Role of CTCF Mutations in Altering Nuclear Architecture and Gene Regulation in Cancer

Daniel Bower

Mentor: Jane Skok, PhD

May 2017

A thesis in fulfillment of the Masters in Biology Degree

ACKNOWLEDGEMENTS

I am incredibly grateful to Jane Skok, a wonderful advisor. Without her ambitious vision for my project I would not be half as pleased with the final product. Her outstanding guidance has made me into a better scientist. A special thanks is due to Nishana Mayilaadumveettil, my mentor in the Skok lab, for her patience and mentorship. This thesis could not have been done without her.

I am also grateful to all the members of the Skok lab: Sana Badri, Pedro Rocha, Jason Wong, Lili Blumenberg, Erica Chio, Priscillia Lhoumaud, Gunjan Sethia, Deirdre Greenridge, Nikolaos Vosniakis, and Yi Li. Thank you so much for making this possible.

I thank Marcus B. Noyes for his sharing his bacterial one hybrid assay system, as well as for guidance and scientific advice, and the members of the Noyes lab for their gracious and welcoming attitude. I am also grateful to Elphège Nora, for his kindness in sharing the vectors and cells for the auxin-inducible degron system.

I am also thankful for my loving girlfriend, Laura, for putting up with me through these long and stressful graduate school years. I would also like to thank my parents, Carol & Ernie, for accepting me and supporting me in my journey as a scientist.

1.0 SUMMARY

The genome is regulated at several different levels, at the nuclear scale, the supranucleosomal, and the nucleosomal scale. CTCF, an 11 Zinc finger protein, is deeply involved at almost all these levels of organization, because of its fundamental function in chromatin loop formation. Changes to the function of CTCF have broad effects due to its importance. Because of the importance of CTCF, it is implicated in cancer as a haploinsufficient tumor suppressor. Some mutations in the zinc fingers of CTCF can mimic this effect, due to causing a complete loss of binding of CTCF. However, mutations that do not cause total depletion of CTCF binding are also implicated in cancer, but it is less clear how this phenotype is related to the change of function caused by these mutations. Performing bacterial one-hybrid binding assays on CTCF mutations can uncover how the binding pattern has been affected and can uncover reasons for the observed ontogenetic properties of these mutations. Additionally, the effects of CTCF depletion on genome architecture can help uncover what role CTCF plays in regulating the three-dimensional structure of the genome.

2.0 INTRODUCTION

The three-dimensional (3D) structure of the genome is maintained by complex interactions of proteins and chromatin. One major focus of research is uncovering how these DNA-protein interactions influence nuclear organization and gene regulation. Eukaryotic genomes are organized into higher order structures, which are known to influence gene expression (Bortle and Corces 2012). A critical component of the 3D genome architecture is an insulator binding protein known as CCCTC-binding factor (CTCF). CTCF is a highly conserved, ubiquitously expressed nucleic acid binding zinc finger protein. It is involved at several layers of chromatin organization, including in the formation of chromatin loops and regulation of enhancer-promoter relationships. Because of its importance in genome structure, dysfunction of CTCF has also been implicated in cancer.

2.1 NUCLEAR ORGANIZATION:

Microscopy studies have provided evidence that organization of chromatin within the nucleus is nonrandom and chromosomes occupy distinct territories throughout interphase known as "chromosome territories" **[Figure 1]**. However these studies did not provide enough evidence to conclusively prove the existence of chromosome territories (Stack et al. 1977). In fact, for some time the existence of chromosome territories was thought to be experimentally disproved (Wischnitzer 1973). Direct evidence for their existence was provided later by the advent of a new type of visualization in microscopy, fluorescent in-situ hybridization (FISH), in the 1980s (Cremer, T., & Cremer, M. 2010).

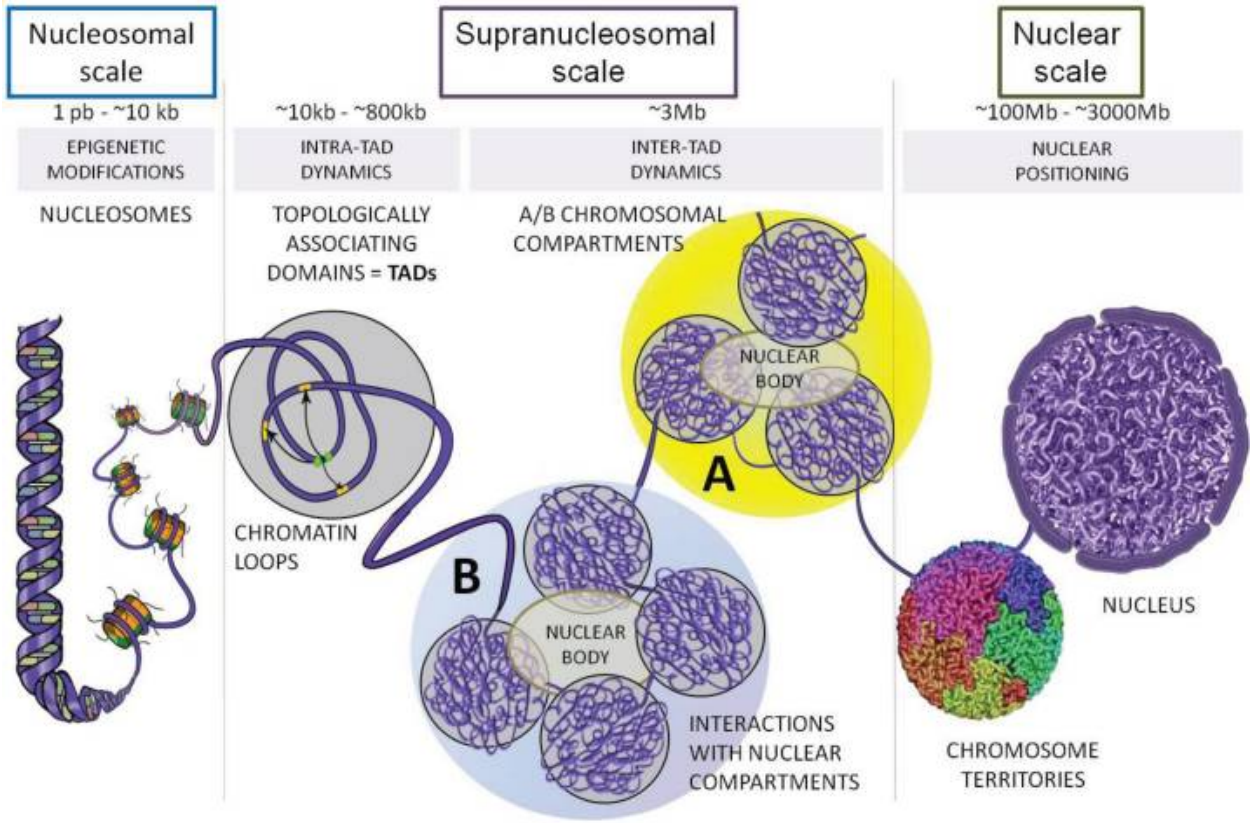


Figure 1. Chromatin organization. Chromatin is organized into two compartments, the "A" compartment in the center and the "B" compartment on the outside of the nucleus. Chromosome territories are the locations of the different chromosomes, shown in different colors. On a smaller scale, DNA forms loops mediated by CTCF and cohesin (Ea et al., 2015).

While chromosome territories are now accepted as a key feature of nuclear organization, FISH and related microscopy methods do not show the molecular mechanism behind their formation. The advent of molecular and sequencing-based approaches was instrumental in further understanding of spatial interactions in the chromosome. These techniques are based on "chromosome conformation capture" (3C), initially developed by Dekker et al. (2002). Extensions of this technique include circular chromosome conformation capture (4C), (Zhao et al., 2006, Simonis et al., 2006), chromosome conformation capture carbon copy (5C) (Dostie et al., 2016), and Hi-C (Lieberman-Aiden et al., 2009). Unlike microscopy techniques, 3C and its derivatives infer proximity based on the frequency of two sequences being in close contact and ligating to-

gether. Sequences can be far apart on the linear chromosome but still have a high frequency of ligation, implying that they are close together in space. Hi-C, an "all-versus-all" technique, visualizes all potential interactions in the genome (de Wit & de Laat, 2012). 3C, Hi-C, and related techniques are used to determine the location of genomic interactions in space, which may be separated by many kilobases or megabases on the linear chromosome or even be located on a different chromosome.

The nucleus is segregated into two “compartments”, A and B corresponding to open and closed chromatin respectively. Nuclear compartments were discovered through Hi-C studies. Loci in A compartments are generally transcriptionally active, located in the interior of nucleus and DNase I hypersensitive. In contrast those in the B compartment are transcriptionally repressed, located at the nuclear periphery and resistant to DNase I. As a general rule, interactions are largely constrained to occur between loci belonging to the same compartment **[Figure 1]** (Lieberman-Aiden et al. 2009, Hansen & Forten 2015). However topologically associating domains (TAD) structure can interfere with compartmentalization as shown by the impact of cohesin removal (Rao et al., 2017).

Nuclear compartments are further divided into megabase or submegabase domains called topologically associating domains (TADs). TADs are highly self-interacting regions of chromatin with the same epigenetic marks. Almost all cell specific long-range interactions occur within TADs (Dixon et al. 2012). TADs are conserved evolutionarily, and across cell types. The borders of TADs are enriched with CTCF. In mice, CTCF sites are found at 75% of TAD boundaries (Dixon et al. 2012). CTCF plays an important role in the formation of TADs. Deletion of a CTCF binding site at a TAD boundary leads to invasion of DNA methylation state from inside the TAD

to nearby sequences (Gombert and Krumm 2009). Additionally, deletion of CTCF binding sites within Hox clusters causes active chromatin to invade into previously repressed chromatin, causing transcriptionally aberrant cells when they were differentiated. This shows that CTCF is necessary for TAD formation, and that TAD formation is necessary for transcriptional programs to proceed properly (Narendra et al 2015). In other words, at TAD borders, CTCF is associated with a strong insulating role and prevents interactions between enhancers and inappropriate gene targets (Ea et al 2015).

Sub-TAD level organization is also important for maintaining genome architecture. These sub-TAD structures called “contact domains” have been studied in high resolution in *in situ* Hi-C experiments and have a median size of 185 kilobases. There are several types of contact domains, and about 65% of them correspond to chromatin loops, which are associated with CTCF. Thus, CTCF is not only associated with TAD borders, but also with the ends of chromatin loops (Ea et al. 2015). CTCF and cohesin are the key players in the formation of loops. The most accepted model for CTCF and cohesin in loop formation is called is the "loop extrusion model.” Accordingly, cohesin slides along DNA actively extruding it into loops until it encounters two convergently orientated CTCF sites **[Figure 2]**. (Lieberman-Aiden et al., 2009, Fudenberg et al., 2016, Sanborn et al., 2015).

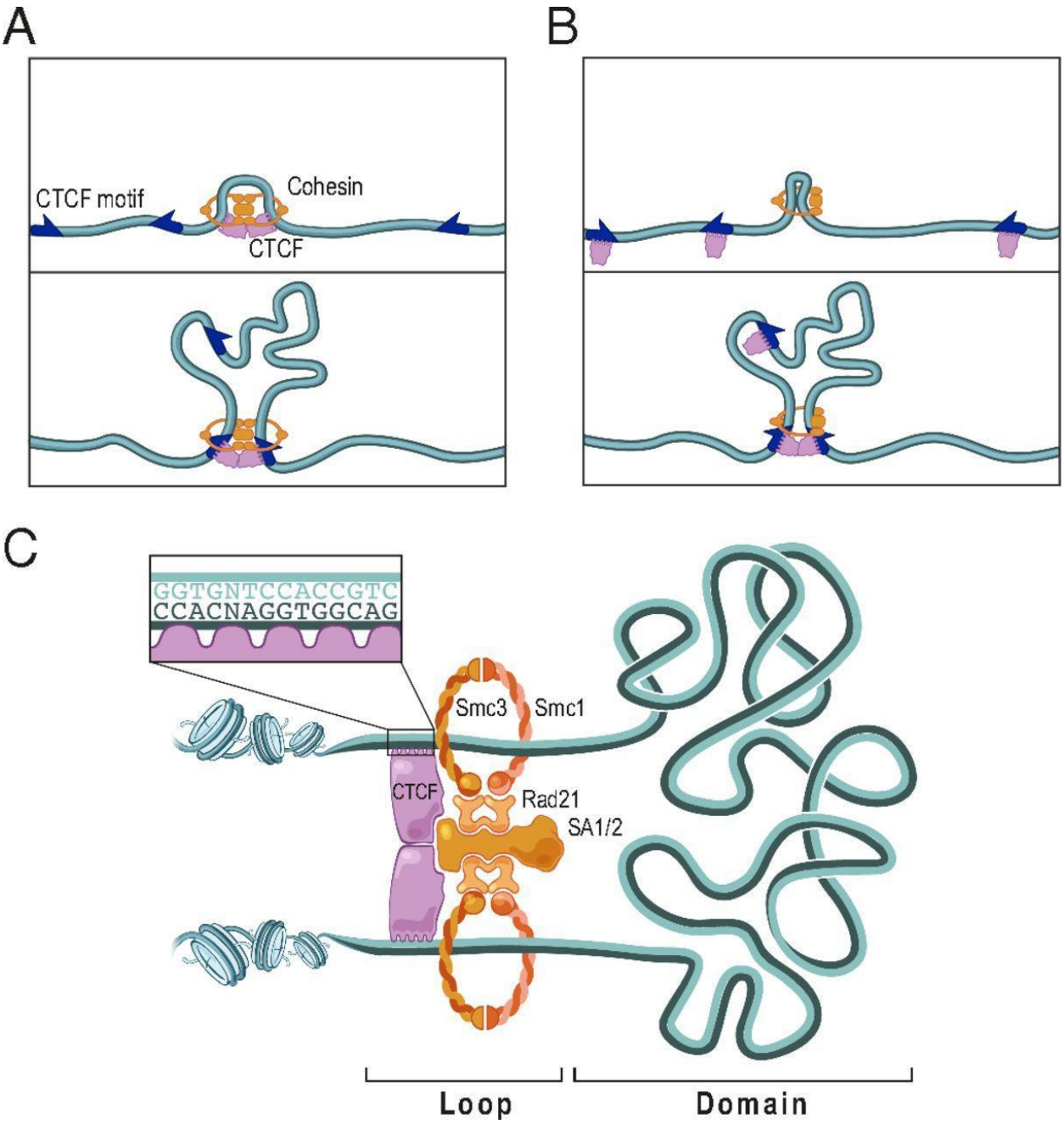


Figure 2. Loop extrusion model. The formation of loops is mediated by an extrusion complex formed by cohesin: Smc3 and Smc1, Rad21 and SA1/2. The extrusion complex moves along the DNA until it reaches a CTCF site occupied by a CTCF protein. Two molecules of CTCF seal the loop. While this figure shows two cohesin molecules at the loop anchor, it may be that only one is necessary for the formation of loops (figure adapted Sanborn et al., 2015).

2.2 STRUCTURE OF CTCF:

CTCF, a 727 amino acid long protein, is highly conserved from *Drosophila* to human. One third of the protein is made up of 11 zinc fingers (ZFs) that are involved in binding toof ~60,000 CTCF binding sites in the genome. The crystal structure of the CTCF ZF peptides

shows that ZFs 3-7 recognize a 15 base pair consensus sequence in the major groove of DNA. ZF8 is not involved with base recognition, rather it recognizes the minor groove and assists in binding stability (Hashimoto et al., 2017) **[Figure 3]**. However, another crystal structure of CTCF shows that ZFs 9-11 do make contact with DNA, however they only bind with one base as opposed to three as in the case of the other ZFs. In this second crystal structure, ZF8 assists by forming a flexible spacer element that does not recognize any bases, but instead forms a bridge across the phosphate backbone, allowing for ZFs 4-7 and ZFs 9-11 to bind to their respective sections of the CTCF binding site. This is consistent with the finding that CTCF binding sites have non-conserved regions between the site where ZFs 4-7 and ZFs 9-11 recognize DNA (Yin et al., 2017).

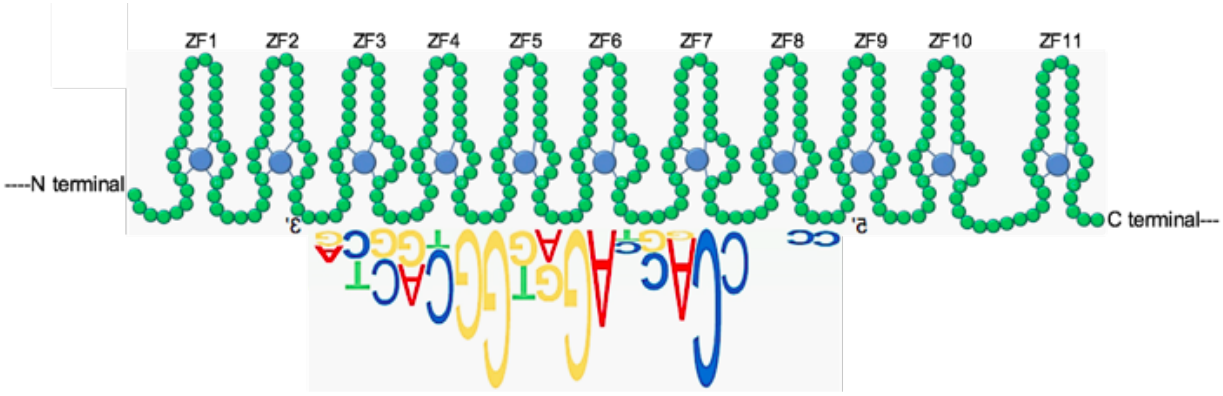


Figure 3. Structure of CTCF, with binding site sequence (below).

2.3 ZINC FINGERS OF CTCF:

The structure of zinc finger is important in considering how different mutations in the ZF region of CTCF affect its binding and function. CTCF contains 10 C2H2 ZFs and one CH2C ZF that help it to bind DNA (Hashimoto et al., 2017, Yin et al., 2017). Each zinc finger has histidine/cysteine that forms co-ordinate bonds with zinc ions, amino acids that bind to the bases of the recognition sequence and those that form the linker region connecting adjacent ZF **[Figure**

4]. A zinc finger domain is composed of two beta-pleated sheets and an alpha helix. The binding site recognition amino acids are in position -1 to 6 of the zinc finger, with position 1 indicating the start of the alpha helix. Each zinc finger recognizes three bases. Amino acids in position -1, 2, 3, and 6 are base-recognition sites, and thus changes in these amino acids can affect specificity of binding of CTCF, leading to changes in affinity of binding to consensus binding site or even altered specificity. Each base-recognition amino acid of the ZF makes a hydrogen bond on its respective nucleotide residue. Additionally, there are two cysteines and two histidines that form a zinc coordinating complex, and any mutation that disrupts this interaction is thought to render the entire zinc finger, and in some cases the entire protein, unable to bind (Hashimoto et al 2017).

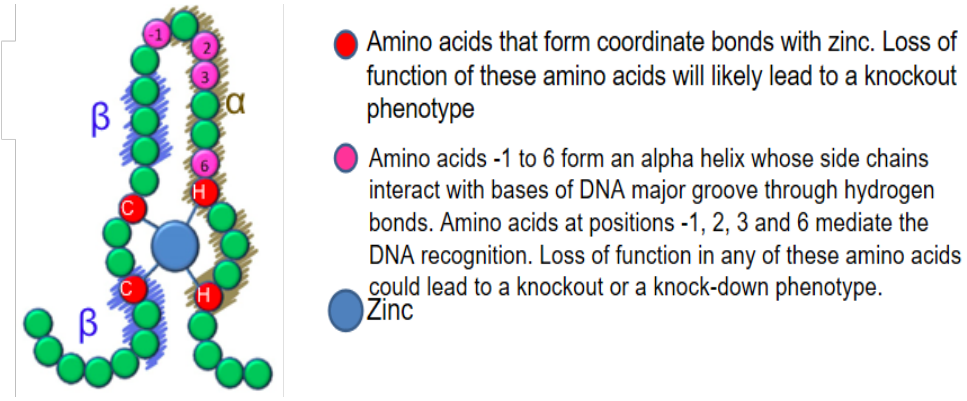


Figure 4. Structure of a C2H2 zinc finger.

2.4 ROLE OF CTCF PROTEIN MUTATIONS IN CANCER:

Mutations that affect the function of CTCF can have wide-ranging effects on gene function due to the protein's importance in TAD architecture and gene regulation (Phillips and Corces, 2009). Loss of heterozygosity at the 16q22 locus, where CTCF is encoded, is commonly observed in human cancers **[Figure 5]** (Filippova et al., 2002, Loukinov et al., 2016). In addition,

mutations can lead to loss of recognition of CTCF binding sites, and thus the loss of an insulated neighborhood, leading to an alteration in interactions and proto-oncogenes coming in contact with regulatory elements in a pathological manner (Hnisz et al 2016).

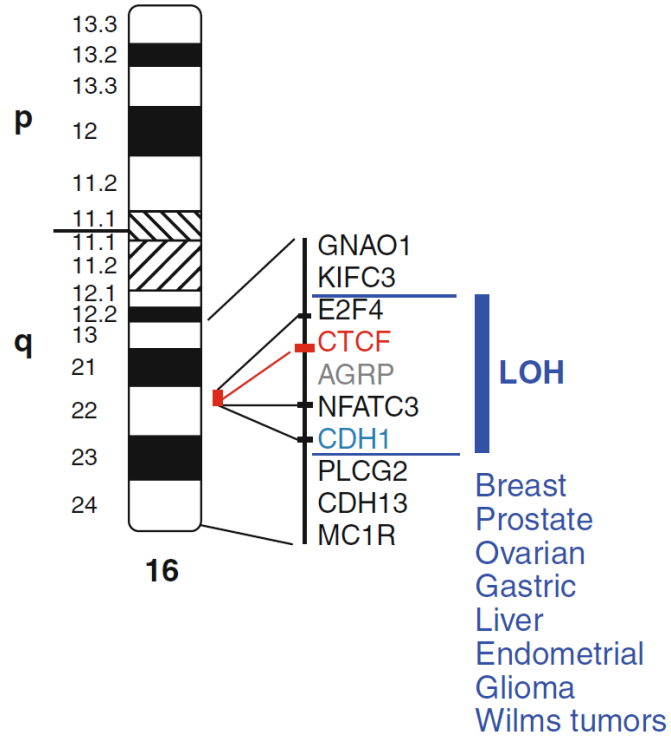
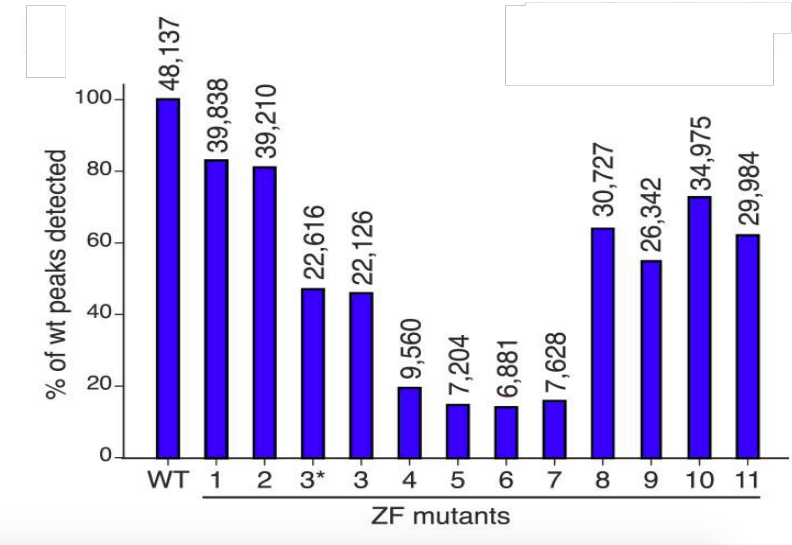


Figure 5. 16q22 locus. Loss of heterozygosity at chromosome 16, at the q.22 locus is found in several different cancers (adapted from Loukinov et al., 2016).

An earlier study explored the role of four point mutations in CTCF that were reported in breast, prostate, and Wilms' tumors. Using electrophoretic mobility shift assays (EMSAs), it was shown that mutations of CTCF disrupted its binding to some CTCF binding sites but not others. Further, using reporter assays loss of binding of CTCF at the promoters of several growth regulating genes such as *MYC*, *p19ARF*, and *Igf2* was demonstrated, while no loss of binding was seen in a β-globin insulator, lysozyme silencer and APP promoter. This suggests that CTCF mutations may deregulate activity of some but not other regulators (Filippova et al., 2002). Genome-wide binding profiles of biotin labeled CTCF with mutations in the zinc-coordinating histidines of ZFs showed a decrease in the number of sites occupied by CTCF. Mutations in the core ZFs (ZFs 4-7) are most greatly affected (Nakahashi et al., 2014) **[Figure 6]**. However this

study was performed in cells that harbor wild type CTCF along with the mutants, so it is likely



underestimating the effect of these mutants.

Figure 6. Mutations at the Zinc-coordinating amino acid reduce CTCF binding throughout the genome compared to WT CTCF. 1-11 implies mutations in the zinc coordinating amino acids of ZF1-11. 3* indicates R339W mutation, which is located on ZF3 but not a Zinc-coordinating amino acid (Nakahashi et al., 2014).

Further evidence for CTCF being a contributing factor to oncogenesis comes from mouse studies that show that hemizygous CTCF knockout mice are highly susceptible to cancer in many tissues. Additionally, tumors found in CTCF hemizygous mice have enhanced malignant progression. Total knockout CTCF is embryonically lethal in mice, confirming the importance of CTCF for development. CTCF haploinsufficient mouse genomes were found to have alterations in CpG methylation preceding development of cancer, suggesting that this altered epigenetic state may be leading to the development of cancer. In other words, CTCF is a haploinsufficient tumor suppressor and a better understanding of CTCF mutations may lead to new avenues for cancer therapy (Kemp et al. 2014) **[Figure 7]**.

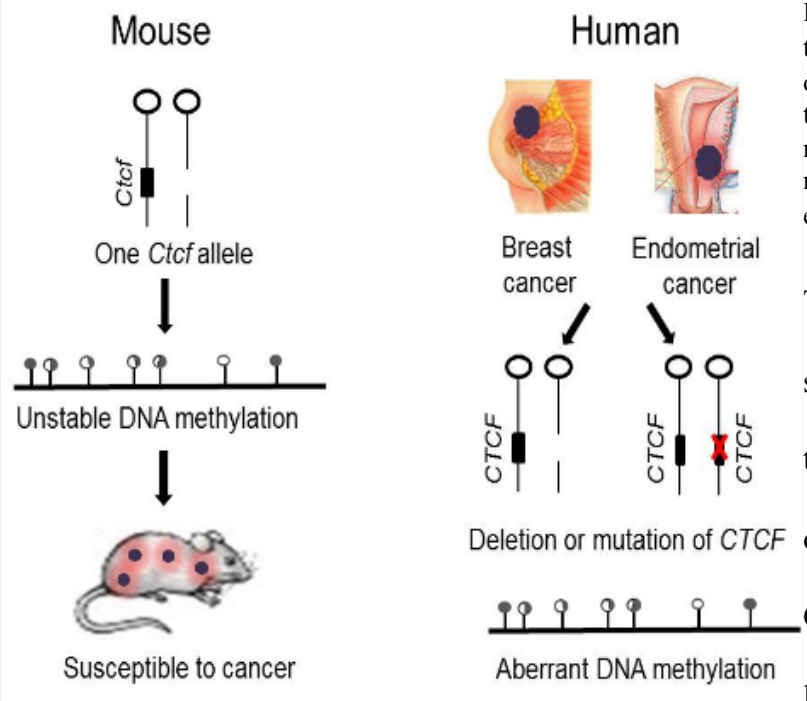


Figure 7. CTCF is a haploinsufficient tumor suppressor. Loss of function in one CTCF allele leads to susceptibility to cancer in mice due to unstable DNA methylation. In humans, deletions and mutations are associated with breast and endometrial cancer (Kemp et al., 2014).

There are several mutations associated with the amino acids that coordinate the zinc ion. Loss of coordination of the zinc ion in C2H2 and CH2C ZFs causes a loss of function of the entire molecule. Thus, these mutations likely mimic loss of heterozygosity. Mutations in the sites that make contact with DNA are also associated with cancer. It is also likely that these mutations mimic loss of heterozygosity because CTCF is unable to bind. But what is more interesting is the change of specificity in binding site recognition by mutated CTCF. This change in specificity, as shown in reporter assays, can lead to deregulation of chromatin throughout the genome. There are also mutations in the linker region between ZFs , but it is unclear how mutations in these sites affects the binding of the protein. It is possible that these amino acids are critical in interactions between two zinc fingers, or even other cofactors. Finally, there are cancer-associated mutations in other regions, such as the C-terminus, the N-terminus, or non-base associated and non-zinc-coordinating amino acids.

Knockout studies of CTCF are difficult because CTCF is a necessary factor for cell growth. Knockdown studies have been performed with siRNA, but inadequate levels of depletion

were seen as only a small amount of CTCF is required for TAD formation, and thus the results were inconclusive (Zuin et al., 2014) Depletion of CTCF with an auxin-inducible degree (AID) system, which leads to a more complete depletion of CTCF, removes insulating neighborhoods, and restoration of CTCF restores the loop structure **[Figure 8].** Interestingly, depletion of CTCF does not strongly affect nuclear compartmentalization - only about ten percent of compartmentalization is affected. However, loss of cohesin strengthens compartment structure suggesting there is some kind of opposition between TAD structure and compartmentalization (Rao et al., 2017). Clearly, the formation of TADs and nuclear compartments are governed by separate mechanisms (Nora et al. 2017).

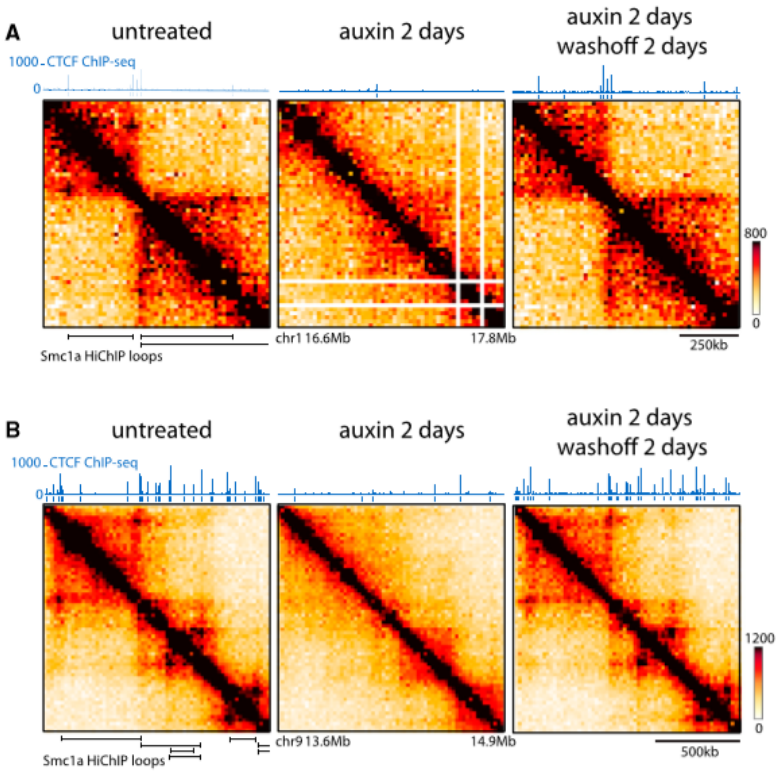


Figure 8. TADs disappear after auxin induced degradation of CTCF, and after removal of auxin TAD structure returns. **A)** 250kb scale, removal of a TAD is observed after 2 days of auxin treatment. **B)** 500kb scale,TADs are removed and CTCF no longer appears in ChIP-seq data (Adapted from Elphege et al., 2017).

2.5 MUTATIONS AT AMINO ACID R339:

My work has been focused on one mutation in ZF3, R339. R339 makes base-specific contact with DNA **[Figure 9]**. Mutations of R339 have been found in Wilms' tumor, B-cell progenitor acute lymphoblastic leukemia, and stomach and colorectal cancers. Previously, using an EMSA against

CTCF target sites, it has been shown that missense mutations of R339 abrogate binding to some sites. This leads to a selective loss of CTCF activity, which may affect tumor progression and generally lead to dysregulation of the 3D structure of the genome. Reporter assays have found that R339W mutation greatly reduces CTCF binding at the p19ARF promoter compared to wild type, implicating this mutation in affecting a growth-regulating gene (Filippova et al. 2002). R339W also greatly reduced the number of binding sites genome wide. While the wild type CTCF was bound to 48,137 sites across the genome, the mutation at R339 resulted in binding of the protein to only 22,616 sites (Nakahashi et al., 2014) **[Figure 9]**. Interestingly, none of the Wilms' tumors sampled in Filippova et al. 2002 were found to have hemizygous loss of function in the CTCF allele - every tumor had a ZF base recognition site altering mutation. It is possible that the alteration of the binding site provides a selective growth advantage in malignant cells (Kemp et al., 2014).

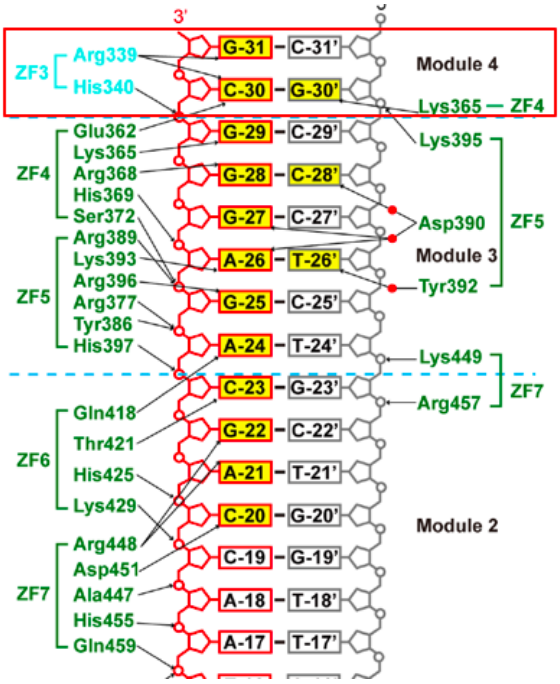


Figure 9. R339 makes base specific contact with DNA (adapted from Yin et al., 2017).

In the present study my aim is to decipher the role of R339Q, a missense mutation in the third zinc-finger of CTCF for its effect on binding affinity, binding site specificity, and cellular functions. This will initially be assessed using a bacterial one hybrid system that allows screening of a large synthetic library of sequences. The mutations will then be studied in mouse embryonic stem cells where endogenous CTCF is depleted with a degron and rescued with mutant transgenes.

3.0 MATERIALS & METHODS

3.1 BACTERIAL STRAINS AND CELL LINES

3.1.1 BACTERIAL STRAINS

The *E.coli* strain (USOΔ*hisB*Δ*pyrF*Δ*rpoZ)* that lacks *hisB* (the bacterial homolog of *HIS3*), *pyrF* (the bacterial homolog of *URA3*) and omega subunit of RNA polymerase (as *rpo* gene is replaced with zeocin) that is used for the bacterial one-hybrid assay, was kindly provided by the Marcus Noyes (MB Noyes, 2008, Wolfe et al., 2005).

E. coli strains XL1-Blue and stabl2 were used for construction and extraction of plasmids.

3.1.2 MAMMALIAN CELL LINES

EN 156 CTCF-AID-eGFP mouse embryonic stem cells were kindly provided by Elphège Nora at the Bruneau laboratory (Nora et al., 2017).

3.2 MAINTENANCE OF MOUSE EMBRYONIC STEM CELL (MESC) CULTURE

Feeder free culture of mouse embryonic cells was performed on 10 cm plates pre-coated with 0.1% gelatin. The cells were cultured in DMEM (Life Technologies, 11965-118), supplemented with 10% Fetal Bovine Serum (Fisher, SH30071.03), 550 μM beta-mercaptoethanol

(Sigma, M6250), 1 mM Sodium Pyruvate (Life Technologies, 11360-070), glutamax (Life Technologies, 35050-061), MEM non-essential amino acids (Life Technologies, 11140-050) and Leukemia Inhibitory Factor (LIF) (Millipore, ESG1107). A density of 0.2-1.5 x 10^5 cells / cm^2 was maintained by passaging cells every 24 to 48 hours using TrypLE™ at 37 ^{O}C and 5% CO_2. When required the culture media was supplemented with puromycin (1 μg/mL).

3.2.1 FREEZING & THAWING OF MESCS

To freeze mESCS, cells were dissociated using TrypLE™ Express Enzyme, phenol red (Thermo Fisher, 12605010), washed in PBS and resuspended ESGRO complete serum-free cell culture freezing medium (Millipore Sigma, SF005). The cryovials containing the cells were placed in an isopropanol chamber and stored at -80°C overnight before transferring to liquid nitrogen.

Frozen cells were thawed rapidly at 37 ^{O}C and plated in pre-warmed complete media to ensure survival.

3.3 PREPARATION OF COMPETENT CELLS

E.coli in log phase was used for preparation of competent cells. On reaching log phase, cell division was halted by incubating cells on ice for 30 minutes. Cells were then harvested by centrifugation at 3000g for 15 minutes at 4°C. Cells were resuspended in 10 mL cold 0.1 M $MgCl_2$, centrifuged and resuspended in 10 mL cold 0.1 M $CaCl_2$, and centrifuged again. Cells were then resuspended in a solution of ice cold 85 mM $CaCl_2$ and 15% glycerol and subsequently snap frozen in liquid nitrogen and stored at -80°C.

3.4 CONSTRUCTION OF PLASMIDS

3.4.1 CONSTRUCTION OF PLASMID HARBORING ZINC FINGERS 2 TO 9 OF CTCF

Zinc fingers 2 to 9 of CTCF were cloned as a fusion to the omega-subunit of RNA polymerase under the control of the lac5 promoter.

The region corresponding to zinc fingers 2 to 9 of CTCF was amplified from a plasmid that harbors human CTCF cDNA (Sino Biological, HG15017-G) using CTCF ZF2 forward and CTCF ZF-9 reverse primers harboring KpnI and XbaI restriction sites respectively **(Table 1)**. The PCR product was purified, digested with Kpn1 (NEB, R3142) and XbaI (NEB, R0145) in CutSmart (NEB, B7024) buffer for one hour at 37 ^{O}C, gel purified and ligated into a pB1H2ω5 vector using T4 DNA Ligase (NEB, M0202L) for 30 minutes at room temperature. The plasmid was then transformed into calcium competent cells through a heat shock method. The resulting cells were plated on LB plates containing ampicillin to select for cells containing the plasmid. The clone was confirmed by sequencing (Macrogen).

3.4.2 SITE DIRECTED MUTAGENESIS OF PLASMID FOR BACTERIAL ONE-HYBRID ASSAY

Site-directed mutagenesis of the R339Q mutation at ZF3 was performed using a QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent, 210518) following the manufacturer's protocol. Oligonucleotides were designed using the QuikChange Primer Design tool (Agilent) and analyzed to ensure efficiency using the Integrated DNA technologies web tool. Plasmids were transformed into *E.coli* XL1-Blue Competent cells.

Plasmid vector harboring the binding site library was kindly provided by the Noyes laboratory.

3.4.3 SITE DIRECTED MUTAGENESIS OF PLASMID FOR EXPRESSION IN MESC

A pEN366 plasmid containing a doxycycline-inducible CTCF (pEN366) with Flag tag and mRuby2 was provided by the Bruneau laboratory.

Site-directed mutagenesis of plasmids for use in mESCs was performed according to the protocol described by Heckman & Pease, 2007.The R339Q mutation was introduced using a two-tiered site-directed mutagenesis system. Primers were created using Benchling and analyzed using Integrated DNA technologies web tool. Mutations were introduced using an internal forward (DB1) and reverse primer (DB2) to create overlapping PCR products harboring the missense mutation. A subsequent PCR was performed with these segments as template (primers FLSH and p366 Reverse) in order to introduce the mutation into the complete cDNA **[Figure 10]**. Primer sequences used are shown in **Table 1**.The PCR product with the mutation, R339Q was digested using AflII (NEB, R0520) and NotI-HF (NEB, R3189) for one hour at 37 ^{O}C in CutSmart buffer and gel purified. The insert was ligated to the pEN366 vector backbone from which WT CTCF was removed following digestion with NotI and AflII (NEB, M0202L). The resulting plasmid was transformed into *E. coli* stabl2 competent cells and grown on LB plates with ampicillin. Plasmid was purified from isolated colonies using Qiagen MiniPrep Plasmid Purification kit (Qiagen, 27106) and sequenced.

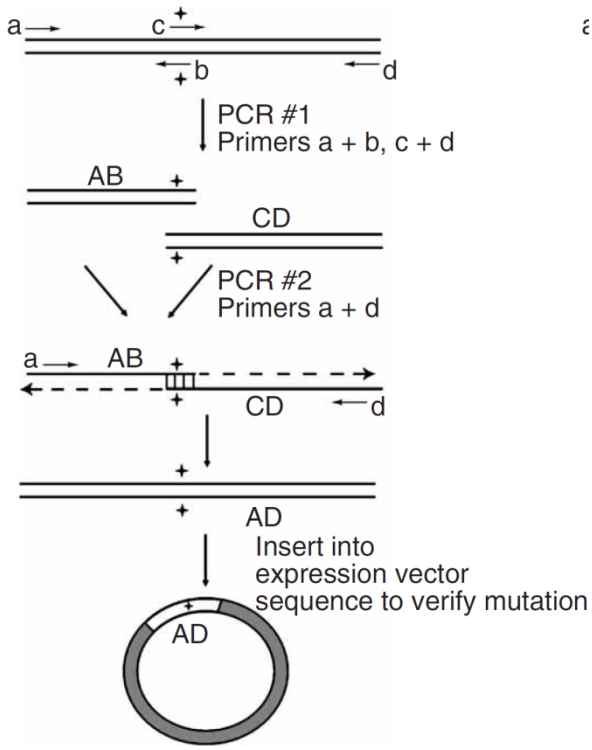


Figure 10. Site directed mutagenesis using PCR overlap extension. PCR#1 is performed using primers "a" and "b" and primers "c" and "d". The products of this PCR are used as a template for PCR#2, which is performed using primers "a" and "d". This amplifies the mutation introduced by the internal primers "b" and "c". (Adapted from Heckman & Pease, 2007)

3.5 BACTERIAL ONE-HYBRID ASSAY

3.5.1 BACTERIAL ONE-HYBRID ASSAY FOR DETERMINING AFFINITY OF BINDING OF CTCF R339Q MUTANT TO CONSENSUS CTCF BINDING SITE

Binding affinity of R339Q for the consensus CTCF binding site was assayed using a protocol adapted from Marcus Noyes (MB Noyes, 2008). A GHUC/Binding site reporter vector with CTCF consensus binding site was transformed along with the pB1H2ω5 expression vector harboring WT CTCF 2-9/mutant R339Q/Zif 268/OriO into US0Δ*hisB*Δ*pyrF*Δ*rpoZ* by electroporation (1800V) and recovered in 10 mL of SOC media for 1 hour at 37°C. Resulting cultures were diluted and grown overnight at 37°C on plates containing 2xYT media, 1% glucose, 1 µg/ml kanamycin, and 1 µg/ml ampicillin.

Three colonies from these plates were inoculated again and grown for 8 hours to ensure cells were in log phase. 2.5 µl of this culture was inoculated in 5 ml NM media **(Table 2)** with 50 µl of 10% histidine and 25 µl of 20 mM uracil. Resulting cultures were grown overnight at 37°C.

The samples were then centrifuged, washed with PBS, and resuspended in 1 ml of PBS+FBS and 100 µl were transferred to FACS tubes.

Fluorescence-assisted cell sorting was performed in Sony SH800Z using the corresponding excitation wavelength for GFP and mCherry. Gates were applied to forward-scatter/side-scatter dot plots to exclude nonviable cells and cell debris. FACS results were analyzed using FlowJo version 10.4.2. The percentage of positive cells was calculated against the background set on negative control (OriO). The percentage of cells expressing GFP was plotted using ggplot2 version 2.2.1.

3.5.2 BACTERIAL ONE-HYBRID ASSAY FOR DETERMINING SPECIFICITY OF BINDING OF CTCF R339Q MUTANT

The binding site of R339Q was determined with a library that harbors a 28bp randomized region in H3U3 vectors. This library was kindly provided by the Marcus Noyes laboratory. Expression vectors were prepared using a Qiagen Plasmid Maxi Plus Kit (12963) and precipitated using dry ice to ensure high concentration.

Library plasmids and CTCF R339Q-pB1H2ω5 expression vectors were transformed into competent cells by electroporation (1800V). Cells were rescued using 5 ml pre-warmed SOC media and recovered for 1 hour at 37°C. Cells were then pelleted, resuspended in 5 ml NM media with10% histidine, 1 mM IPTG, 100 µg/ml ampicillin, and 30 µg/ml kanamycin, and expanded for 1 hour at 37°C. Cells were then pelleted again, resuspended in NM without histidine, transferred to clean tubes, pelleted and washed. Next, cells were resuspended in NM media and

stored overnight at 4°C. Serial dilutions of a small aliquot of resuspended cells were performed and plated on LB media (with 100 µg/ml ampicillin and 30 µg/ml kanamycin) to determine the number of cells that have taken in both plasmids.

Appropriate volumes of cells corresponding to 5 X 10^7 colonies were then plated on NM media with ampicillin, kanamycin and 5 mM 3AT. The cells from the plate were scraped into a single tube and the plasmid was prepared using a Qiagen MiniPrep Plasmid kit. The resulting plasmids were then amplified by PCR with barcoded primers and sent for Illumina sequencing.

3.6 GENE TARGETING IN MOUSE EMBRYONIC STEM CELLS

3.6.1 NUCLEOFECTION

The plasmids for nucleofection were prepared using a Qiagen Plasmid Maxi Plus Kit (12963) and precipitated according to manufacturer's protocol. In order to knock-in CTCF R339Q, 20 µg of the R339Q plasmid, containing a doxycycline-inducible promoter, and 2.5 µg of pX330-EN1201 containing an sgRNA (ACTGCCATAACACCTAACTT) to the Tigre locus (provided by the Bruneau lab) were introduced into mESCs using a SF Cell Line 4D-Nucleofector® X Kit (Lonza, V4XC-2024) according to manufacturer's procedure.

3.6.2 COLONY SELECTION

Following nucleofection, cells were seeded into a 10 cm plate for 48h in media without antibiotic, after which the cells were split into a 1:10 and 1:100 dilution and puromycin was introduced into the media to select for cells which expressed pEN366 CTCF R339Q and thus had puromycin resistance. A plate containing non-nucleofected mESCs grown in the presence of puromycin served as a control. When cells were completely selected from the control plate by puromycin, the cells remaining in the nucleofected plate were ready to be screened. The individ-

ual colonies were then picked under a microscope and transferred to wells of a 0.1% gelatin coated 96-well plate containing TripLE. Media was then added to the wells and cells were allowed to grow for 5 to 7 days.

3.6.3 GENOTYPING

Genotyping was performed as per Elphège Nora and Edith Heard (Nora & Heard, 2013). The 96 well plates containing nucleofected colonies were frozen for 24 h to crack the cells. Cells were then treated with a lysis buffer containing 10 mM Tris-HCl, 10 mM EDTA, 0.5% SDS, 10 mM NaCl, and 1 mg/µL Proteinase K, and incubated at 60^{O}C overnight. Wells were then treated with a 75 uM NaCl and ethanol precipitation solution and incubated at room temperature for 30 min. The plates were then spun at 4000 rpm at 4^{O}C for 30 min, after which 150 ul 70% ethanol was added to each well, and the cells were spun for a further 15 min. PCR amplification was performed with DNA from each well with a primer mix containing three primers, AU1, AU2, and EN1566. AU1 and AU2 encode a sequence at the Tigre locus where pEN366 is targeted to integrate into the genome. EN1566 targets a site on pEN366. Because the third primer targets a site only found on pEN366 which is inserted into the genome, genotyping samples which yield a band at 1.4 kb contain the introduced CTCF at the Tigre locus. Cells which have a band at 1.2 kb do not have the plasmid introduced at the Tigre locus, and cells which have both bands are heterozygous. Wells containing the 1.4 kb strand were selected and sent for sequencing to confirm the result.

4.0 RESULTS

4.1 CREATION OF INSERT BY SITE-DIRECTED MUTAGENESIS.

The R339Q insert was successfully created by site directed mutagenesis. The reason a kit was not used was because of a high number of repeats on the pEN366 plasmid which renders it unsuitable to replication. The PCR amplification product was successfully run on 1.5% agarose to a band corresponding to the correct protein length and was sequenced by Illumina sequencing to confirm the existence of the mutation, as shown in **Figure 11.**

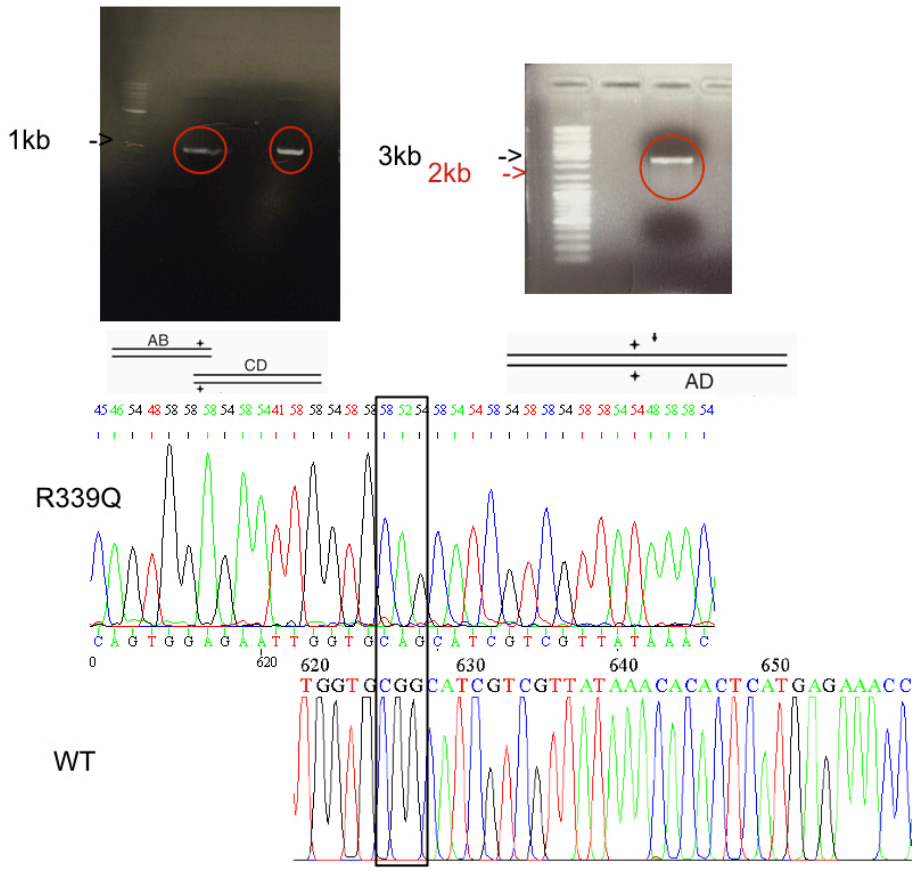


Figure 11. Site directed mutagenesis result. The site directed mutagenesis was performed successfully, with the band size in the second gel ~2.5kb. The sequencing result (below) contains the mutation.

4.2 CTCF MUTATIONS ARE FOUND IN SEVERAL CANCERS.

To determine the frequency of occurrence of cancer mutations in CTCF, we looked on cBioPortal, a cancer study database **[Figure 12]**. Mutations in CTCF were found in 951 samples. Of these, 600 were missense mutations, 434 were truncating mutations, 4 were classified as in-frame, and 8 were classified as "other." The "other" category included one unknown mutation, one fusion mutation, and 6 non-start mutations.

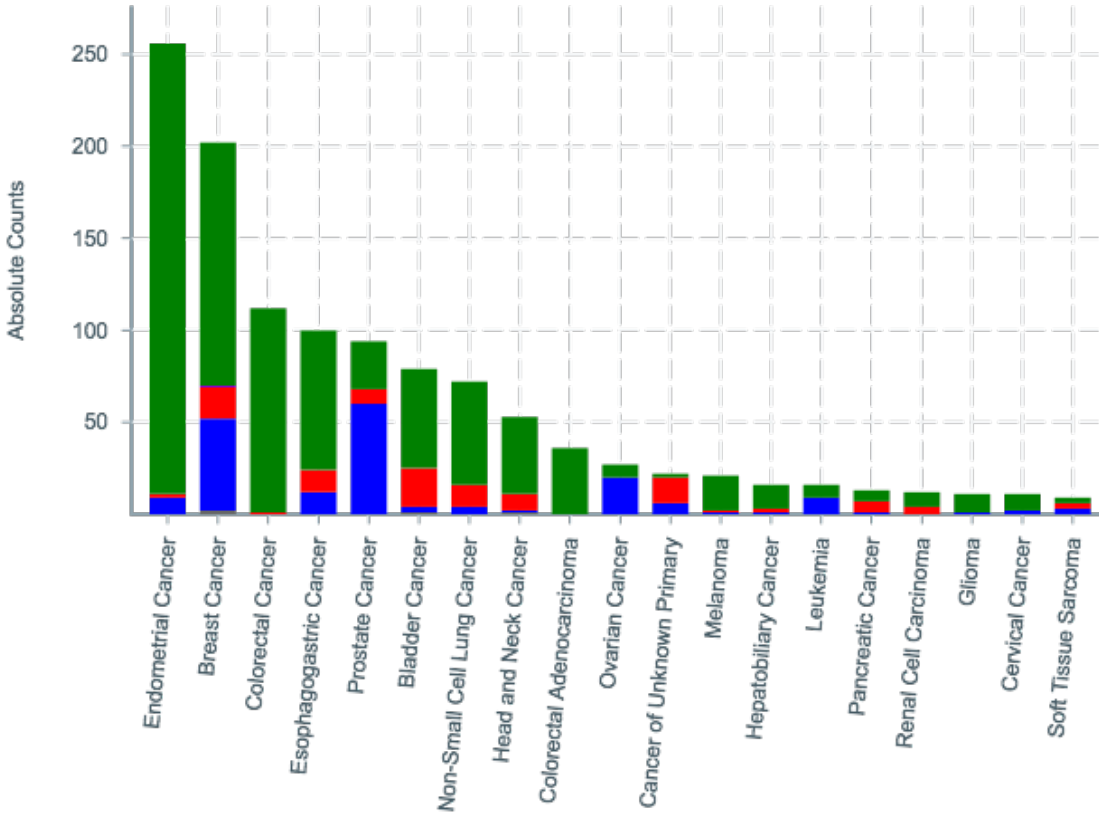


Figure 12. cBioPortal cancer list. This chart shows a portion of the 54640 samples for "CTCF" available on cBioPortal. Most cancers involving CTCF are related to mutations.

The analysis identified 64 mutations in amino acids in the ZFs of CTCF. All 64 of these mutations are predicted to be deleterious through the Sorting Intolerant From Tolerant (SIFT) algorithm. 34 mutations were predicted to be oncogenic by OncoKB and the remaining 30 were "Unknown."

We decided to study a subset of CTCF ZF mutations identified from the databases like TumorPortal (http://www.tumorportal.org/), ICGC Data Portal (https://dcc.icgc.org/) and COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/) as well as published literature based on position in zinc finger **(Table 3)** and frequency of occurrence in cancer [**Figure 13]**.

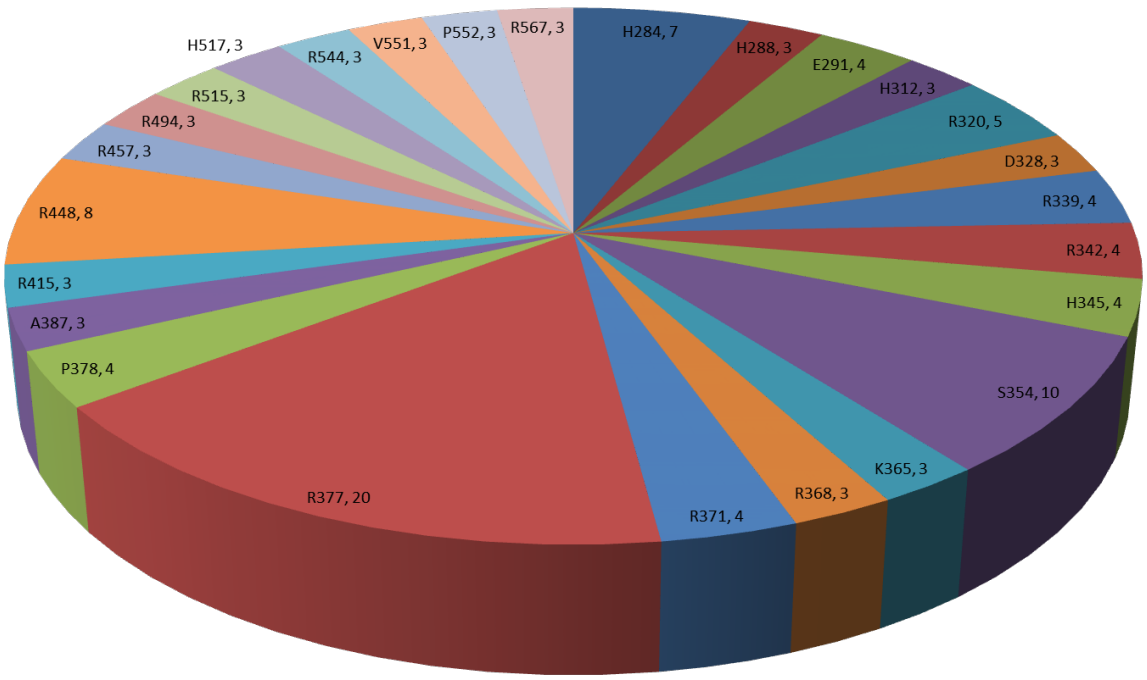


Figure 13. The proportion of mutations found in CTCF ZF.

The mutations were identified in several cancers including breast carcinoma, endometrioid carcinoma, squamous cell carcinoma, urothelial bladder carcinoma, adenocarcinoma, kidney renal clear cell carcinoma, Wilms' tumors, acute lymphoblastic T and B cell leukemia, multiple myeloma, acute myeloid leukemia, diffuse large B-cell lymphoma, prostate, colorectal, lung, stomach, large intestine, liver, serous and skin carcinoma, malignant melanoma, medulloblastoma and head and neck cancer **(Table 3)**.

4.3 MUTATIONS IN ZFS OF CTCF HAVE VARYING IMPACT ON BINDING AFFINITY TO ITS CONSENSUS BINDING SITE.

A bacterial one-hybrid (B1H) system was used to determine the impact of CTCF ZF mutations on binding affinity to the consensus binding sequence. B1H assays uses the protein of interest as 'prey' and its binding site as 'bait.' This system contains three key components: the expression vector fused with the omega-subunit of RNA polymerase ('bait'), a reporter vector that contains the binding site as well as GFP, HIS3, and URA3 ('prey'), and finally, the bacterial cell strain that is auxotrophic for HIS3 and URA3 (US0Δ*hisB*Δ*pyrF*Δ*rpoZ*). If the protein encoded by the expression vector recognizes the sequence on the reporter vector, the omega-subunit will initiate transcription of the reporter plasmid leading to expression of GFP **[Figure 14]** and HIS3 giving it a survival advantage in media lacking HIS in the presence of 3AT, an inhibitor of histidine synthesis. We adapted the system to harbor CTCF ZF 2-9 as the 'protein of interest' and incorporated the consensus CTCF binding site (CCAGCAGGGGGCGCT) in the reporter vector. The expression of GFP was analyzed by FACs to determine the binding affinity of wild-type or mutant CTCF to the consensus binding site. Wild type CTCF along with its binding site reporter was used as positive control in B1H assay. OriO was used as a negative

control. OriO encodes a ZF protein with a frameshift, leading to an inactive ZF protein and thus it should not activate the reporter cassette.

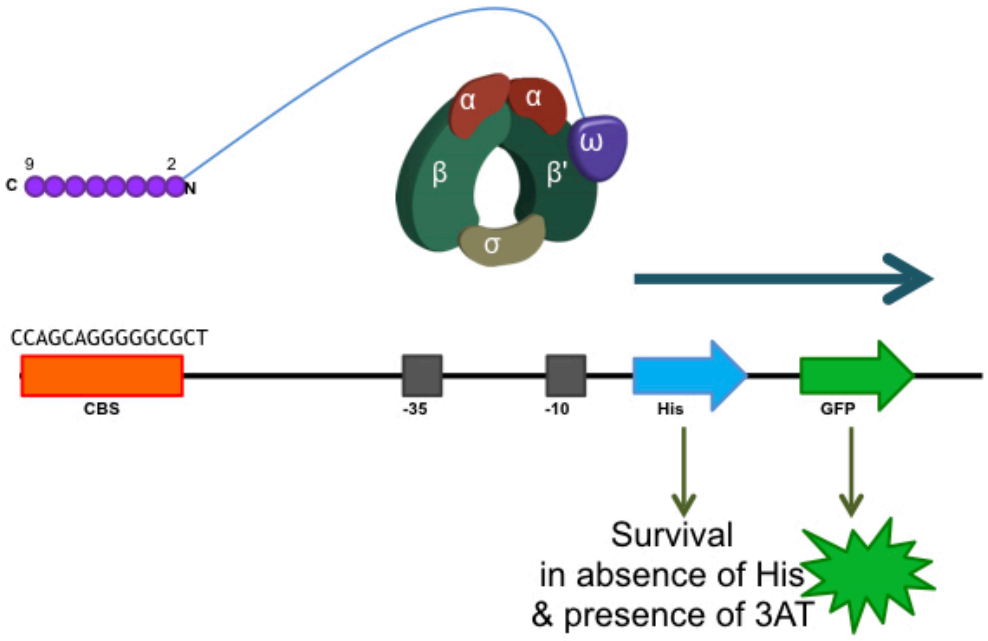


Figure 14. Bacterial one-hybrid for affinity assay. The His region of the plasmid allows for the screening, and the GFP is what causes fluorescence. An interaction is necessary between the CTCF 2-9 in the expression vector and the CTCF binding site in the reporter vector. CTCF 2-9 was also substituted with our mutated versions of CTCF.

4.3.1 MUTATIONS IN ZINC-COORDINATING AMINO ACIDS ABROGATE CTCF BINDING TO THE CTCF BINDING SITE.

Two mutations C356F [ZF4] and H455R [ZF7] that coordinate the zinc ion of ZF were tested in the B1H assay. 1.89% of cells expressed GFP for C365F and 1.18% of cells expressed GFP for H455R, compared to 88.9% for wild-type CTCF ZFs 2-9. The near total loss of binding caused by both mutations supports previous findings that suggest these mutations lead to a loss

of zinc coordination and thus loss of ZF function **[Figure 15]**.

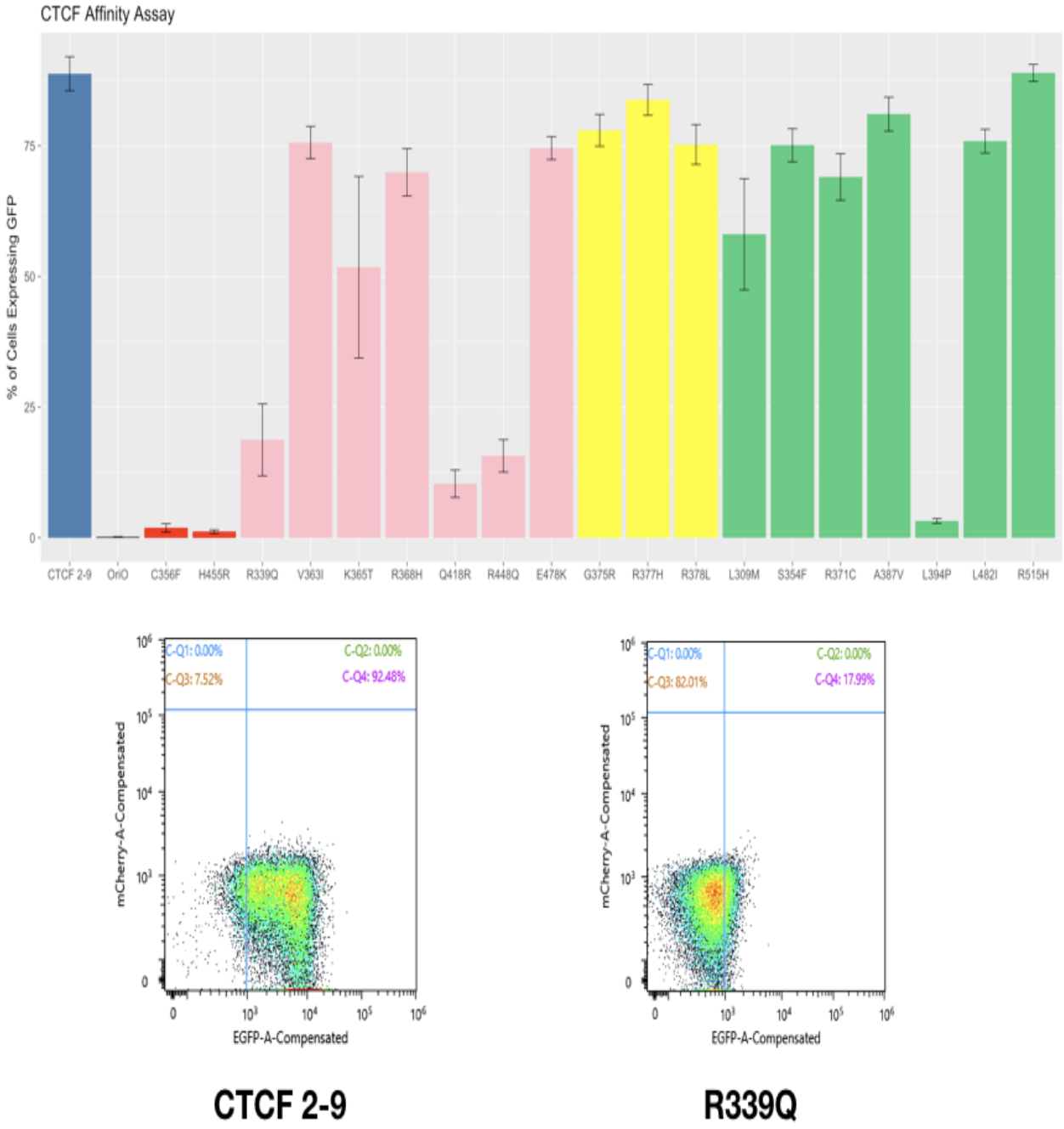


Figure 15. Bacterial one-hybrid GFP expression for WT and R339Q. A) R339Q had a greatly reduced strength of binding compared to the wild type CTCF. CTCF 2-9 = 88.8%, R339Q = 18.7%, OriO = 0.17%. B) Results from one run of FACS showing 92.48% of cells were in quadrant 4 (Q4) in the wild type, whereas only 17.99% of cells were in Q4 in R339Q, indicating a change in GFP expression in the two samples.

4.3.2 MUTATIONS IN BASE-RECOGNITION AMINO ACIDS HAVE MODERATE TO LARGE REDUCTION IN CTCF BINDING.

Base-recognition amino acids occupy positions -1, 2, 3, and 6 of a ZF. Mutations in base recognition amino acids, R339Q [ZF3], V363I [ZF4], K365T [ZF4], R368H [ZF4], Q418R [ZF6], R448Q [ZF7], and E478K [ZF8], lead to a moderate to severe loss of GFP expression. Q418R had the lowest GFP expression of the cohort, followed by R448Q and R339Q. These amino acids are near to the "core" of the ZF region, which has been shown to be more important in CTCF binding than peripheral ZFs.

Using the bacterial one-hybrid system and FACS to measure binding affinity of R339Q to the CTCF binding site, we found that R339Q has a marked decrease in GFP expression compared with WT CTCF **[Figure 14]**. An average of 18.7% of bacteria with the R339Q mutation expressed GFP, compared to 88.8% of cells for wild type CTCF 2-9. This shows that R339Q does not strongly bind the CTCF consensus binding site, consistent with previous results showing that mutations at R339 lead to a change in binding site preference (Filippova et al., 2002). For the rest of my project I decided to focus on this specific mutation.

4.3.3 MUTATIONS ELSEWHERE DID NOT GENERALLY REDUCE BINDING.

Mutations in the linker region and elsewhere did not have a large reduction in CTCF binding, except for L394P. L394P is a proline substitution, which likely introduces a disruption in the alpha helix and causes the entire ZF to be unable to bind. Comparing L394P (3.82% of cells expressed GFP) to L482I (75.9% of cells expressed GFP), which both occupy the same position on their relative ZFs (ZF5 and ZF8, respectively), provides evidence for this, as the Leucine to Isoleucine mutation in L482I does not introduce a helix disruption, and both amino

acids are hydrophobic, aliphatic and nonpolar. Therefore L482I is likely a silent mutation **[Figure 16]**.

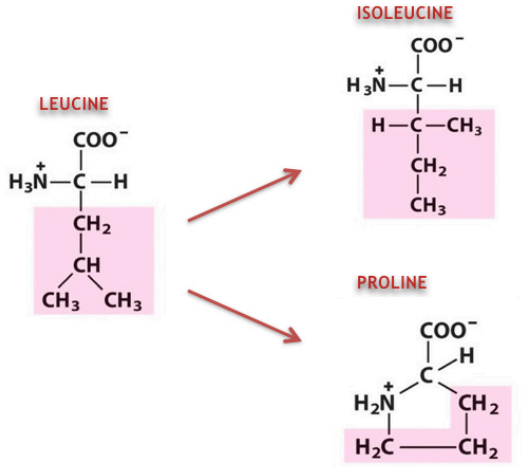


Figure 16. L394P mutation. The change from leucine to proline likely causes a disruption in the alpha helix of ZF5, whereas a mutation from leucine to isoleucine causes little effect.

4.4 R339Q HAS AN ALTERED BINDING SITE CONSENSUS SEQUENCE.

An additional use for the B1H system is to determine how effectively a zinc finger binds to the randomized library of DNA sequences that will turn characterize the binding specificity. Here, a library of $>10^7$ different sequences in a 28 bp window was used as the bait instead of the consensus CTCF binding site. The library sequences that R339Q binds can be found through sequencing the PCR products of the reporter plasmids using primers flanking these sequence. The cells are selected in HIS negative, 3AT positive media. Different concentrations of 3AT provide information about the strength of binding of the colonies: at high concentrations of 3AT only cells with the strongest expression/reporter plasmid interactions will survive, whereas in lower concentrations cells with relatively weak interactions can also survive **[Figure 17].**

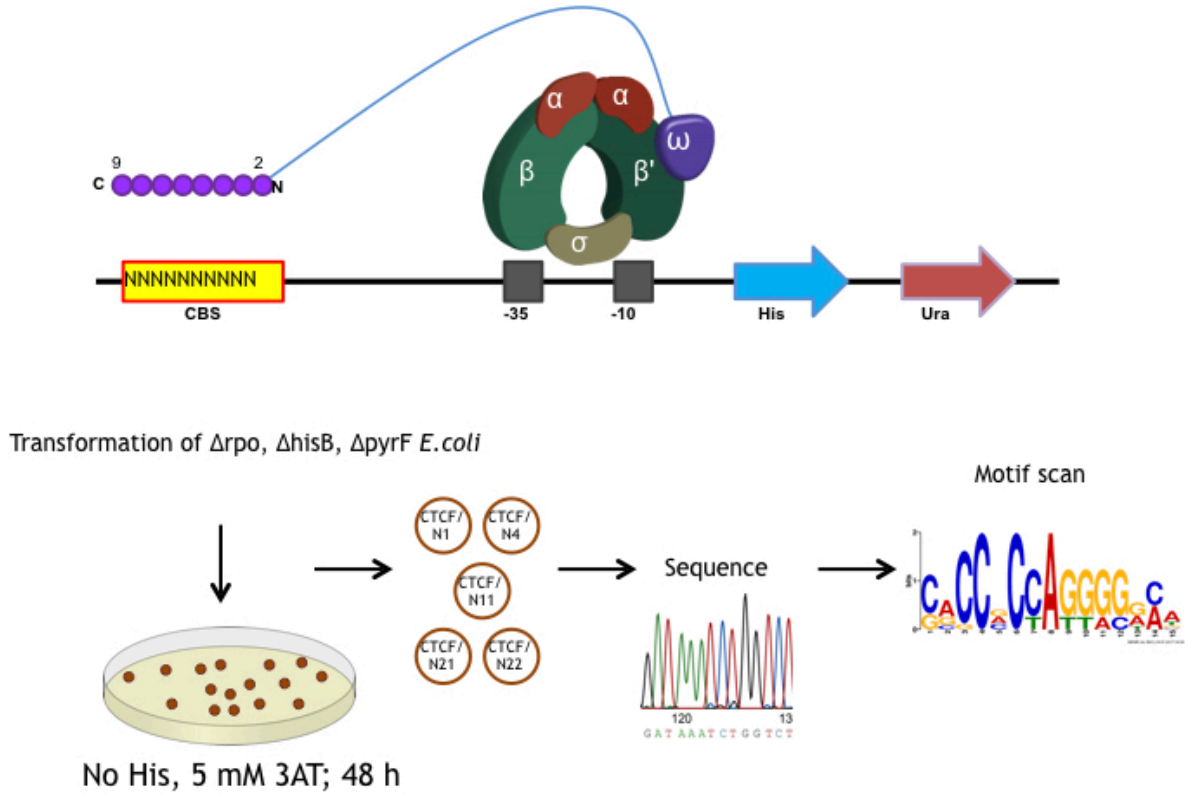


Figure 17. The bacterial one-hybrid system for the specificity assay.

Self-activating sequences in the library interfere with this system, so they must be removed prior to selection. Reporter cells were grown on minimal medium with uracil and 5-fluoro-orotic acid in order to remove cells which self-activate the reporter gene. This counter-selection eliminates self-activating sequences from the library and allows for the proper characterization of the expression vector binding site without interference.

Using this system, the binding site motif of CTCF 2-9 was determined and found to match that of the ENCODE motif **[Figure 18]**.

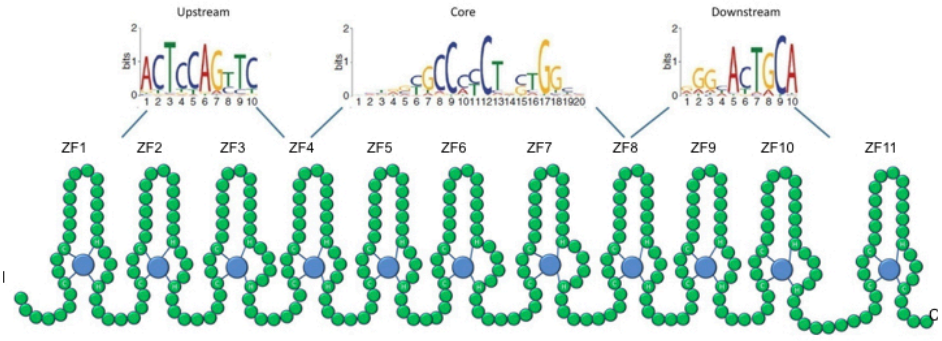


Figure 18. CTCF consensus sequence.

The binding site motif of R339Q was also determined using this system **[Figure 19]**. Interestingly, there was no change in the specificity of the amino acids recognized at the ZF3 recognition site as predicted. However, there was an increased GC requirement in the mutant ZF3 compared to wild type. Additionally, a change was found instead at ZF6 and ZF7, indicating that there may be an interaction between these ZFs and ZF3 harboring the R339Q mutation.

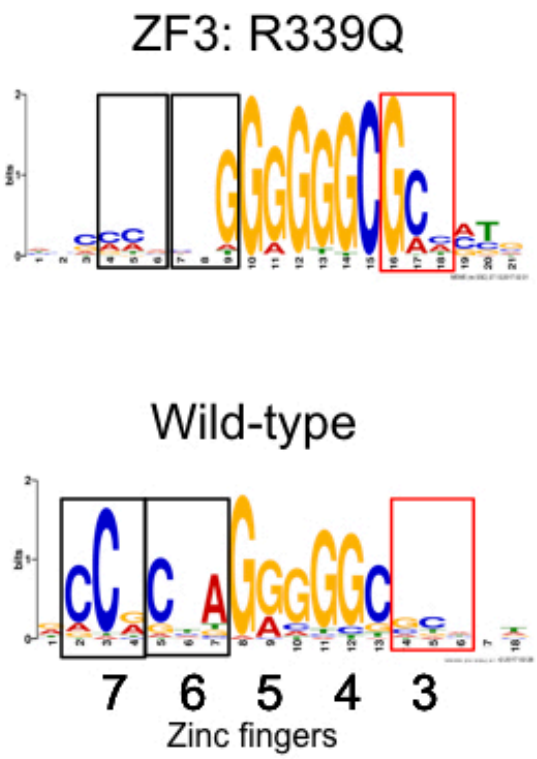


Figure 19. The R339Q consensus sequence from the B1H specificity assay.

4.5 ESTABLISHING A DEGRON SYSTEM TO STUDY THE IMPACT OF THE R339Q MUTATION ON CTCF FUNCTION

Mouse embryonic stem cells were nucleofected with R339Q-pEN366 and pX330-EN1201 and colonies were selected as described in the methods. The auxin-inducible degron (AID) system for targeted depletion of CTCF was employed to fully degrade CTCF and rescue expression with our mutated transgene. The cells have an AID and GFP tag that is expressed with their CTCF, as well as a Tir1 F-box protein **[Figure 20]**.

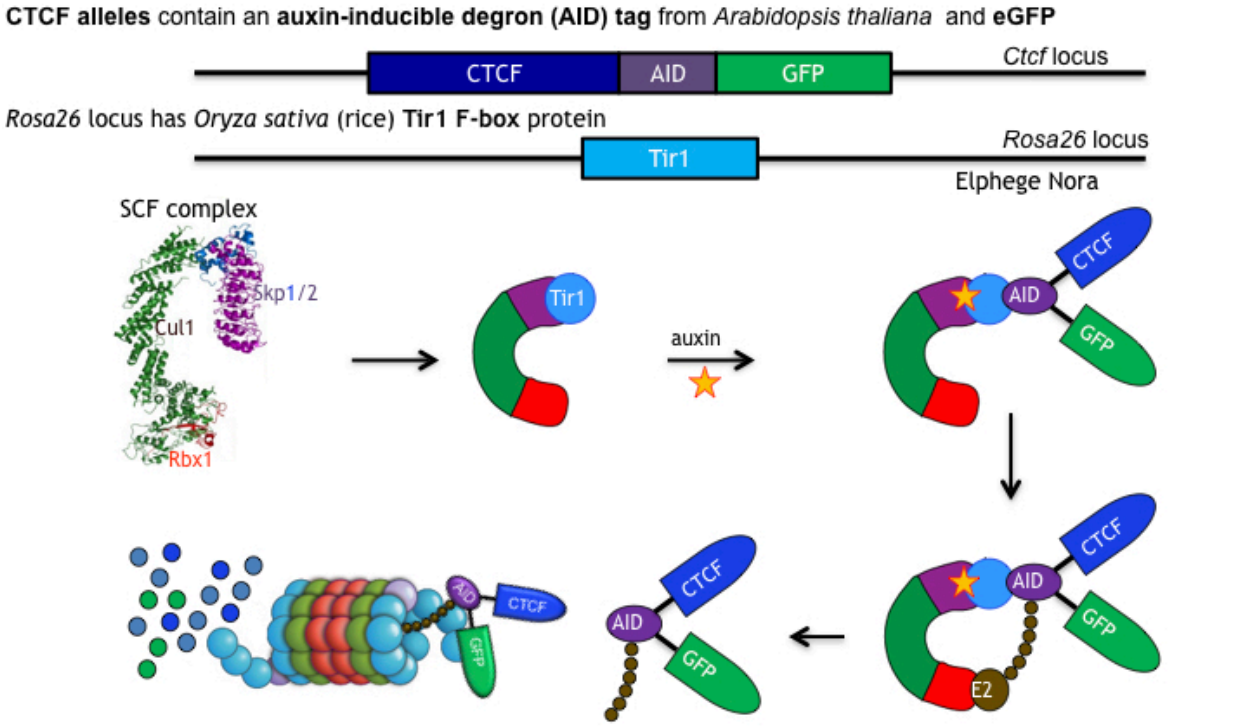


Figure 20. Auxin-inducible degron system. The SCF complex with Tir1 polyubiquitinates CTCF-AID-eGFP in the presence of auxin.

Tir-1 allows for the polyubiquitination of the CTCF-AID-eGFP fusion protein in the presence of auxin, degrading it nearly completely. In order to confirm that the nucleofection was successful, genotyping of isolated colonies were performed. These PCR products were used to determine whether the well contains a heterozygous WT CTCF/R339Q, an R339Q/R339Q, or WT CTCF/ WT CTCF. The results of the genotyping are shown in **Figure 21**.

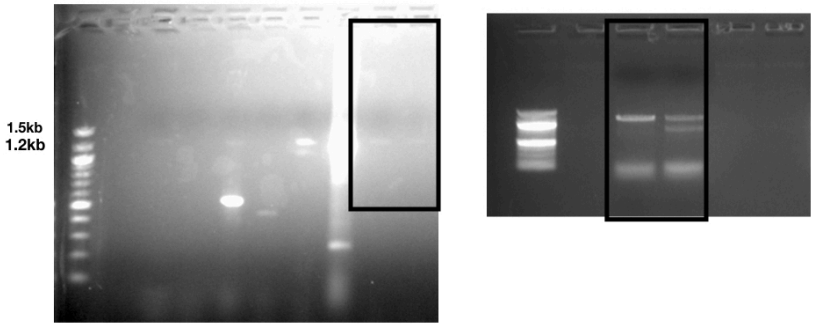


Figure 21. Genotyping of mESC at the Tigre locus. The gel was repeated with the sections highlighted. It appears one colony has the right insert resulting in a band at 1.4 kb while the other is heterozygous with bands at 1.4 and 1.2 kb.

5.0 DISCUSSION

CTCF is an important regulator of 3D genome architecture. Mutations in ZF 2-9 of CTCF can cause a change or abrogation of the CTCF binding site. These mutations are implicated in cancer due to their potential for altering the function of CTCF, which may lead to alteration of enhancer-promoter relationships of oncogenes. Additionally, mutations in zinc-coordinating amino acids on ZFs leads to a total abrogation of binding of CTCF, which is equivalent phenotypically to a heterozygous knockout of CTCF, which leads to altered methylation of the genome and various cancers.

Previous studies of the binding affinity of R339Q, a mutation in ZF3 that occurs in cancer patients, have found a reduced affinity to the CTCF binding site, but the technique used, EMSA, provides only limited sensitivity and can only identify whether or not binding is completely abrogated. Using a bacterial one-hybrid system, we found that R339Q causes a decrease of binding

affinity to the CTCF binding site. R339Q is a missense mutation at position 6 of the base-recognition site, and this result implies that CTCF undergoes a change in binding due to ZF3 now recognizing a new binding site. Because the affinity is not reduced to zero, as with other mutations and our negative control, we can see that R339Q does not completely abrogate binding to the CTCF binding site, and thus has a more subtle phenotype than knockout mutations. This change in binding could provide a selective growth advantage for cancer cells that would be lost with a more severe mutation. This change in binding site could lead to CTCF affecting promoters of genes it otherwise would not, and this may be why it is seen in some cancers. Previous studies show that mutated CTCF is a potential intervention target for cancer (Filippova et al., 2002), however, further characterization of the binding sites of these mutations is necessary to understand the effect mCTCF has on genome architecture.

The R339Q consensus sequence from the B1H specificity assay shows that there is a change in the binding site, although in an unexpected fashion. The binding affinity at module 3 for R339Q is similar to the wild type, but has a greater level of GC requirement, but at modules 6 the consensus sequence changes from C_A to __G as shown in **Figure 19**. Perhaps this change could be caused by ZF3 altering its specificity, shifting the binding of non-mutated ZFs, without altering the specificity of ZF3. However, this cannot be confirmed without the crystal structure of R339Q and comparing it to the wild type crystal structure.

A future direction of this research is to perform auxin degradation of CTCF on the mESCs using the cells that were developed over the course of this study. This system will be used to test the ability of R339Q to rescue survivability and gene expression from cells that have been completely deprived of wild type CTCF. Due to time constraints, this study could not be

performed at the time of writing, but the Skok lab will continue using the auxin-inducible degron system to study the effects of CTCF depletion on genome architecture.

5C and Hi-C can be performed to analyze how this mutation affects TAD structure. Interaction data from these tests will allow researchers to understand how these mutations are affecting the 3D structure of the genome at a high resolution. ATAC-seq, a technique using transposons to characterize DNA accessibility, can help uncover changes in gene activity in cells with active R339Q. ChIP-seq can be performed to find sequences where R339Q is binding, potentially uncovering cancer-driving promoters that may be affected. Finally, RNA-seq can be performed to analyze differential expression in R339Q cells, potentially finding cancer-associated genes becoming differentially expressed. Gene set enrichment analysis will be performed on this data in order to understand what gene sets may be more active in R339Q cells.

Further study of CTCF and its effects on genome architecture is important to uncover its effects on cancer. It is perhaps more important, however, to study how these genome architecture changes alter expression and DNA accessibility, as understanding this has broader implications. Alterations of genome architecture caused by other proteins, such as cohesin, will help synthesize this knowledge to more broadly understand the 3D genome.

6.0 REFERENCES

1. Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. *2*, 292–301.

2. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

3. Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. *16*, 1299–1309.

4. Fortin, J.-P., and Hansen, K.D. (2015). Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol. *16*, 180.

5. Filippova, G.N., Qi, C.-F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., et al. (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer Res. *62*, 48–52.

6. Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. Nature *447*, 413–417.

7. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

8. Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. Genes Dev. *30*, 881–891.

9. Gombert, W.M., and Krumm, A. (2009). Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. PLoS One *4*, e6109.

10. Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G., and Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol. Cell *66*, 711–720.e3.

11. Heckman, K.L., and Pease, L.R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. Nat. Protoc. *2*, 924–932.

12. Kemp, C.J., Moore, J.M., Moser, R., Bernard, B., Teater, M., Smith, L.E., Rabaia, N.A., Gurley, K.E., Guinney, J., Busch, S.E., et al. (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. Cell Rep. *7*, 1020–1029.

13. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

14. Lutz, M., Burke, L.J., Barreto, G., Goeman, F., Greb, H., Arnold, R., Schultheiss, H., Brehm, A., Kouzarides, T., Lobanenkov, V., et al. (2000). Transcriptional repression by the insulator protein CTCF involves histone deacetylases. Nucleic Acids Res. *28*, 1707–1713.

15. Meng, X., Brodsky, M.H., and Wolfe, S.A. (2005). A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. Nat. Biotechnol. *23*, 988–994.

16. Nakahashi, H., Kieffer Kwon, K.-R., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. *3*, 1678–1689.

17. Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

18. Nichols, M.H., and Corces, V.G. (2015). A CTCF Code for 3D Genome Architecture. Cell *162*, 703–705.

19. Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

20. Noyes, M.B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M.H., and Wolfe, S.A. (2008). A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. Nucleic Acids Res. *36*, 2547–2560.

21. Persikov, A.V., and Singh, M. (2014). De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. Nucleic Acids Res. *42*, 97–108.

22. Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U. S. A. *112*, E6456–E6465.

23. Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). Nat. Genet. *38*, 1348.

24. de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev. *26*, 11–24.

25. Xu, D.J., and Noyes, M.B. (2015b). Understanding DNA-binding specificity by bacteria hybrid selection. Brief. Funct. Genomics *14*, 3–16.

26. Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q., and Wang, Y. (2017). Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. Cell Res. *27*, 1365–1377.

27. Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

28. Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc. Natl. Acad. Sci. U. S. A. *111*, 996–1001.

7.0 TABLES

Table 1. Primers

Primer Name	Sequence	Description
CTCF ZF2 For	AATAAAGGTACCACTGATG AGAGACCACACAAG	The forward primer for ZF 2-9
CTCF ZF9 Rev	CCGCCCTCTAGAGTTAGGC GTAAGGCTTCTCCCCGGT	The reverse primer for ZF 2-9
DB1	AGTGGAGAATTGGTGC AGCATCGTCGTTATAAA CA	The forward internal primer containing an R-Q mutation at amino acid 339. Primer "c" in Figure 10
DB2	TGTTTATAACGACGATG CTGCACCAATTCTCCAC T	The reverse internal primer containing an R-Q mutation at amino acid 339. Primer "b" in Figure 10
AU1	GGGTGGGTACAAAGGT ACCAG	Genotyping forward primer
AU2	CCAGGACCCACAAAGT GAAG	Genotyping reverse primer
EN1566	GCGGTACCGCAAGCTT GATA	Genotyping reverse insert primer
FLSH	CGTGCTAGCGCGGCCG CA	Forward primer for entire CTCF protein with FLAG tag for use in Degron system. Primer "a" in Figure 10

Primer Name	Sequence	Description
p366 Reverse	TGGTTTGAGTTCCCATGTACGGAT	Reverse primer for entire CTCF protein with FLAG tag for use in Degron system Primer “d” in Figure 10
R339Q For	CAGTGGAGAATTGGTTCAGCATCGTCGTTACAAAC	Forward primer for R339Q in human CTCF, used in B1H system.
R339Q Rev	GTTTGTAACGACGATGCTGAACCAATTCTCCACTG	Reverse primer for R339Q in human CTCF, used in B1H system.

Table 2. Composition of NM media

Component	Volume (mL) for 500 ml media
10% Histidine	0.5
20nM Uracil	2.5
10X M9 Salts	50
20% Glucose	10
20mM Adenine HCl	5
Amino Acid Mixture (-His)	15
1M $MgSO_4$	0.5

Component	Volume (mL) for 500 ml media
10mg/mL Thiamine	0.5
10mM $ZnSO_4$	0.5
100mM $CaCl_2$	0.5
Bacto Agar (g)	7.5
H_20	370 (or 420 if no Histidine added)
Kanomycin	0.5
Ampicillin	0.5
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	0.5

Table 3. Mutations and corresponding zinc fingers.

Amino acid	Zinc finger	Associated cancer
Mutations in the amino acids that co-ordinate zinc ion		
C356	4	T cell acute lymphoblastic leukemia, head and neck cancer, stomach cancer
H455	7	Diffuse large B cell lymphoma
Mutations of the amino acids that make direct contact with DNA (positions -1, 2, 3 and 6)		
R339	3	Stomach, colorectal, Wilms', B-cell progenitor acute lymphoblastic leukemia
V363	4	T cell acute lymphoblastic leukemia
K365	4	endometrioid cancer
R368	4	serous carcinoma (ovary), Large Intestine carcinoma, prostate cancer
Q418	6	Primitive neuroectodermal tumour-medulloblastoma, Liver Carcinoma
R448	7	Wilms' tumor, endometrioid carcinoma, breast cancer
E478	8	Skin Carcinoma
Mutations of amino acids in the linker region between two zinc fingers		
G375	4-5	Lung carcinoma, endometrioid carcinoma, breast carcinoma
R377	4-5	Endometrium, stomach, Acute myeloid leukaemia, High-risk childhood B-precursor Acute lymphoblastic leukaemia, large intestine carcinoma, breast cancer
P378	4-5	Breast carcinoma, endometrioid carcinoma , Acute lymphoblastic leukaemia
Mutations of amino acids in the other regions of the zinc fingers		
L309	2	endometrioid cancer
S354	4	Breast carcinoma, squamous cell carcinoma, melanoma, endometrioid carcinoma, malignant melanoma, urinary tract carcinoma, Cholangiocarcinoma , head & neck cancer
R371	4	Malignant melanoma, endometroid carcinoma, Atypical chronic myeloid leukemia
A387	5	endometrioid cancer, breast carcinoma
L394	5	Stomach carcinoma
L482	8	kidney renal clear cell carcinoma
R515	9	endometrioid carcinoma, pancreatic carcinoma