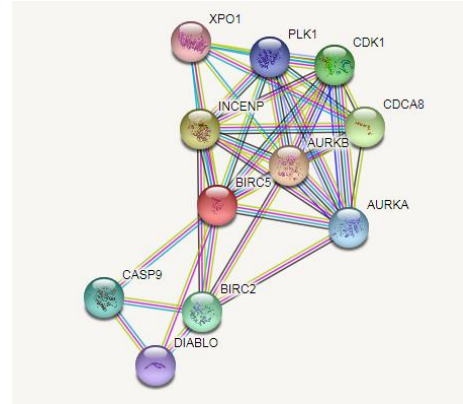
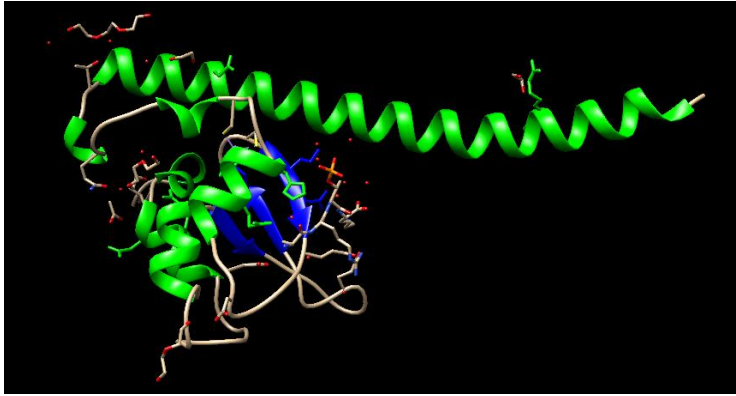
**Candidate gene:** BIRC5 (Baculoviral IAP Repeat Containing 5) – protein coding

**Location:** q25.3

**GO Molecular Function:** Ran GTPase binding – interacts with Ran, a GTP-binding protein involved in cell cycle progression, spindle assembly and nuclear organization assembly.

The full map of the gene ontology analysis for this function can be found in Appendix 5.

BIRC5 gene, a protein-coding gene with PDB ID: 3UEC was fetched in the UCSF Chimera and the protein structure was retrieved as can be seen on the left image. On the right image a predicted and known protein interaction map retrieved by the STRING database is presented.



**Figure 8:** An UCSF Chimera representation of the BIRC5 protein structure showing the alpha helices (green) and beta sheets (blue) and an interaction map between BIRC5 protein and other proteins.

For more detailed data on specific interactions and the confidence score of each interaction, a link is provided in Appendix 2.

In the following section, an extensive analysis of each candidate gene supported by peer review literature is presented, as well as recommendations for future work.

## 5.0 Discussion

In this section peer review literature discussed in Section 2.0 is used to interpret the findings of this study and further support the significance of the three identified genes as candidate genes that play a key role in stem cell *in vitro* culture.

The three genes that are proposed in this study to be candidate genes are: CCNH, ZMYND11 and BIRC5 gene in chromosomes 5, 10, and 17. On a larger picture, mutations in these chromosomes cause developmental and cognitive disabilities *in vivo*.

Located on chromosome 5 (region q14.3), CCNH gene is a protein coding gene part of the RNA polymerase II kinase activity of the cell cycle process as revealed by the cytogenetic studies. Its direct interaction with CDK7 to form a complex as a CDK-activating kinase was confirmed by STRING analysis as presented in Figure 6 and the whole complex together with other CDKs and MAT1 protein is part of transcriptional regulation. Since this gene has lost heterozygosity as revealed by CGH analysis and is previously linked with tumor growth in glioma cells (Weber et. al., 2000), it is a strong candidate for the abnormalities of H9 stem cell line.

Located on chromosome 10 (region p15.3), ZMYND11 gene is a protein coding gene part of the RNA polymerase II regulatory region sequence-specific DNA binding as revealed by cytogenetic studies. It acts as a transcriptional repressor by preventing tumor cell growth through its specific binding with histone H3 subtype (Wen et. al., 2014). STRING analysis revealed a direct interaction of ZMYND11 protein with DIP2C protein which has been previously mentioned in a 2012 study (DeScipio et. al., 2012). Both these genes are deleted due to random mutation acquired during cell culture, therefore confirming their role as strong candidate genes for the abnormalities of H9 stem cell line. 10p15.3 deletions are linked with speech delay and language

disorders, motor delays and brain anomalies (DeScipio et al., 2012; Szalontai and Csiszar, 2013; Cobben et al., 2014).

Located on chromosome 17 (region q25.3), BIRC5 gene is a protein coding gene part of the Ran GTPase binding as revealed by cytogenetic studies. This gene belongs to the IAP (inhibitor of apoptosis) family that negatively regulate cell death, meaning that higher expression of this gene can be found in tumors. Based on literature, GTPase binding proteins in 17q amplified regions have been linked to breast tumors (Montagna et. al., 2003). Its interaction with CASP9 as shown by the STRING analysis and other proteins forms a complex that promotes cell proliferation. Gains in 17q chromosome are related to early stage tumors of neuroblastoma (Abel et al., 1999; Krona et al., 2008). Since BIRC5 gene is amplified on chromosome 17, it is a strong candidate for the abnormalities of H9 stem cell line.

Multiple cancers are associated with a combination of these mutations and others, indicating that the effects of random mutation of stem cells *in vitro* can have larger and fatal implications *in vivo* such as heart defects, growth delays, cognitive disorders and other severe conditions (Baker et al., 2007).

## 6.0 Future studies

The genes proposed here as candidate genes in stem cell culture are protein coding genes, but the CGH analysis of the cell line revealed other non-coding sequences, mRNAs or microRNAs that were prone to mutations. Based on literature, microRNAs expression analysis can provide crucial information on the differentiation stages of the tumors and development processes (Lu et. al., 2005). Hence, as a future recommendation, an extensive focus on these sequences, non-coding regions, microRNAs, and their interaction with protein coding genes is required to create a wide profile of candidate sequences that can contribute to the degeneration of stem cells into cancer cells.

In this study, a bioinformatics and data analysis approach was taken to identify potential key genes. While this is the first and most important step in managing big and raw data, experimental approaches are required to verify the predictions and hypothesis stated in this paper. For the mentioned candidate genes, the next step would be to run transfection assays and record the changes through a second CGH analysis. Therefore, more data will be available to further support or oppose the role of the proposed genes. To investigate the functional relevance of these (and other) genes in transformation of the hESCs, expressing them in the cell through overexpression or knockdown will indicate the reaction of the cell toward differentiation. For example, since ZMYND11 gene is deleted from Ch10, overexpression can lead to phenotypical rescue. If the cells with the overexpressed ZMYND11 gene will be placed into neuronal differentiation environment and will be able to differentiate into neurons, it can be concluded that ZMYND11 indeed plays the key role in acquisition of transformed cell undifferentiated phenotype. Yet, there could be other factors that will need to be expressed in order to observe a

gain of function because as mentioned above about the role of microRNAs genes may colocalize with other expressed genes to carry their function.

Lastly, considering the large number of cell lines used in research and their *in vitro* culturing, mutations may arise even through the differences in cell culture conditions (Tosca et. al., 2015), hence specific characterization of growing conditions is important for maintenance and improvement of cell line culture *in vitro*. In order to minimize random mutations acquired by genes in *in vitro* stem cell cultures, researchers should focus on the combined use of experimental and computational approaches in order to classify and identify these genomic abnormalities to assure safe use of hESCs.

## 7.0 References

1. Abel, F., Ejeskär, K., Kogner, P., & Martinsson, T. (1999). Gain of chromosome arm 17q is associated with unfavourable prognosis in neuroblastoma, but does not involve mutations in the somatostatin receptor 2 (SSTR2) gene at 17q24. *British journal of cancer*, 81(8), 1402. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2362984/>
2. Abeyta, M. J., Clark, A. T., Rodriguez, R. T., Bodnar, M. S., Pera, R. A. R., & Firpo, M. T. (2004). Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Human molecular genetics*, 13(6), 601-608. Retrieved from: <https://www.ncbi.nlm.nih.gov/pubmed/14749348>
3. Agilent Technologies. (2017). Human Genome CGH+SNP Microarrays. Retrieved from: <http://www.genomics.agilent.com/en/CGH-SNP-Microarrays/Human-Genome-CGH-SNP-Microarrays/?cid=AG-PT-112&tabId=AG-PR-1080>
4. Avery, S., Hirst, A. J., Baker, D., Lim, C. Y., Alagaratnam, S., Skotheim, R. I., ... & Andrews, P. W. (2013). BCL-XL mediates the strong selective advantage of a 20q11. 21 amplification commonly found in human embryonic stem cell cultures. *Stem cell reports*, 1(5), 379-386. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S2213671113000994>
5. Baker, D. E., Harrison, N. J., Maltby, E., Smith, K., Moore, H. D., Shaw, P. J., ... & Andrews, P. W. (2007). Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nature biotechnology*, 25(2), 207-215. Retrieved from: <http://www.nature.com/nbt/journal/v25/n2/full/nbt1285.html>
6. Bryndorf, T., Kirchhoff, M., Rose, H., Maahr, J., Gerdes, T., Karhu, R., ... & Philip, J. (1995). Comparative genomic hybridization in clinical cytogenetics. *American journal of human genetics*, 57(5), 1211. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1801381/pdf/ajhg00037-0227.pdf>
7. Cobben, J. M., Weiss, M. M., Van Dijk, F. S., De Reuver, R., de Kruiff, C., Pondaag, W., ... & Yntema, H. G. (2014). A de novo mutation in ZMYND11, a candidate gene for 10p15. 3 deletion syndrome, is associated with syndromic intellectual disability. *European journal of medical genetics*, 57(11), 636-638. Retrieved from: <http://www.sciencedirect.com.ezproxy.wpi.edu/science/article/pii/S1769721214001724>
8. Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., ... & Abdel-Hamid, H. (2011). A copy number variation morbidity map of developmental delay. *Nature genetics*, 43(9), 838-846. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3171215/>
9. Cytogenetic Services WiCell. (2012). Retrieved from: <http://www.wicell.org/home/cytogenetic-services/cytogenetic-services.cmsx>
10. DeScipio, C., Conlin, L., Rosenfeld, J., Tepperberg, J., Pasion, R., Patel, A., ... & McGuire, M. (2012). Subtelomeric deletion of chromosome 10p15. 3: clinical findings and molecular cytogenetic characterization. *American Journal of Medical Genetics Part*



- A, 158(9), 2152-2161. Retrieved from:  
<http://onlinelibrary.wiley.com.ezproxy.wpi.edu/doi/10.1002/ajmg.a.35574/full>
11. Desmarais, J. A., Hoffmann, M. J., Bingham, G., Gagou, M. E., Meuth, M., & Andrews, P. W. (2012). Human embryonic stem cells fail to activate CHK1 and commit to apoptosis in response to DNA replication stress. *Stem Cells*, 30(7), 1385-1393. Retrieved from: <https://www.ncbi.nlm.nih.gov/pubmed/22553144?dopt=Citation>
  12. Gene Ontology Consortium. (2017). Ontology Structure. Retrieved from: <http://www.geneontology.org/page/ontology-structure>
  13. Guo, R., Zheng, L., Park, J. W., Lv, R., Chen, H., Jiao, F., ... & Wang, Z. (2014). BS69/ZMYND11 reads and connects histone H3. 3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Molecular cell*, 56(2), 298-310. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S1097276514006777>
  14. Henegariu, O., Heerema, N. A., Lowe Wright, L., Bray-Ward, P., Ward, D. C., & Vance, G. H. (2001). Improvements in cytogenetic slide preparation: controlled chromosome spreading, chemical aging and gradual denaturing. *Cytometry*, 43(2), 101-109. Retrieved from: [http://onlinelibrary.wiley.com/doi/10.1002/1097-0320\(20010201\)43:2%3C101::AID-CYTO1024%3E3.0.CO;2-8/full](http://onlinelibrary.wiley.com/doi/10.1002/1097-0320(20010201)43:2%3C101::AID-CYTO1024%3E3.0.CO;2-8/full)
  15. The International Stem Cell Initiative, 2011. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nat Biotechnol* (12):1132-44). Retrieved from: <http://www.nature.com/nbt/journal/v29/n12/abs/nbt.2051.html>
  16. International Stem Cell Registry. (2017). hES Cell Line: H9 (WA09). Umass Medical School. Retrieved from: <http://www.iscr-admin.com/Default.aspx?Action=viewsc&StemCellLine=89>
  17. Kole, D., Grella, A., Dolivo, D., Shumaker, L., Hermans, W., & Dominko, T. (2017). High molecular weight FGF2 isoforms demonstrate canonical receptor-mediated activity and support human embryonic stem cell self-renewal. *Stem Cell Research*. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S1873506117300727>
  18. Krona, C., Carén, H., Sjöberg, R. M., Sandstedt, B., Laureys, G., Kogner, P., & Martinsson, T. (2008). Analysis of neuroblastoma tumour progression; loss of PHOX2B on 4p13 and 17q gain are early events in neuroblastoma tumorigenesis. *International journal of oncology*, 32(3), 575-584. Retrieved as paper copy.
  19. Lefort, N., Feyeux, M., Bas, C., Féraud, O., Bennaceur-Griscelli, A., Tachdjian, G., ... & Perrier, A. L. (2008). Human embryonic stem cells reveal recurrent genomic instability at 20q11. 21. *Nature biotechnology*, 26(12), 1364. Retrieved from: <http://search.proquest.com/openview/a437e8b74161b2a9b243ab192106d3ea/1?pq-origsite=gscholar&cbl=47191>
  20. LifeMap Discovery. (n.d). The Embryonic Development, Regenerative Medicine and Stem Cell Database. Retrieved from: <http://discovery.lifemapsc.com/>
  21. Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., ... & Downing, J. R. (2005). MicroRNA expression profiles classify human cancers. *nature*, 435(7043), 834-838. Retrieved from: <http://www.nature.com/nature/journal/v435/n7043/full/nature03702.html>

22. McIlhatton, M. A., Burrows, J. F., Donaghy, P. G., Chanduloy, S., Johnston, P. G., & Russell, S. H. (2001). Genomic organization, complex splicing pattern and expression of a human septin gene on chromosome 17q25. 3. *Oncogene*, 20(41), 5930. Retrieved from: <https://www.nature.com/onc/journal/v20/n41/full/1204752a.html>
23. Mendes-da-Silva, P., Moriera, A., Duro-da-Costa, J., Matias, D., & Monteiro, C. (2000). Frequent loss of heterozygosity on chromosome 5 in non-small cell lung carcinoma. *Journal of Clinical Pathology*, 53(4), 184. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1186967/>
24. Montagna, C., Lyu, M. S., Hunter, K., Lukes, L., Lowther, W., Reppert, T., ... & Ried, T. (2003). The Septin 9 (MSF) gene is amplified and overexpressed in mouse mammary gland adenocarcinomas and human breast cancer cell lines. *Cancer research*, 63(9), 2179-2187. Retrieved from: <http://cancerres.aacrjournals.org/content/63/9/2179.long>
25. NCBI. (2017). What is NCBI Remap? Retrieved from: <https://www.ncbi.nlm.nih.gov/genome/tools/remap/docs/whatis>
26. Nguyen, H. T., Geens, M., Mertzaniidou, A., Jacobs, K., Heirman, C., Breckpot, K., & Spits, C. (2013). Gain of 20q11. 21 in human embryonic stem cells improves cell survival by increased expression of Bcl-xL. *Molecular human reproduction*, gat077. Retrieved from: <https://academic.oup.com/molehr/article/20/2/168/998526/Gain-of-20q11-21-in-human-embryonic-stem-cells>
27. NIH Human Embryonic stem Cell Registry. (2011). National Institute of Health (NIH). Retrieved from: [https://grants.nih.gov/stem\\_cells/registry/current.htm](https://grants.nih.gov/stem_cells/registry/current.htm)
28. Noshio, K., Igarashi, H., Nojima, M., Ito, M., Maruyama, R., Yoshii, S., ... & Yamamoto, E. (2013). Association of microRNA-31 with BRAF mutation, colorectal cancer survival and serrated pathway. *Carcinogenesis*, bgt374. Retrieved from: <https://academic.oup.com/carcin/article/35/4/776/268932/Association-of-microRNA-31-with-BRAF-mutation>
29. Ozaki, T., Schaefer, K. L., Wai, D., Buerger, H., Flege, S., Lindner, N., ... & Juergens, H. (2002). Genetic imbalances revealed by comparative genomic hybridization in osteosarcomas. *International journal of cancer*, 102(4), 355-365. Retrieved from: <http://onlinelibrary.wiley.com/doi/10.1002/ijc.10709/full>
30. Puppini, C., Durante, C., Sponziello, M., Verrienti, A., Pecce, V., Lavarone, E., ... & Russo, D. (2014). Overexpression of genes involved in miRNA biogenesis in medullary thyroid carcinomas with RET mutation. *Endocrine*, 47(2), 528-536. Retrieved from: <https://link.springer.com/article/10.1007/s12020-014-0204-3>
31. RBVI. (2016). UCSF Chimera, an Extensible Molecular Modeling System. Retrieved from: <https://www.cgl.ucsf.edu/chimera/>
32. Riaz, M., van Jaarsveld, M. T., Hollestelle, A., Prager-van der Smissen, W. J., Heine, A. A., Boersma, A. W., ... & Wiemer, E. A. (2013). miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs. *Breast Cancer Research*, 15(2), R33. Retrieved from: <https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr3415>
33. Saito-Ohara, F., Imoto, I., Inoue, J., Hosoi, H., Nakagawara, A., Sugimoto, T., & Inazawa, J. (2003). PPM1D is a potential target for 17q gain in neuroblastoma. *Cancer*

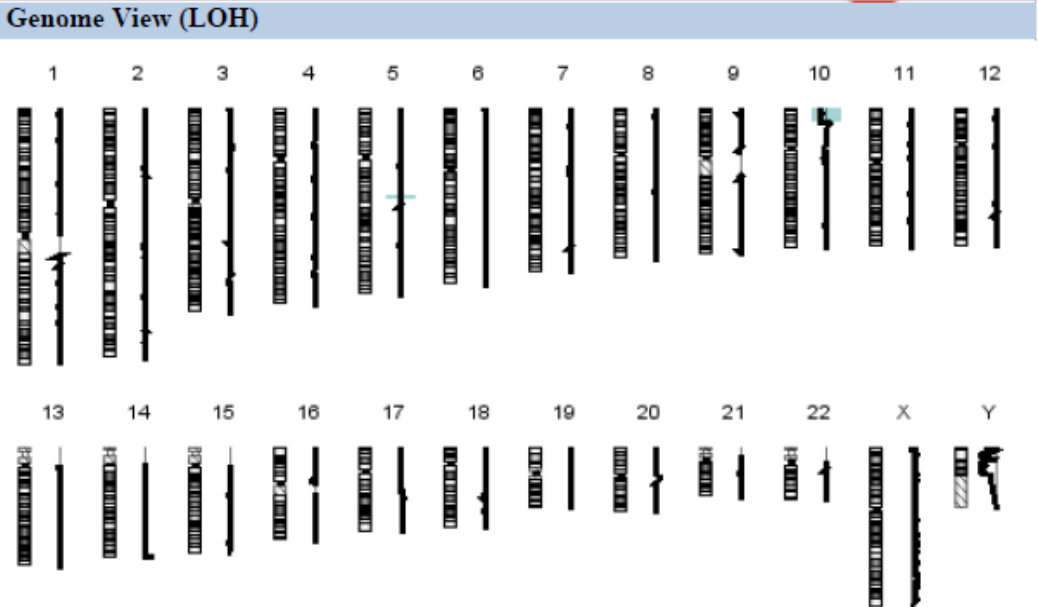
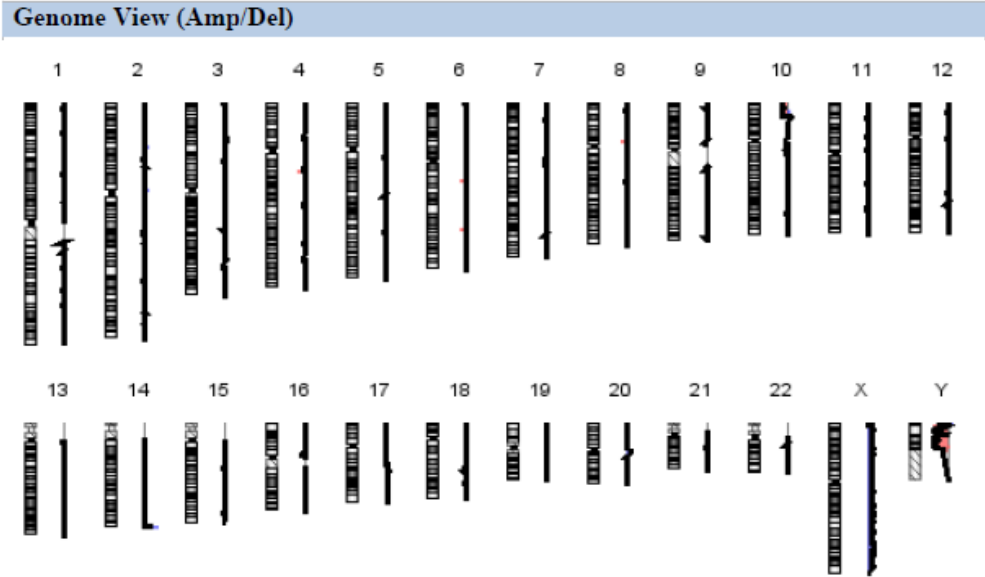
- research*, 63(8), 1876-1883. Retrieved from:  
<http://cancerres.aacrjournals.org/content/canres/63/8/1876.full.pdf>
34. Shadeo, A., & Lam, W. L. (2006). Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Research*, 8(1), R9. Retrieved from: <https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr1370>
  35. Solomon, E., Voss, R., Hall, V., Bodmer, W. F., Jass, J. R., Jeffreys, A. J., ... & Rider, S. H. (1987). Chromosome 5 allele loss in human colorectal carcinomas. Retrieved from: <http://www.nature.com/nature/journal/v328/n6131/abs/328616a0.html>
  36. Spits, C., Mateizel, I., Geens, M., Mertzanidou, A., Staessen, C., Vandesselde, Y., ... & Sermon, K. (2008). Recurrent chromosomal abnormalities in human embryonic stem cells. *Nature biotechnology*, 26(12), 1361. Retrieved from: <http://search.proquest.com/openview/d6fe67f1a7aed7dabe63c3749e096aa2/1?pq-origsite=gscholar&cbl=47191>
  37. Storlazzi, C. T., Brekke, H. R., Mandahl, N., Brosjö, O., Smeland, S., Lothe, R. A., & Mertens, F. (2006). Identification of a novel amplicon at distal 17q containing the BIRC5/SURVIVIN gene in malignant peripheral nerve sheath tumours. *The Journal of pathology*, 209(4), 492-500. Retrieved from: <http://onlinelibrary.wiley.com/doi/10.1002/path.1998/full>
  38. STRING. (2015). Database issue: D447-52. Retrieved from: <http://string-db.org/cgi/help.pl?UserId=cgE7XuJLT5ux&sessionId=Rvs1fMAC34Py>
  39. Sun, W. C., Liang, Z. D., & Pei, L. (2015). Propofol-induced rno-miR-665 targets BCL2L1 and influences apoptosis in rodent developing hippocampal astrocytes. *Neurotoxicology*, 51, 87-95. Retrieved from: <http://www.sciencedirect.com/science/article/pii/S0161813X15001163>
  40. Szalontai, A., & Csiszar, K. (2013). Genetic insights into the functional elements of language. *Human Genetics*, 132(9), 959-986. Retrieved from: <https://link.springer.com.ezproxy.wpi.edu/article/10.1007%2Fs00439-013-1317-0>
  41. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *science*, 282(5391), 1145-1147. Retrieved from: <http://science.sciencemag.org/content/282/5391/1145.full>
  42. Tosca, L., Feraud, O., Magniez, A., Bas, C., Griscelli, F., Bennaceur-Griscelli, A., & Tachdjian, G. (2015). Genomic instability of human embryonic stem cell lines using different passaging culture methods. *Molecular cytogenetics*, 8(1), 30. Retrieved from: <https://molecularcytogenetics.biomedcentral.com/articles/10.1186/s13039-015-0133-8>
  43. Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., ... & Kallioniemi, A. (2002). Targets of gene amplification and overexpression at 17q in gastric cancer. *Cancer research*, 62(9), 2625-2629. Retrieved from: <http://cancerres.aacrjournals.org/content/canres/62/9/2625.full.pdf>
  44. Ware, C. B., Nelson, A. M., & Blau, C. A. (2006). A comparison of NIH-approved human ESC lines. *Stem cells*, 24(12), 2677-2684. Retrieved from: <http://onlinelibrary.wiley.com/doi/10.1634/stemcells.2005-0452/pdf>

45. Weber, R. G., Rieger, J., Naumann, U., Lichter, P., & Weller, M. (2001). Chromosomal imbalances associated with response to chemotherapy and cytotoxic cytokines in human malignant glioma cell lines. *International journal of cancer*, 91(2), 213-218. Retrieved from: [http://onlinelibrary.wiley.com/doi/10.1002/1097-0215\(200002\)9999:9999%3C::AID-IJC1036%3E3.0.CO;2-M/full](http://onlinelibrary.wiley.com/doi/10.1002/1097-0215(200002)9999:9999%3C::AID-IJC1036%3E3.0.CO;2-M/full)
46. Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., ... & Li, B. (2014). ZMYND11 links histone H3. 3K36me3 to transcription elongation and tumour suppression. *Nature*, 508(7495), 263-268. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4142212/>
47. WiCell. (2012). Retrieved from: <http://www.wicell.org/home/about-wicell/news-room/42710/>
48. WiCell Feeder Based (MEF) Pluripotent Stem Cell Protocols. (2003). WiCell®. Retrieved from: <http://www.wicell.org/product-files/cultureProtocols/SOP-SH-001.pdf>

# 8.0 Appendices

## Appendix 1

Genomic view of the CGH+SNP microarray results. The upper picture indicates the deletion and amplification mutations, while the bottom picture indicates the loss of heterozygosity mutation.



## **Appendix 2**

### GeneAnalytics Database

Full analysis for chromosome 5 mutated region can be found at:

<https://ga.genecards.org/#results>

Full analysis for chromosome 10 mutated region can be found at:

<https://ga.genecards.org/#results>

Full analysis for chromosome 17 mutated region can be found at:

<https://ga.genecards.org/#results>

*Note: To access these links it is required to have an account.*

### STRING Database

STRING analysis and detailed information on protein interaction for the candidate gene on chromosome 5 can be found at:

<http://string-db.org/cgi/network.pl?taskId=i2JQBT92dij2>

STRING analysis and detailed information on protein interaction for the candidate gene on chromosome 10 can be found at:

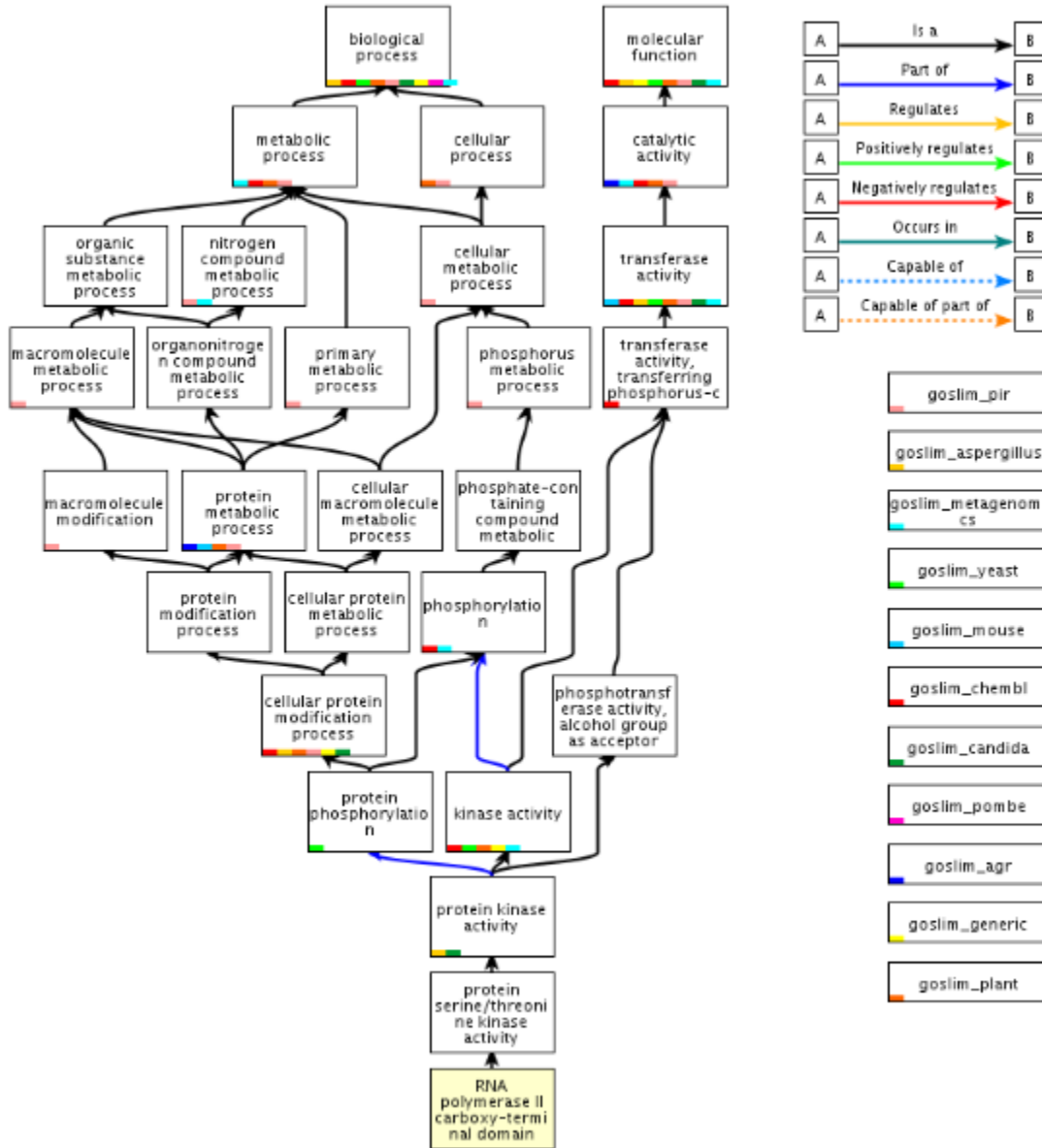
<http://string-db.org/cgi/network.pl?taskId=I7XLE0eNfrVw>

STRING analysis and detailed information on protein interaction for the candidate gene on chromosome 17 can be found at:

<http://string-db.org/cgi/network.pl?taskId=kPfN1KTbnAYr>

### Appendix 3

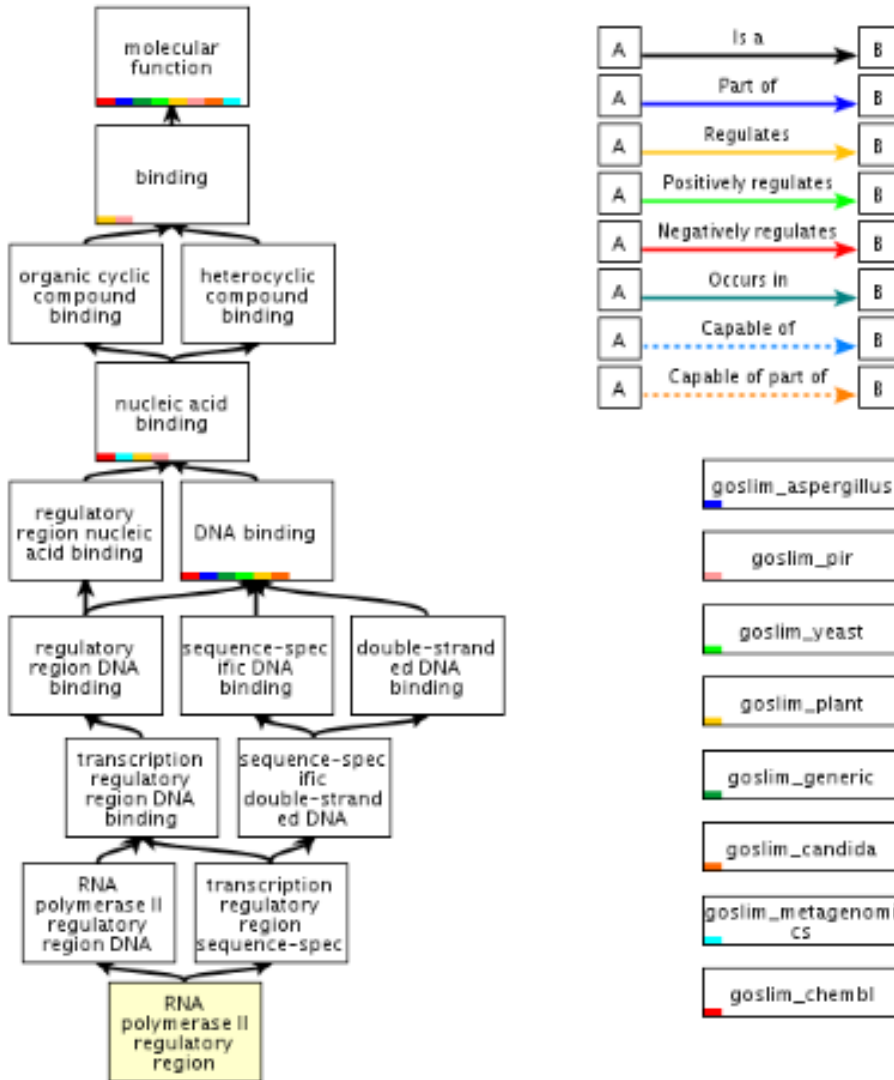
Graph view of the gene ontology molecular function of the candidate gene (CCNH) of chromosome 5 (GO: 0008353).



Retrieved from: <http://amigo.geneontology.org/amigo/term/GO:0008353>

## Appendix 4

Graph view of the gene ontology molecular function of the candidate gene (ZMYND11) of chromosome 10 (GO: 0000977).

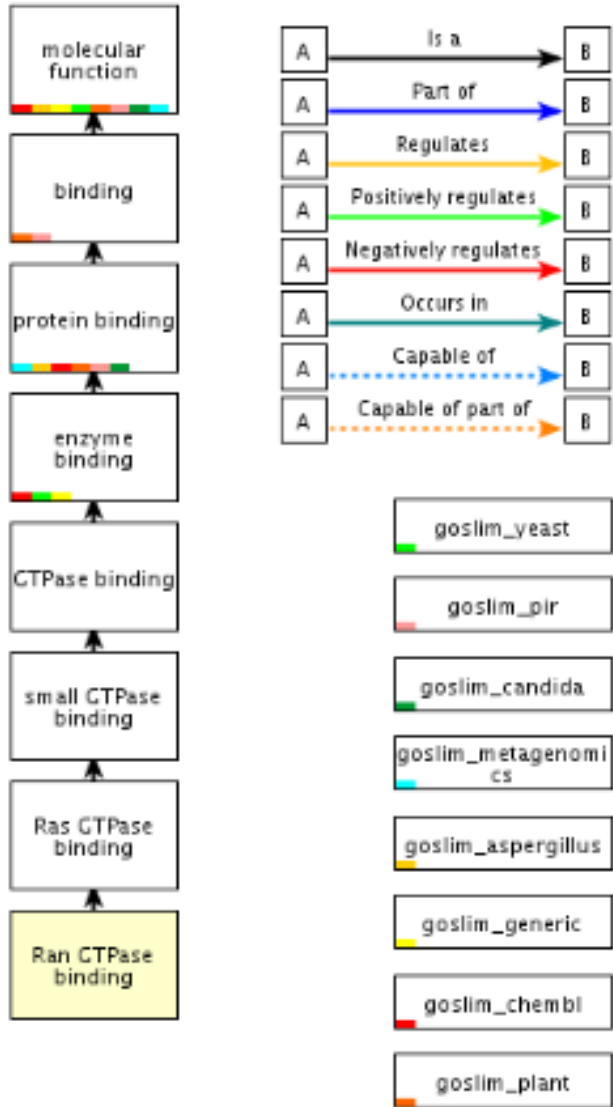


Retrieved from: <http://amigo.geneontology.org/amigo/term/GO:0000977>



## Appendix 5

Graph view of the gene ontology molecular function of the candidate gene (BIRC5) of chromosome 17 (GO: 0008536).



Retrieved from: <http://amigo.geneontology.org/amigo/term/GO:0008536>

