

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Jennifer Gerfen

Date

Telomere position effect influences in human terminal deletions

By

Jennifer Bosse Gerfen
Master of Science

Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

M. Katharine Rudd, Ph.D.
Advisor

Tamara Caspary, Ph.D.
Committee Member

Roger Deal, Ph.D.
Committee Member

Carlos Moreno, Ph.D.
Committee Member

Paula Vertino, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Telomere position effect influences in human terminal deletions

By

Jennifer Bosse Gerfen
A.B., Bryn Mawr College, 2006

Advisor: M. Katharine Rudd, Ph.D.

An abstract of

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of
Emory University in partial fulfillment of the requirements for the degree of
Master of Science in Graduate Division of Biological and Biomedical Science Genetics
and Molecular Biology

2015

Abstract

Telomere position effect influences in human terminal deletions

By Jennifer Bosse Gerfen

Copy number variation (CNV) such as deletions and duplications are a well-known cause of intellectual disability and congenital anomalies. In most cases altered expression of the genes that are gained or lost explain the phenotype. However, alternate mechanisms are necessary to further explain other pathogenic CNV. Telomere position effect (TPE) is the silencing of intact genes adjacent to telomeres as a result of heterochromatin spreading from the telomere, which is known in artificially generated truncated chromosomes from yeast to humans. Here eight cell lines with different terminal deletions were studied to understand how gene expression and chromatin environment is changed as a result of the deletion on naturally occurring chromosomes. In a cell line from a patient with a 5.5-megabase deletion of chromosome 4 there was a ~300-kilobase (kb) region extending from the breakpoint that was enriched for heterochromatic modifications associated with the telomere. Analysis of gene expression within and outside this region shows silencing and allele-skewing of a gene that is ~150-kb from the breakpoint in the region of increased heterochromatin, but normal biallelic expression from a gene ~650-kb from the breakpoint, beyond the region of altered chromatin environment. Analysis by ChiP-chip did not reveal this classical TPE near the breakpoint. This demonstrates that a new telomere can have a cis-effect on nearby genomic regions.

Telomere position effect influences in human terminal deletions

By

Jennifer Bosse Gerfen
A.B., Bryn Mawr College, 2006

Advisor: M. Katharine Rudd, Ph.D.

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of
Emory University in partial fulfillment of the requirements for the degree of
Master of Science in Graduate Division of Biological and Biomedical Science Genetics
and Molecular Biology

2015

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: PROJECT	
1. HYPOTHESIS.....	8
2. MATERIALS AND METHODS	9
3. RESULTS.....	13
FIGURE 1.....	20
FIGURE 2.....	21
FIGURE 3.....	22
FIGURE 4.....	23
FIGURE 5.....	24
FIGURE 6.....	25
TABLE 1.....	26
TABLE 2.....	27
TABLE 3.....	28
TABLE 4.....	29
CHAPTER 3: DISCUSSION.....	30
FIGURE 7.....	37
FIGURE 8.....	38
REFERENCES.....	39

CHAPTER 1: INTRODUCTION

1. Copy Number Variation and human disease

Copy number variation (CNV) results from the gain or loss of genetic material greater than one kilobase. CNVs contribute prominently to normal human polymorphism(1-5). They also have the potential to produce diverse phenotypic consequences such as intellectual disability, autism, and congenital abnormalities (2, 6-9). Usually, CNV-associated phenotypic changes are attributed to altered gene expression within the duplicated or deleted regions of the genome. However, this is only one mechanism that can explain disease presentation (10). Ongoing work is beginning to uncover other ways in which CNV can lead to disease, such as cis-regulation of gene expression near CNV breakpoints.

CNV can potentially influence gene expression via effects on the function of long-range regulatory elements. Indeed, there are increasingly frequent reports of mutations in cis-regulatory elements that can affect phenotype (11). Heterozygous deletions upstream of the *SOX9* gene lead to campomelic dysplasia (12). Similarly, duplication of a limb-specific regulatory element downstream of regulator of *BMP2* has been linked to brachydactyly(13).

Position effect refers to changes in gene expression due to alteration of the chromosomal environment rather than the transcription unit (14). CNV can lead to a change in organization of the genomic structure juxtaposing two regions that are not usually in contact and disrupting normal chromatin environment. In a case report of a reciprocal chromosomal translocation with between the heterchromatic rDNA array on the short arm of chromosome 15 into a euchromatic region on the

long arm of chromosome 16 there were alterations to the chromatin environment and gene expression near the breakpoint on the derivative 16. However, there were no changes to the chromatin environment or gene expression to the region of chromosome 16 on the derivative 15 (15).

In sum, CNV can have long-range cis effects that can alter gene expression through different methods including disruption of regulatory elements or the alteration of the chromatin environment. Clearly understanding the full impact of CNV on gene expression throughout the genome will be required to understand fully the implications of CNV for human genetic disease.

2. Telomere position effect

A particular type of position effect is known as telomere position effect (TPE), wherein intact genes are silenced due to the proximity of a repositioned telomere (16-18). This phenomenon has been reported across many species from yeast to humans using reporter genes adjacent to an artificially truncated chromosome. This has the potential to act on naturally occurring human chromosomal abnormalities and result in a human phenotype.

In most eukaryotes, telomeres are long stretches of short tandem repeats that cap the ends of eukaryotic linear chromosomes. They maintain the structural integrity of chromosomal ends, by preventing DNA damage repair machinery from misidentifying ends as double-strand breaks and circumventing the problem of progressive chromosomal shortening during replication(19, 20). Regulation of telomere length is critical to normal cellular function as abnormally long and short telomeres either allow for cellular over-proliferation or promote senescence (21).

Telomere length is tightly regulated and although the mechanisms are not fully understood, establishment of heterochromatin at the telomeres has been implicated in the process (22).

In humans, telomeres are comprised of 10-15 kb of the six base pair tandem repeat TTAGGG(23). The telomeres are enriched for the histone modifications marked by histone H3 lysine 9 trimethylation (H3K9me3) and histone H4 lysine 20 trimethylation (H4K20me3) (24-26). Directly adjacent to the telomere is the subtelomere, which is comprised of approximately 200-300 kb of segmental duplications. The same heterochromatin modifications associated with telomeres, as well as DNA methylation occur in the subtelomere segmental duplications. However, due to their repetitive nature, it is difficult to precisely map the exact location of modifications within the segmental duplications with respect to how far they spread from the end of the chromosome(27). One possible function of the subtelomere is that there are insulators within these repeats, which protect the rest of the chromosome from the heterochromatin of telomere.

Although telomeres were long thought to be transcriptionally silent, more recent studies have shown that there is expression of telomeric repeat-containing RNA, or TERRA(28). This non-coding RNA is transcribed from just upstream of the telomere. TERRAs negatively regulate telomere length, wherein cell lines with longer telomeres have greater expression of TERRAs and increased H3K9me3(26). The function of TERRA is not fully understood, but it appears to bind to telomeric DNA and recruit factors that facilitate heterochromatin formation on the telomere via a feedback loop.

Propagation of telomeric heterochromatin has been shown to silence nearby genes through the process of TPE(16-18). This has been studied, across many species, using artificially constructed genetic material in which a truncated chromosome contains a telomere positioned directly next to a reporter gene. The mechanism for TPE was first characterized in yeast (17). Initially, the amount of gene silencing from TPE was shown as a linear function of distance from the telomere, with the greatest silencing closest to the telomere, and normal gene expression when positioned 5 kb from the telomere. Subsequent studies determined that establishment of heterochromatin is required for telomeric silencing of genes (29). At native yeast chromosome ends, insulators within the subtelomeric repeat elements prevent the silencing of genes on the chromosome (30, 31).

Similar silencing has been seen in mammalian cell lines with reporter genes. The mechanism for this silencing has been shown to be regulated by both telomere length and heterochromatin. Reduction of histone methylation, or mutations in genes responsible heterochromatin formation and maintenance, restores gene expression (16, 32). In particular studies focused on H3K9me3 methylation and subsequent binding of HP1(33). These studies suggest that telomere length and the specific modifications of the heterochromatin associated with telomeres are contributing factors to TPE. Interestingly, processes for regulation of TPE is species-dependent, as DNA methylation plays a critical role in silencing genes near mouse telomeres, but does not affect TPE in some human systems (18, 32).

At native human chromosomes the focus of understanding TPE has been on the identification of genes that are deregulated as a result of telomere length.

Targeted microarray analysis of more than 1,300 genes near the ends of chromosomes have identified a single change in gene expression from cells with shortened telomeres(34). The silencing of *ISG15*, ~1-Mb from the telomere was not due to spreading of heterochromatin from the telomere since genes closer to the telomere did not exhibit silencing. The mechanism for this silencing has been through recent studies have shown that telomeres can participate in long-range interactions that extend far into the chromosome. 3D mapping of chromosome location has established that the length of the telomere affects the regions of the genome it will bind (35). These changes in chromatin architecture influence gene expression up to 10 Mb from the telomere.

The ability of telomere proximity to influence gene expression is a proposed mechanism for facioscapulohumeral muscular dystrophy (FSHD). This disease that is associated with a reduction in the number of small tandem repeats (3.3 kb) found about 300 kb from the end of chromosome 4q(36, 37). In unaffected individuals there are >100 repeats, however, in patients with FSHD there are <11 repeats. One current hypothesis for the genetic mechanism of FSHD is that the repeat array normally acts as an insulator that prevents toxic overexpression of the gene *DUX4*, which is located in the final repeat(38). As telomeres gradually shorten with age, there is a greater expression of *DUX4*, accounting for both the disease, and its late age of onset.

Thus, the ability of telomeres to silence nearby genes is an important process. There is much currently unknown in the understanding of the role of TPE

in human chromosomes. It is important to further understand the multitude of ways in which the telomere affects local genomic structure and gene expression.

3. Telomere position effect and CNV

Terminal deletions are a specific category of CNV, in which the deletion of one end of a chromosome results in addition of a new telomere to an internal region of the genome. These chromosomal rearrangements have been studied for many years, demonstrating a significant role for their role in human disease. Terminal deletions were shown to account for 1.5% of cases that were referred for clinical cytogenetic testing that did not have an abnormality detected by G-banded chromosomes (39). For some chromosome ends, overlapping deletions have made it possible to identify specific genes associated with a genomic disorder, for example, deletion of *EHMT1* on chromosome 9q and *SHANK3* on chromosome 22(40, 41). However, identification of a single gene deletion responsible for a phenotype is not always possible even when overlapping deletions can be identified and correlated with disease (42, 43)

There is the potential for TPE to play a role in the pathogenicity of terminal deletions since a telomere is added into a new genomic context. Without the subtelomere to act as a boundary for the heterochromatin the telomere has the potential to silence intact genes near the breakpoint. A case study looked for TPE in a human cell line in a patient with the terminal deletion of chromosome 22q, and qualitatively found biallelic expression of and a gene ~50-kb from new telomere through RFLP (44). This does not preclude TPE in human disease since the effect of

the addition of a new telomere could be dependent on the new genomic location or alterations in gene expression might need sensitive, quantitative assays.

Our study aims to better characterize how the repositioned telomeres in terminal deletions affect neighboring chromatin environment and gene expression across the genome. Here we have used a collection of cell lines derived from patients with fine-mapped terminal deletions to determine if there is classical TPE (39, 45, 46). Using these cell lines it is possible to map chromatin modifications and gene expression at known distances from the telomere. We have identified the first reported case of classical TPE with the spreading of heterochromatic marks associated with the telomere (H3K9me3 and H4K20me3) ~300-kb proximal to the breakpoint on a single allele. Within this region of heterochromatin the gene *EVC*, ~150-kb from the breakpoint had a decrease in expression in an allele-specific manner. The deletion of another cell line disrupted a highly organized region of H3K9me3. The seven other cell lines revealed no change in chromatin environment or gene expression. Comparison of changes, or the lack of change, between the deletion cell lines and normal controls allow for the speculation of factors that could influence position effects in other CNV.

CHAPTER 2: PROJECT

1. Hypothesis

This project sets out to better understand the role that CNV play in human disease. I hypothesized that expression of genes near the deletion breakpoint of terminal deletions could be altered. One mechanism that might account for aberrant chromatin modifications or gene expression is telomere position effect (TPE). There is potential for truncated chromosomes to mimic other examples of TPE with the spreading of heterochromatin associated with the telomere to extend into the nearby chromosomal regions, silencing normally expressed genes. We therefore utilized terminal deletion cell lines with known breakpoints to see how a telomere moved to the interior of the chromosome affects gene expression and chromatin environment at known distances to the new telomere.

Eight lymphoblastoid cell lines (LCLs) derived from patients with terminal deletions on six chromosomal ends were studied with ChIP-chip to compare and contrast how a relocated telomere affects different genomic regions (**Figure 1, Table 1**). While it is possible that we would see similar gene silencing and heterochromatin spreading at all chromosome ends, there is a greater chance that each cell line will be unique based on the normal chromatin environment and regulatory elements near the breakpoint. Comparing how different cell lines are influenced by a new telomere would allow for a comparison of nearby genomic features that influence any changes that are observed.

2. Materials and Methods

A. Samples and Tissue Culture

Lymphoblastoid cell lines (LCLs) were obtained from Emory Genetics Lab (EGL) and Ledbetter, Martin, and Bonaglia Labs from patients who had subtelomeric deletions that had previously sequenced breakpoint junctions(45, 46). A control cell line was obtained from Corriell Institute for Medical Research (GM06990). The LCLs were cultured in RPMI Media (Corning, Manassas, VA) supplemented with 15% fetal bovine serum (Atlanta Biologicals, Flowery Branch, GA), and penicillin/streptomycin/L-glutamine (Life Technologies, Grand Island, NY).

B. Chromatin immunoprecipitation

We performed chromatin immunoprecipitation (ChIP) according to the protocol outlined by the Agilent Mammalian ChIP-on-chip v9.2 protocol. Cells were fixed with formaldehyde in a salt solution of 50 mM Hepes-KOH, pH 7.5; 100 mM NaCl; 1 mM EDTA, pH 8.0; and 0.5 mM EGTA, pH 8.0 for 10 minutes and subsequently washed with cold PBS before storage at -80°. Fifty million cells were pooled and lysed with a buffer of 50 mM Hepes-KOH, pH 7.5; 140 mM NaCl; 1 mM EDTA, pH 8.0; 10% glycerol; 0.5% NP-40; and 0.25% Triton X-100. The chromatin was extracted with a solution of 10 mM Tris-HCl, pH 8.0; 200 mM NaCl; 1 mM EDTA, pH 8.0; and 0.5 mM EGTA, pH8.0. Chromatin was then suspended in 10 mM Tris-HCl, pH 8.0; 100 mM NaCl; 1mM EDTA, pH 8.0; 0.5 mM EGTA, pH 8.0; 0.1% Na-deoxycholate; and 0.5% N-lauroylsarcosine. The chromatin was sheared to ~100-bp fragments using a Covaris E4 sonicator with 5% duty, 4 intensity, 200 cycles for 2400 seconds. A

fraction of the lysate was removed at this step as an “input” control. The remainder of the lysate was rocked with antibody bound to Protein G Dynabeads (Life Technologies, Grand Island, NY) that had been blocked with 0.5% BSA in 1X PBS, for 18 hours at 4°. The antibodies used were H3K9me3 (ActivMotif, Carlsbad, CA) and H4K20me3 (Abcam, Cambridge, MA).

(a) CHIP-chip

A custom microarray was designed for each terminal deletion cell line with 180k probes tiled over ~3 Mb proximal and ~500 kb distal to the breakpoint. Arrays were designed with Agilent’s eArray software (<https://earray.chem.agilent.com/earray/>).

Libraries for hybridization to the microarray were created by annealing linkers to the DNA fragments from the enriched chromatin and the input and amplification through two rounds of PCR. The libraries were then labeled with Oligo aCGH/ChIP-on-chip Hybridization Kit (Agilent). The input was labeled with Cy5 and the ChIP fraction was labeled with Cy3. The libraries were then hybridized to each array and scanned with in an ozone-controlled room. The program Feature Extraction (Agilent) was used to convert the signal intensity to a value comparing the ChIP fraction to the input, and Genomic Workbench (Agilent) was used to normalize values of each probe based on median intensity on the array using the Lowess intra-array normalization across multiple arrays with the Whitehead error model (47). The logR values were then extracted and plotted using the UCSC genome browser (<http://genome.ucsc.edu/>).

(b) SNP analysis

Heterozygous SNPs were identified in cell lines using Illumina IlluminaOmniQuad1M genotyping arrays at either the Emory Genomics Core or Hudson Alpha Institute of Biotechnology. Primers were designed to amplify ~100-bp amplicons around each heterozygous SNP. Amplicons were initially derived from samples of input DNA that was sheared in the same manner as ChIP fractions in order to determine our ability to accurately amplify both alleles at equal levels. Amplicons from the same source were pooled, barcoded, and sequenced using an Illumina MiSeq at EGL. This generated 150-bp paired-end reads. The barcode from the end of each read was removed and the samples were treated as single end reads. These reads were aligned to the human genome build GChr37/hg19 using the burrows wheeler alignment (BWA)(48). The frequency of each allele was calculated using a combination of samtools(49) and VarScan(50) to calculate all calls at each SNP of interest that had a read depth greater than 100.

C. RNA-seq

Samples of RNA were isolated from each cell lines (5Prime, Gaithersburg, MD) and sent to Hudson Alpha for RNA-seq. Each sample was sequenced on a full lane of an Illumina HiSeq generating ~200 million reads per sample. Reads from each sample were mapped to the genome using the program Tophat(51). Expression levels were calculated with the program Cufflinks run through the Galaxy server (<https://usegalaxy.org/>) by calculating the fragments per kilobase per million reads (FPKM) (52). Relative amounts of expression were taken by normalizing the FPKM values of each gene to the FPKM of the control gene *GAPDH* in

the same sample. Allele percentages were visualized with Integrative Genome Viewer (IGV).

In order to link multiple SNPs to the same allele, the 3' UTRs of *EVC* and *WFS1* were amplified and cloned using a TOPO vector (Life Technologies, Grand Island, NY). These samples were then sent for sequencing at Beckman Coulter. The 3' UTR from each allele was sequenced to confirm the accuracy of each SNP.

3. Results

A. Heterochromatin enrichment near breakpoint

In a cell line derived from a patient with an ~5.5-Mb deletion of the short arm of chromosome 4 we identified a region adjacent to the new telomere that contained an increase of heterochromatin. DNA from chromatin enriched for H3K9me3 was hybridized to a custom ChIP-chip array with coverage ~500-kb telomeric and 3-Mb centromeric to the breakpoint. ChIP-chip revealed an ~300-kb region of enriched H3K9me3 in the 4p deletion cell line that was not in the normal cell line (**Figure 2**). Beyond this region the deletion cell line appears similar to the normal cell line.

ChIP-chip can detect a net increase of chromatin marks, but is unable to differentiate between the deleted chromosome with the new telomere and the intact chromosome. To show that this increase in H3K9me3 was associated with the new telomere it was necessary to show enrichment in an allele-specific manner. We used the Illumina OmniQuad1M genotyping chip to identify heterozygous SNPs that distinguish the two homologous chromosomes. Primers were designed for ~100-bp amplicons of 36 heterozygous SNPs between 1.6 kb and 2.3 Mb. For each SNP we identified one of three other cell lines with two full-length copies of chromosome 4 that had a corresponding heterozygous SNP.

To prove the validity of the method amplicons from input DNA of the 4p deletion cell line we PCR amplified from input DNA. Of the 36 SNPs that that were interrogated 31 provided reliable data with approximately equal amplification from both alleles <60:40 ratio and 20 of these had a ratio closer to 50:50 with a range of

55:45 (**Table 2**). Two additional SNPs had a greater skewing than 60:40. Read depth ranged from ~100 reads per SNP to in excess of 10,000 reads per SNP. We were unable to link all of the SNPs to each other; however, the two SNPs closest to the breakpoint were amplified from the same chromosome as the new telomere(46).

We amplified DNA from chromatin enriched from both H3K9me3, which we had previously detected a net enrichment near the breakpoint, and from the heterochromatic mark H4K20me3 (**Figure 3b**). Within the 300-kb region that we saw a net increase of H3K9me3 there was skewing of chromatin enriched from one allele over the other. Plotting the allele with the greatest frequency of H3K9me3 enrichment in the deletion cell line across 10 SNPs showed that within this region there was greater chromatin enrichment of one allele over the other with an average ratio of 68:32 but beyond this region there was a similar frequency between both alleles across 22 SNPs with a ratio of 55:45 (**Table 2**). Within the ~300-kb region near the breakpoint chromatin enriched for H4K20me3 was also biased on the same alleles as the enrichment of H3K9me3. In the deletion cell line across an allelic ratio of 73:27 was observed in the DNA enriched for H4K20me3 across 7 SNPs. For SNPs that were >301 kb from the breakpoint a 50:50 ratio was detected across 19 SNPs. The control cell lines had chromatin enriched for H3K9me3 from two full-length copies of chromosome 4. In these cell lines there was an average ratio of 52:48 across five SNPs within ~300 kb of the breakpoint and an average ratio of 53:47 across the 22 of the SNPs beyond this region.

The 4p deletion cell line is associated with a 300-kb region of allele-specific increased H3K9me3 and H4K20me3 next to the breakpoint. This is consistent with the heterochromatin spreading from the telomere that is seen in classical TPE.

B. Gene expression near breakpoint.

To measure the expression levels of genes near the breakpoint, we used RNA-seq, which detects the total amount of gene expression and distinguishes expression from different alleles. Samples were sequenced at the Genomic Services Lab at Hudson Alpha Institute for Biotechnology using an Illumina HiSeq machine. This generated $\sim 200 \times 10^6$ 100-bp paired-end reads per sample that were mapped back to the genome using Tophat (51).

Total gene expression was calculated using Cufflinks to determine the fragments per kilobase per million reads (FPKM)(52). Relative amounts of each gene were normalized to *GAPDH* (**Table 3**). The closest normally expressed gene to the breakpoint was *EVC*, which was ~ 150 -kb from the breakpoint. The cell line with the 4p deletion showed a 49% reduction of *EVC* as compared to the control cell line and a cell line that has a deletion on a different chromosome (**Figure 4A**). Another gene, *WFS1*, is ~ 650 -kb from the breakpoint and has similar expression levels between the deletion and the control cell lines.

We analyzed RNA-seq data to identify heterozygous SNPs in the 3'UTRs of genes near the breakpoint and separate allele-specific gene expression. Visualization of the reads within IGV identified heterozygous SNPs in genes near the breakpoint (**Figure 4B**). The 3' UTR from *EVC* and *WFS1* was cloned in order to

identify SNPs linked to each copy of the gene to use the number of reads to calculate the percentage of transcripts arising from each chromosome.

In the 4p deletion cell line, 80% of the reads that mapped to the 3'UTR of *EVC* were expressed from a single allele, as determined by seven linked SNPs that had approximately 32 reads per SNP (**Figure 4C**). In two control cell lines both copies of *EVC* were expressed at equal levels of reads from each allele with ratios of 51:49 and 55:45 with an average read depth of 120 and 64 reads per SNP. For *WFS1* the alleles showed approximately equal expression of eight SNPs with a read depth of approximately 42 reads per SNP with a ratio of 58:42. Reads from the control cell lines were also show to be present in equal frequency with a ratio of 54:46 and 51:49 in the controls with average read depths of 49 and 24 reads per SNP (**Table 3**).

C. Correlation of chromatin and expression analysis

Integration of the data from the chromatin and expression assays demonstrates that the 300-kb region adjacent to the breakpoint is enriched in heterochromatin and associated with the decreased *EVC* expression. This region encompasses several genes; however, the only gene that is expressed in lymphoblastoid cell lines is *EVC*, which is expressed primarily from one allele in the cell line with the relocated telomere. Beyond this region the chromatin environment is similar between the deletion cell line and the control cell lines, and genes are expressed at similar levels of relative gene expression in a biallelic manner.

For the cell line with the 4p deletion there was a greater amount of heterochromatin on one allele. This same allele had a paucity of reads from RNA,

showing that the decrease of expression was correlated with the increase of heterochromatin (**Figure 5**). The control cell line show approximately equal reads from both alleles from DNA amplified from H3K9me3 enrichment and RNA across *WFS1*, the gene within a similar net chromatin environment between the deletion and the control cell line. Both cell lines showed similar levels of heterochromatin enrichment and gene expression from both *WFS1* alleles. These allele-specific changes unique to the deletion cell would allow us to believe that the change suggest that they are present on the chromosome with the deletion, which is consistent with the new telomere exerting a cis-effect on the nearby genomic region.

D. Differences on other chromosome ends

Analysis of the chromatin environment and gene expression near other terminal deletion breakpoints allows comparison of the effects of a relocated telomere on different genomic environments. In total, I studied eight cell lines with terminal deletions on six different chromosome ends (**Figure 1, Table 1**). As determined by ChIP-chip, the only cell line with obvious heterochromatin spreading from the telomere was the cell line with the deletion of 4p. The analysis of these chromosome ends can help to identify factors that are present in the 4p deletion cell line that allow for TPE.

In one case, existing chromatin organization was affected. Proximal to the breakpoint of a 4.5-Mb deletion on chromosome 21 was a large organized region of H3K9me3 in normal cell lines. When we analyzed ChIP-chip data from a wild type cell line there was a highly organized 1-Mb region of H3K9me3. Analysis of H3K9me3 enrichment in this deletion showed disorganized levels H3K9me3 near

the breakpoint, including disruption of an organized region of the H3K9me3-associated region (**Figure 6A**). Disruption of the organization in this region could result in loss of heterochromatin across genes in the region and allow for gene expression. However, *DSCAM*, the gene that is located within this heterochromatic region, was not expressed, nor was there any allele-skewing in the three genes with heterozygous SNPs out of six expressed genes near the breakpoint. This included the gene *C2CD2*, which is the closest expressed gene to the breakpoint which showed similar relative expression levels by FPKM in the deletion as compared to the control to the control, 0.56 and 0.60 respectively. Allele-specific analysis demonstrated minor skewing with a ratio of 61:39 across 6 SNPs; however, one of the two control lines had the same ratio, with 61:39 and other had a less skewed ratio of 55:45.

In six cases, ChIP-chip data did not reveal any difference between the chromatin environment between the control cell lines and the terminal deletion cell lines. These included a 1.3-Mb deletion of the long arm of chromosome 9, two deletions of the long arm of chromosome 12 (1.9-Mb and 4.8-Mb), a 2.0-Mb deletion of the short arm of chromosome 16, a 300-kb deletion of chromosome 21, and a 130-kb deletion of chromosome 22. For one of these cell lines, the 4.8 Mb deletion of chromosome 12, the breakpoint was located within a region of enrichment for H3K9me3. The boundaries of this region remained the same. In the 130-kb deletion of 22q deletion, the breakpoint was located at a breakage hotspot, with the same deletion as the previous report looking for TPE. Our finding of a lack of change in H3K9me3 enrichment complements the previous report of biallelic expression of

ARSA, a gene ~55-kb from the breakpoint, by RFLP analysis (44). Since analysis of two independent deletions revealed a lack of change in different factors involved in position effect it can be surmised that position effect does not play a role in patients with this deletion.

In two of these cases we had RNA-seq data that did not identify changes in gene expression levels or allele-skewing. One example was a 1.9-Mb deletion on the long arm of chromosome 12 that was located in a gene poor region. This breakpoint of this deletion was located near two regions of organized heterochromatin with of H3K9me3 enrichment, but these were not altered in this deletion cell line (**Figure 6B**). In another case there was a 1.3-Mb deletion on the long arm of chromosome 9 that was located in a gene-rich region with the nearest gene ~34-kb from the breakpoint. Neither the H3K9me3 localization nor the gene expression were altered in the deletion cell line. (**Figure 6C**).

The changes associated with a new telomere vary between different deletions. Based on our ChIP-chip experiments most terminal deletions do not show signs of classical TPE. There are also different manners in which deletions can affect nearby chromatin environment.

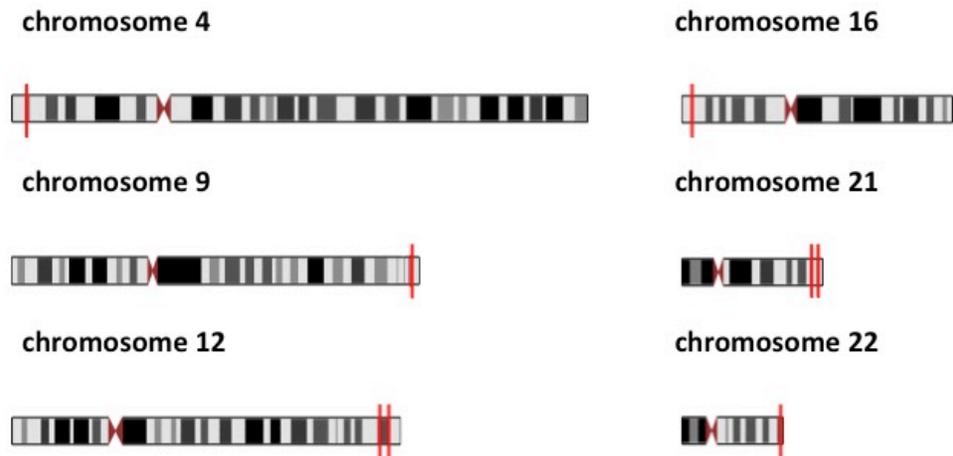


Figure 1: Terminal deletions in study

A schematic showing the locations of breakpoints of the patients involved in this study in relationship to each other. Each red line indicates the location of a breakpoint where a new telomere was added. Each patient has a single truncated chromosome.

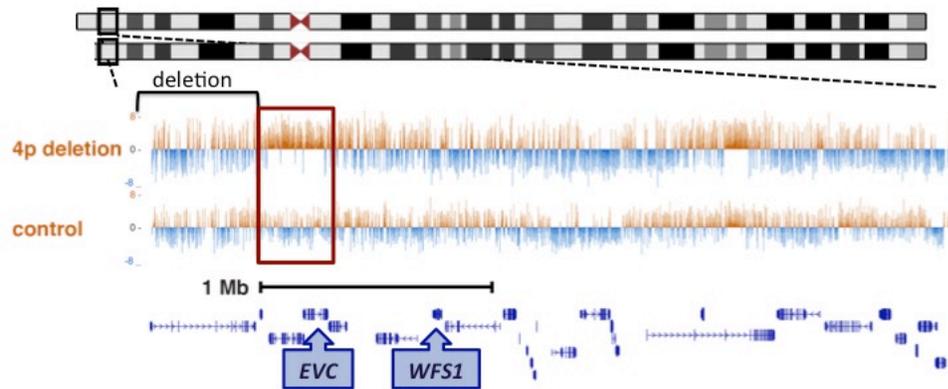


Figure 2: An increase of H3K9me3 enrichment next to the breakpoint in 4p deletion cell line

Data from ChIP-chip for enrichment of H3K9me3 is plotted on the UCSC browser for the deletion cell line and the control cell line. The bracket represents the deleted region. The orange and blue lines represent the data of H3K9me3 enrichment. Positive values (orange) indicate a chromatin enriched for H3K9me3 over the input. Negative values (blue) are regions where there is little or no H3K9me3 bound to the chromatin. The region of enrichment is indicated by the red box. RefSeq genes are indicated in blue, with the genes *EVC* and *WFS1* highlighted.

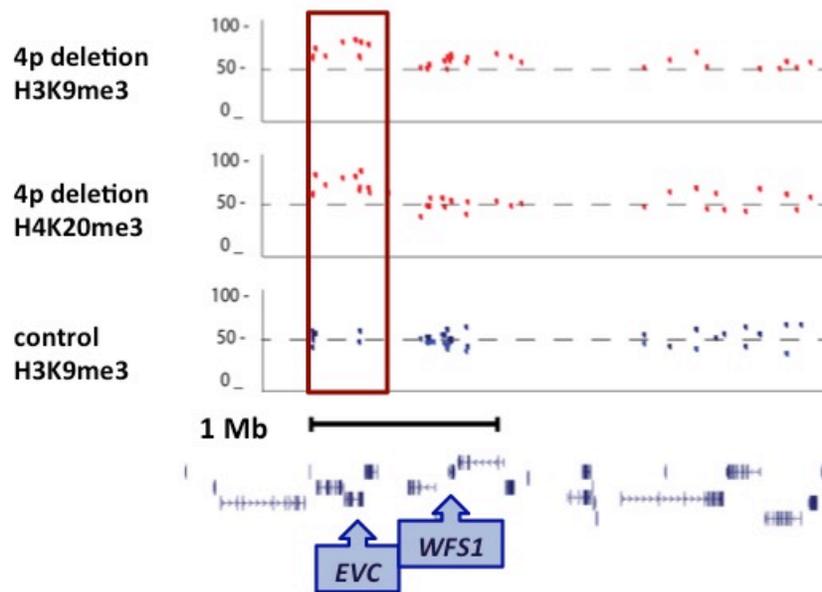


Figure 3: Increased enrichment for H3K9me3 is allele specific.

This shows the plot of the frequency of a single allele in enriched chromatin in the UCSC genome browser. Each point represents a different SNP. The allele plotted for each SNP is the allele with the greatest frequency in the chromatin enriched for H3K9me3 in the 4p deletion. Allelic frequencies shown from chromatin enriched for H3K9me3 and H4K20me3 in the 4p deletion cell line. The data from the control comes from two different cell lines that have two full-length copies of chromosome 4.

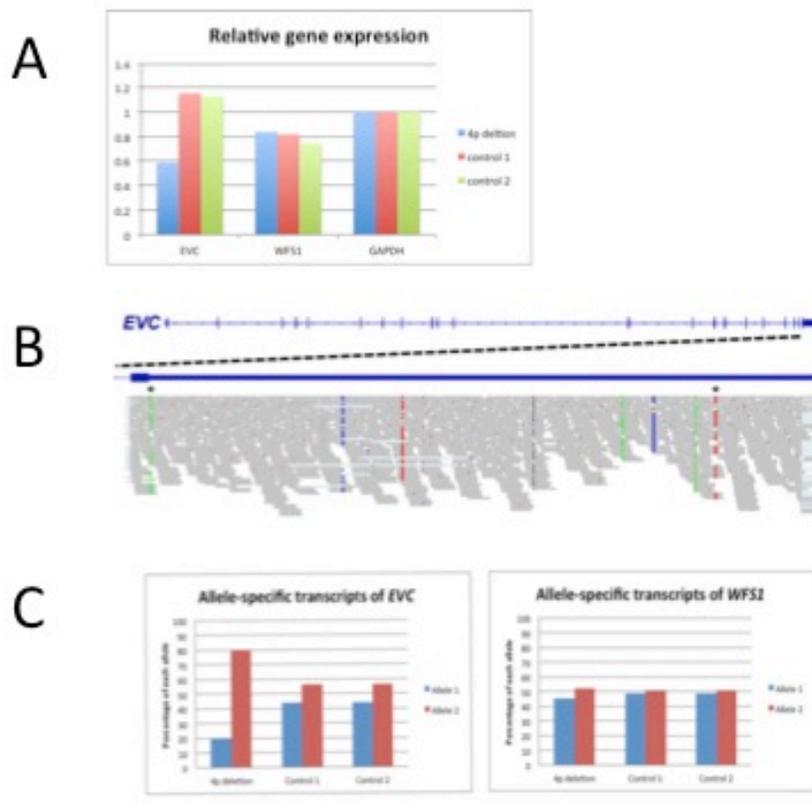


Figure 4: *EVC* expression is decreased in an allele specific manner.

A. This plot shows the relative expression of *EVC* and *WFS1* calculated by FPKM. The values for each gene are shown as normalized to *GAPDH*. The 4p deletion is shown as compared to two control cell lines. **B.** This is a representation of the RNA-seq reads of the 4p deletion in IGV at the 3' UTR of *EVC*. Each asterisk represents one of the heterozygous SNPs analyzed in both RNA-seq and chromatin allelic frequency assays. **C.** Plots showing the frequencies of both alleles of *EVC* and *WFS1* in the 4p deletion and two control cell lines.

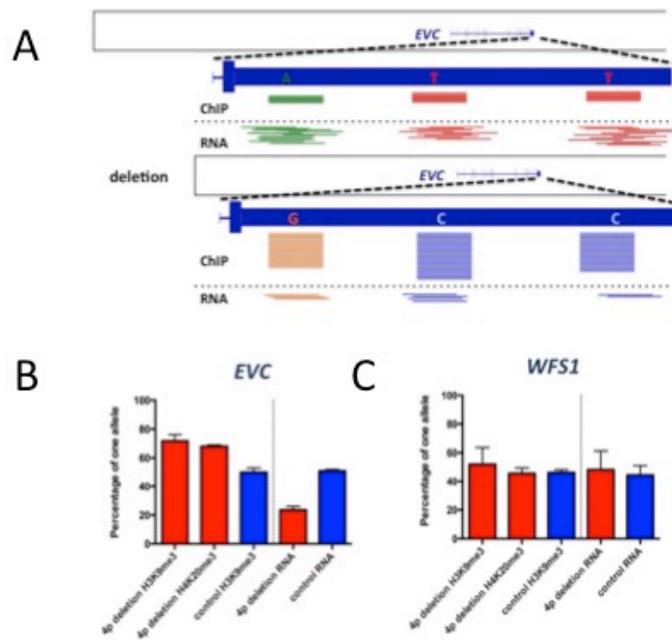


Figure 5: The allele-specific increase of heterochromatin and decrease of gene expression are on the same chromosome.

A. A graphic representation of how the reads mapped to the 3' UTR of *EVC* in the 4p deletion cell line. **B.** Plotting the frequency of one of the alleles for *EVC* in the 4p deletion and control cell line for chromatin enrichment and RNA levels. **C.** Allele-frequency plots for *WFS1*.

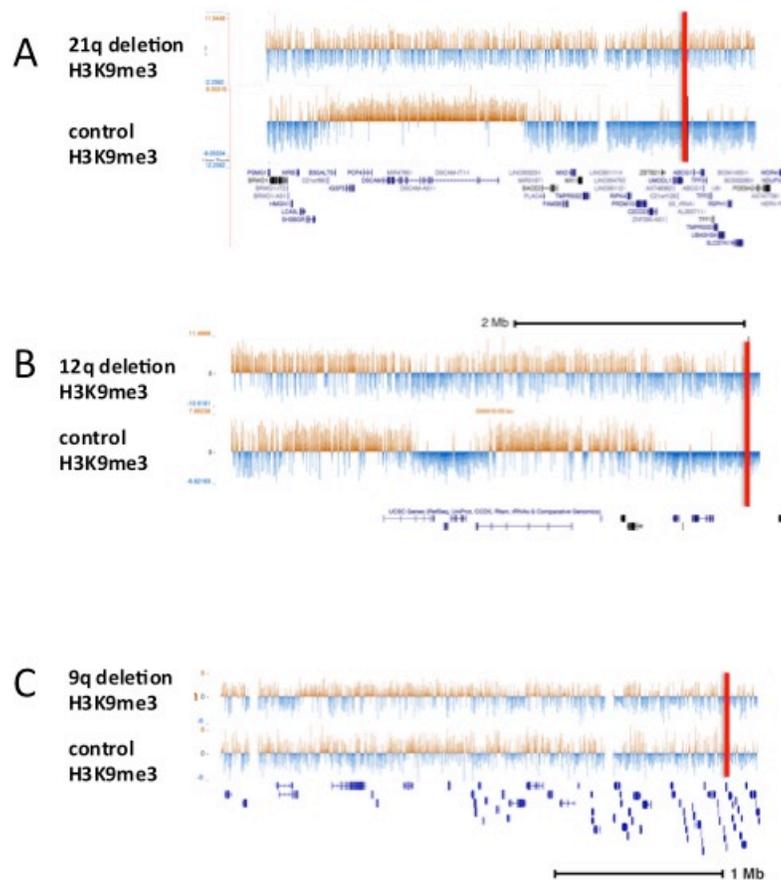


Figure 6: Other cell lines do not have atypical or no TPE near their breakpoints.

A. ChIP-chip data for H3K9me3 chromatin enrichment in a 4.5-Mb 21q deletion cell line and a control cell line plotted in the UCSC genome browser. Enrichment is shown by positive values plotted in orange, whereas negative values in blue show little to no enrichment. The red line indicates the breakpoint, with the region to the right deleted on one chromosome. **B.** ChIP-chip data near the breakpoint of a 1.9-Mb deletion of 12q deletion in the patient and the control cell line plotted in the UCSC browser. **C.** ChIP-chip data from a 1.3-Mb 9q deletion cell line in the patient and the control cell line.

Table 1: Breakpoints of deletion cell lines included in study.

Patient	Breakpoint	Deletion length	Patient ID (45, 46)
4p deletion	chr4:5559710	5.5 Mb	LM208
9q deletion	chr9:139872801	1.3 Mb	LM206
12q deletion 1	chr12:131929173-131929185	4.8 Mb	LM213
12q deletion 2	chr12:129007844	1.9 Mb	LM203
16p deletion	chr16:1934888	1.2 Mb	LM204
21q deletion 1	chr21:43592803	4.5 Mb	EGL205
21q deletion 2	chr21:46645614	130 kb	LM216
22q deletion	chr22:51122204	130 kb	P31

TABLE 2: Allelic frequency of SNPs enriched for heterochromatin marks near the 4p deletion breakpoint. Starred SNPs are located within the 3' UTR of either *EVC* or *WFS1*, and correspond with data from the RNA-seq experiments

SNP	Distance from breakpoint (bp)	4p Deletion H3K9me3	4p Deletion H4K20me3	Control H3K9me3	Input
rs13113867	3,486	66.0%		41.7%	53.0%
rs7670299	16,430	71.0%		55.7%	50.8%
rs7684204	58,597	83.0%	83.3%		53.4%
rs1078758	69,588	63.0%	69.0%	42.0%	48.1%
rs11727214	151,658	60.5%			49.3%
rs6811132	231,496	79.0%	79.0%		55.9%
rs2291151*	253,068	59.0%	64.0%	47.0%	
rs3733186*	255,671	53.0%	65.0%	53.0%	
rs10937681	261,051	83.0%	75.0%		47.2%
rs7356507	301,196	67.0%	74.0%		
rs6838020	307,109	62.0%		48.0%	52.3%
rs6839295	407,091	61.4%		45.2%	52.4%
rs4512066	581,027	51.2%	36.7%	36.9%	43.4%
rs4688970	615,688	52.3%	49.7%	45.3%	50.6%
rs885420	626,106	54.5%	47.8%	51.3%	45.6%
rs1877762	631,341		55.5%	38.6%	48.1%
rs4688982	707,291	59.0%		53.9%	55.3%
rs4689391	720,739	50.0%		40.2%	51.4%
rs10001190	724,923	62.0%		61.4%	59.8%
rs1801206*	742,997	29.0%	44.0%	46.5%	54.0%
rs1046314*	744,245	61.0%	53.0%	46.5%	49.0%
rs1046320*	744,634	63.0%	40.0%	46.0%	53.0%
rs4688999	824,082	58.0%	40.0%	63.0%	62.2%
rs2240268	832,632	62.0%	51.0%	42.4%	38.1%
rs9654056	986,723	65.0%	63.0%	53.0%	55.0%
rs3088248	1,063,622	61.5%	47.8%		49.0%
rs4689510	1,120,867	56.9%	50.7%	53.4%	34.5%
rs34762772	1,776,512	49.0%	52.4%	46.0%	46.5%
rs4495038	1,913,495	59.0%	62.0%	43.0%	40.7%
rs16840670	2,055,667	33.0%	33.0%	41.0%	55.4%
rs11724048	2,113,075	52.0%	45.0%	38.0%	49.0%
rs4689838	2,161,124	58.0%	61.0%	52.0%	65.2%
rs2240055	2,205,333		55.0%	45.0%	45.1%
rs28622214	2,319,730	71.0%	42.0%	43.5%	38.5%
rs11732351	2,395,778	50.0%	65.0%	55.0%	56.1%
rs1281142	2,666,778	42.9%	43.2%		48.1%

Table 3: FPKM values for 4p deletion

	4p deletion	control 1	control 2
<i>EVC</i>	48,544.2	114,848.0	82,484.2
<i>WFS1</i>	69,162.6	81,225.9	54,305.6
<i>GAPDH</i>	82,544.9	99,361.3	73,073.7

Table 4: Allelic frequencies of SNPs from RNA-seq data. The total number of reads across each SNP and the frequency of one of the alleles are shown. Data is only shown for heterozygous SNPs. The asterisks mark SNPs that were interrogated in analysis of chromatin and expression.

		4p Deletion		Control 1		Control 2	
		Total Reads	Allele 1	Total Reads	Allele 1	Total Reads	Allele 1
<i>EVC</i>	rs2291151*	37	30%	119	53%	75	76%
	rs543591683			107	49%	69	38%
	rs6840435	24	25%	125	52%	69	55%
	rs12501119	33	18%				
	rs67547790	39	18%				
	rs4261923	25	24%				
	rs6848370			115	41%	54	54%
	rs3733186*	38	21%	121	50%	72	69%
	rs3733189			138	47%		
	rs3733187			121	30%	62	65%
	rs3733188			113	30%	52	42%
	rs3733189					55	44%
<i>WFS1</i>	rs1801214	40	48%			24	33%
	rs148028521			62	53%		
	rs1805069			42	43%		
	rs2230721			36	44%		
	rs734312			49	43%		
	rs1801206*			48	40%		
	rs1046314*	26	69%	47	49%		
	rs1046316	44	36%				
	rs565619703	37	59%			17	47%
	rs1046319	44	36%			23	57%
	rs1046319	38	39%			28	57%
	rs1046322			74	57%		
	rs1046320*	41	51%			28	57%
	rs573433508						
	rs9457	42	38%	51	37%		
rs3200			32	44%			

CHAPTER 3: DISCUSSION

1. Telomere position effect seen in 4p deletion

This study demonstrates the first example of classical TPE on a human chromosome end, as seen by heterochromatin spreading from the repositioned telomere and gene silencing within the region of the heterochromatin spreading. In the cell line with a 4p deletion, allele-specific sequencing of chromatin revealed heterochromatin marks that extend 300-kb from the telomere on the chromosome with the deletion. Analysis of expression of the gene *EVC*, within the region of the heterochromatin showed decreased expression. Another, *WFS1*, that is outside this region, showed no change in expression compared with expression in control cell lines. These results suggest that heterochromatin spreading from the telomere is responsible for decreased expression gene expression. This is shown as a model in Figure 7.

Ellis-van Crevald syndrome is an autosomal recessive syndrome caused by mutations in either *EVC* or *EVC2*(53, 54). This syndrome presents with skeletal dysplasia, polydactyly, and cardiac malformations, most frequently an atrial septal defect (55, 56). In the literature, a father and daughter with an atrial septal defect and polydactyly have been reported to each carry a heterozygous missense mutation in *EVC* (53). In the 4p deletion cell line we showed a decrease in *EVC* expression. The phenotype of the patient in the study included intellectual disability and an atrial septal defect. While this patient was not diagnosed with Ellis-van Crevald syndrome, it is possible that the decrease in *EVC* expression led to his heart defect.

While TPE was evident in the 4p deletion cell line, it remains unclear why this was the only cell line out of the eight to show spreading of heterochromatin by ChIP-chip. The 21q cell line showed the possibility of disruption of normal chromatin organization near the breakpoint. There is large region of organized H3K9me3 located near the breakpoint in normal cell lines located ~1.1 Mb from the breakpoint.

In the 21q deletion cell line the H3K9me3 enrichment is more diffuse and lacks the organized boundaries present in normal chromosomes 21 (see Figure 6 of Chapter 2). In order to fully understand how the chromatin is affected in these cell lines it would be necessary to more precisely analyze the differences in chromatin in an allele-specific manner as these ChIP-chip assays do not define a clear pattern of how the chromatin is organized in the deletion chromosome.

Interrogation of this region in an allele-specific manner might allow for the distinction of how the new telomere affects the local chromatin environment. However, due to disruption of the H3K9me3 region, it is possible that the deletion chromosome has a decrease of heterochromatin in some regions and an increase in heterochromatin in regions adjacent to it. To properly understand the boundary of the disrupted region would require exact determination of which chromosome each allele was on, rather than just understanding that there was an allelic imbalance. In the few cases in which parents are available it is possible to determine which allele of heterozygous SNPs were inherited from each parent when at least one parent is homozygous for the allele(57). This would allow for the each allele to be assigned to the chromosome with or without the deletion present. If parents are not available

there are methods to sequence each individual chromosome from an individual using a fosmid library to determine the haplotype, but this is not a feasible approach(58).

2. A new method for assessing TPE in human cells

The clustered, regularly interspaced, short palindromic repeat (CRISPR) technology provides a new method of genomic editing involving RNA guided nucleases(59). Use of this technique would allow for the creation of cell lines with terminal deletions from an existing control cell line. This would allow for the overcoming some of the obstacles that are associated with the nature of the current study including determination that changes seen in the chromatin environment and gene expression are a product of the deletion and not the effects of human variation. However, it would not affect our inability to determine whether changes are a direct result of the new telomere. There remains the possibility that the deletion of elements that regulate gene expression or chromatin environment could result in changes on the other side of the breakpoint. It would also be possible to choose a cell line with known parents, which would allow for the determination of which allele was inherited from each parent. That it would be possible to perform allele-specific assays linking each allele to the deleted or non-deleted chromosome. While it was not necessary to determine which chromosome the alleles in the 4p deletion were associated with this was a large increase in heterochromatin across multiple SNPs suggested that they were on the same allele. Linking SNPs to the same chromosome would allow for the understanding of effects beyond the simple spreading of heterochromatin as seen in the 4p deletion. It would also allow for the

detection of smaller effects that might not be seen under the assumption that the alleles with the higher allele frequency were on the same chromosome.

Further, the ability to control the terminal deletions would allow for the creation of deletions at specific genomic locations rather than limiting the study to patients that are willing to enroll in the study. This would enable the correlation of effects with the distance of a repositioned telomere. In the case of the 4p deletion it would be possible to interrogate whether a larger or a smaller deletion would have similar effects. This would allow for determination of the role of the genomic elements near the deletion in comparison with the role of distance from the telomere.

3. Telomere length as a factor

One factor that could play a role in why we did not see TPE in some of our cell lines might be due to telomere length, which has not been measured here. Previous reports have shown the strength of silencing in TPE was proportional to the length of the telomere. HELA cells whose telomeres were lengthened to 14 kb from 5 kb showed an increase of silencing associated with the telomere(16). Short telomeres could account for the fact that we do not observe TPE in some of our cell lines.

Although LCLs are considered “immortalized” they have a finite division rate since there is no mechanism for telomere lengthening as opposed to other true immortal cell lines (60). Data from analyses studying the rapid shortening in patients with a mutation in a component of telomerase demonstrated that telomeres in control LCLs were not shortened as compared to other human subjects

(61, 62). Further after four weeks of passaging there was not a significant decrease in telomere length in these cells.

If the cell lines that we studied behave as other LCLs we should not expect them to have abnormally shortened telomeres. While telomere length is an important factor in TPE it is unlikely the primary cause of why we see differences between our cell lines.

4. Genomic targeting to the nuclear lamina as a possible explanation of TPE in 4p deletion cell line

A study analyzed the effects of different insulators in relieving the effects of TPE by inserting different DNA elements between two reporter genes. These data showed the greatest increase of gene expression and reversal of silencing of the proximal gene were matrix attachment regions(63). This makes sense because telomeres are localized to the nuclear periphery, which is associated with silent genes, and regions of the genome near the new telomere would be expected to be displaced from their normal nuclear location (64, 65). Regions of the genome that are associated with the nuclear lamina show localization of large regions H3K9me3(66). However, genes that are adjacent to the nuclear lamina have active expression due to the insulating regions at these junctions.

It would follow that when in a terminal deletion the effect of the newly positioned telomere would be dependent on nearby genomic features and the normal chromatin state. According to studies in lung fibroblast cells, the 4p deletion breakpoint is located within a region associated with the nuclear lamina (**Figure 8**) (67). Since this region has a propensity to interact with the nuclear lamina it is

possible that the addition of a telomere near this region provides the stabilization for attachment, which would otherwise remain unbound.

Similarly, in the 21q deletion cell line the region of organized H3K9me3 enrichments has been identified as a LAD. Since it is known that telomeres are able to interact with genomic regions that are several megabases away it is likely that this disorganization could be due to changes to a cis-effect from the new telomere (35). Further examination of the changes exerted by the new telomere in that deletion would allow for a better understanding of how or whether telomeres have the ability to affect nearby LADs.

Since the two deletions with altered chromatin structure are located at or near LADs, TPE may only occur in terminal deletions when the breakpoint is located near a region that already has the ability to interact with the nuclear lamina. Whereas, in terminal deletions that are located in a large region of active chromatin the telomere has little effect on the state of the chromatin.

5. Conclusion

In conclusion, a new telomere at the end of a terminal deletion can exert changes in chromatin environment and gene expression. Although TPE is not evident in all cases, one must consider positional effects of intact genes outside of the deletion playing a role in phenotype. Further study is necessary to fully understand the extent of TPE in terminal deletions, and uncover the factors that elicit heterochromatin spreading in this type of CNV. Further, this study shows that deletions have the ability to have cis-effects that are near, but not at, the breakpoint.

In addition, position effects are likely to affect other CNVs that do not involve the telomere.

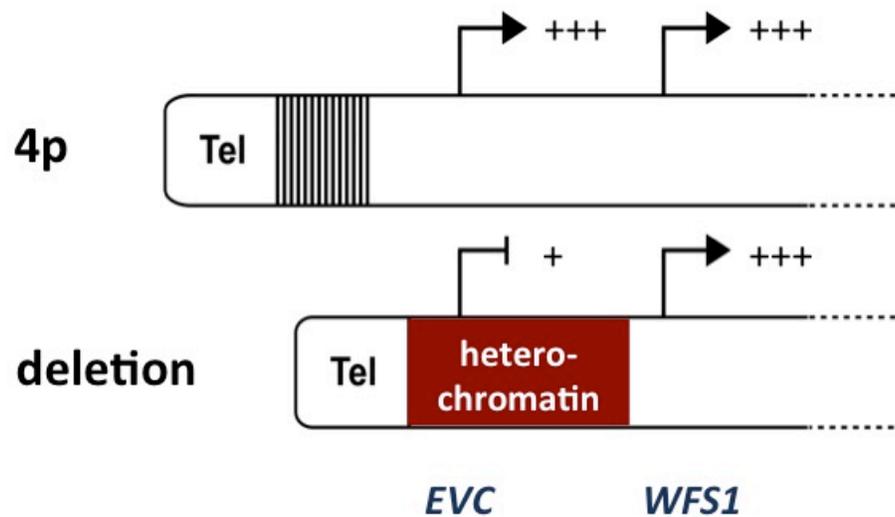


Figure 7: Model of the chromosomes in the 4p deletion cell line.

In this representation of the data in the 4p deletion cell line there are allele-specific heterochromatin marks and decrease in *EVC* expression. These are located to the chromosome with the deletion. The genes *EVC* and *WFS1*, and the plus signs represent the relative amount of gene expression. The presence of heterochromatin is seen as the red box.

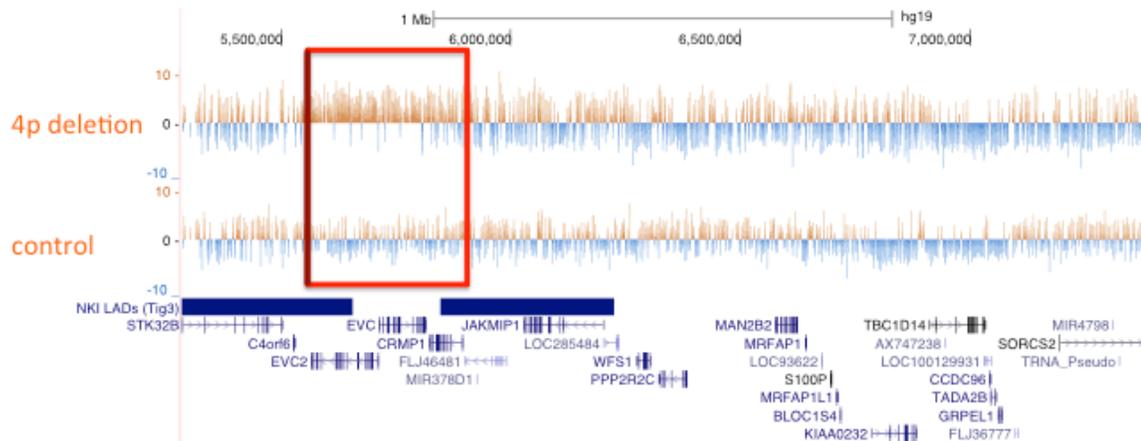


Figure 8: The 4p deletion is located near laminin associated domains (LAD).

A plot from the UCSC genome browser of 2 Mb around the 4p deletion breakpoint showing the ChIP-chip data as depicted before. The red box surrounds the region of heterochromatin enriched in the deletion cell line, with the breakpoint at the left side of the box in dark red. The orange regions representing enrichment of H3K9me3. The orange regions represent increased H3K9me3 enrichment and the blue regions represent little to no enrichment. The UCSC genes are shown. The NKI LADs track shows previously published data of regions of the genome that are bound to LaminB(67).

References

1. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet.* 2006;38(1):75-81. doi: 10.1038/ng1697. PubMed PMID: 16327808.
2. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI, Mefford H, Ying P, Nickerson DA, Eichler EE. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet.* 2009;84(2):148-61. doi: 10.1016/j.ajhg.2008.12.014. PubMed PMID: 19166990; PMCID: 2668011.
3. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM, International HapMap C. Common deletion polymorphisms in the human genome. *Nat Genet.* 2006;38(1):86-92. PubMed PMID: 16468122.
4. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome. *Nature.* 2006;444(7118):444-54. doi: 10.1038/nature05329. PubMed PMID: 17122850; PMCID: 2669898.
5. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. *Science.* 2004;305(5683):525-8. doi: 10.1126/science.1098918. PubMed PMID: 15273396.
6. Lu XY, Phung MT, Shaw CA, Pham K, Neil SE, Patel A, Sahoo T, Bacino CA, Stankiewicz P, Kang SH, Lalani S, Chinault AC, Lupski JR, Cheung SW, Beaudet AL. Genomic imbalances in neonates with birth defects: high detection rates by using chromosomal microarray analysis. *Pediatrics.* 2008;122(6):1310-8. doi: 10.1542/peds.2008-0297. PubMed PMID: 19047251; PMCID: 2795566.
7. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST, Richard G, Compton JG, Fuller AE, Gliem TJ, Huang S, Collinson MN, Beal SJ, Ackley T, Pickering DL, Golden DM, Aston E, Whitby H, Shetty S, Rossi MR, Rudd MK, South ST, Brothman AR, Sanger WG, Iyer RK, Crolla JA, Thorland EC, Aradhya S, Ledbetter DH, Martin CL. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med.* 2011;13(9):777-84. doi: 10.1097/GIM.0b013e31822c79f9. PubMed PMID: 21844811; PMCID: 3661946.
8. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. A copy number variation morbidity

- map of developmental delay. *Nat Genet.* 2011;43(9):838-46. doi: 10.1038/ng.909. PubMed PMID: 21841781; PMCID: 3171215.
9. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LE, Schuurs-Hoeijmakers JH, Hoischen A, Pfundt R, Krumm N, Carvill GL, Li D, Amaral D, Brown N, Lockhart PJ, Scheffer IE, Alberti A, Shaw M, Pettinato R, Tervo R, de Leeuw N, Reijnders MR, Torchia BS, Peeters H, O'Roak BJ, Fichera M, Hehir-Kwa JY, Shendure J, Mefford HC, Haan E, Gecz J, de Vries BB, Romano C, Eichler EE. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet.* 2014;46(10):1063-71. doi: 10.1038/ng.3092. PubMed PMID: 25217958; PMCID: 4177294.
 10. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science.* 2007;315(5813):848-53. Epub 2007/02/10. doi: 10.1126/science.1136678. PubMed PMID: 17289997; PMCID: 2665772.
 11. Epstein DJ. Cis-regulatory mutations in human disease. *Brief Funct Genomic Proteomic.* 2009;8(4):310-6. doi: 10.1093/bfpg/elp021. PubMed PMID: 19641089; PMCID: 2742803.
 12. Sanchez-Castro M, Gordon CT, Petit F, Nord AS, Callier P, Andrieux J, Guerin P, Pichon O, David A, Abadie V, Bonnet D, Visel A, Pennacchio LA, Amiel J, Lyonnet S, Le Caignec C. Congenital heart defects in patients with deletions upstream of SOX9. *Hum Mutat.* 2013;34(12):1628-31. doi: 10.1002/humu.22449. PubMed PMID: 24115316.
 13. Dathe K, Kjaer KW, Brehm A, Meinecke P, Nurnberg P, Neto JC, Brunoni D, Tommerup N, Ott CE, Klopocki E, Seemann P, Mundlos S. Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *Am J Hum Genet.* 2009;84(4):483-92. doi: 10.1016/j.ajhg.2009.03.001. PubMed PMID: 19327734; PMCID: 2667973.
 14. Girton JR, Johansen KM. Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. *Advances in genetics.* 2008;61:1-43. Epub 2008/02/20. doi: 10.1016/S0065-2660(07)00001-6. PubMed PMID: 18282501.
 15. Finelli P, Sirchia SM, Masciadri M, Crippa M, Recalcati MP, Rusconi D, Giardino D, Monti L, Cogliati F, Faravelli F, Natacci F, Zoccante L, Bernardina BD, Russo S, Larizza L. Juxtaposition of heterochromatic and euchromatic regions by chromosomal translocation mediates a heterochromatic long-range position effect associated with a severe neurological phenotype. *Mol Cytogenet.* 2012;5:16. doi: 10.1186/1755-8166-5-16. PubMed PMID: 22475481; PMCID: 3395859.
 16. Baur JA, Zou Y, Shay JW, Wright WE. Telomere position effect in human cells. *Science.* 2001;292(5524):2075-7. doi: 10.1126/science.1062329. PubMed PMID: 11408657.
 17. Gottschling DE, Aparicio OM, Billington BL, Zakian VA. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell.* 1990;63(4):751-62. PubMed PMID: 2225075.

18. Pedram M, Sprung CN, Gao Q, Lo AW, Reynolds GE, Murnane JP. Telomere position effect and silencing of transgenes near telomeres in the mouse. *Mol Cell Biol.* 2006;26(5):1865-78. doi: 10.1128/MCB.26.5.1865-1878.2006. PubMed PMID: 16479005; PMCID: 1430234.
19. O'Sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol.* 2010;11(3):171-81. doi: 10.1038/nrm2848. PubMed PMID: 20125188; PMCID: 2842081.
20. Blackburn EH, Greider CW, Szostak JW. Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. *Nat Med.* 2006;12(10):1133-8. doi: 10.1038/nm1006-1133. PubMed PMID: 17024208.
21. Blasco MA. The epigenetic regulation of mammalian telomeres. *Nat Rev Genet.* 2007;8(4):299-309. doi: 10.1038/nrg2047. PubMed PMID: 17363977.
22. Benetti R, Garcia-Cao M, Blasco MA. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat Genet.* 2007;39(2):243-50. doi: 10.1038/ng1952. PubMed PMID: 17237781.
23. de Lange T, Shiue L, Myers RM, Cox DR, Naylor SL, Killery AM, Varmus HE. Structure and variability of human chromosome ends. *Mol Cell Biol.* 1990;10(2):518-27. PubMed PMID: 2300052; PMCID: 360828.
24. Gonzalo S, Garcia-Cao M, Fraga MF, Schotta G, Peters AH, Cotter SE, Eguia R, Dean DC, Esteller M, Jenuwein T, Blasco MA. Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. *Nat Cell Biol.* 2005;7(4):420-8. doi: 10.1038/ncb1235. PubMed PMID: 15750587.
25. Garcia-Cao M, O'Sullivan R, Peters AH, Jenuwein T, Blasco MA. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat Genet.* 2004;36(1):94-9. doi: 10.1038/ng1278. PubMed PMID: 14702045.
26. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell.* 2009;35(4):403-13. doi: 10.1016/j.molcel.2009.06.025. PubMed PMID: 19716786; PMCID: 2749977.
27. Deng Z, Wang Z, Stong N, Plasschaert R, Moczan A, Chen HS, Hu S, Wikramasinghe P, Davuluri RV, Bartolomei MS, Riethman H, Lieberman PM. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J.* 2012;31(21):4165-78. doi: 10.1038/emboj.2012.266. PubMed PMID: 23010778; PMCID: 3492729.
28. Azzalin CM, Reichenbach P, Khoriatuli L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science.* 2007;318(5851):798-801. doi: 10.1126/science.1147182. PubMed PMID: 17916692.
29. Aparicio OM, Billington BL, Gottschling DE. Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell.* 1991;66(6):1279-87. Epub 1991/09/20. PubMed PMID: 1913809.
30. Pryde FE, Louis EJ. Limitations of silencing at native yeast telomeres. *The EMBO journal.* 1999;18(9):2538-50. Epub 1999/05/06. doi: 10.1093/emboj/18.9.2538. PubMed PMID: 10228167; PMCID: 1171335.

31. Fourel G, Revardel E, Koering CE, Gilson E. Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *The EMBO journal*. 1999;18(9):2522-37. Epub 1999/05/06. doi: 10.1093/emboj/18.9.2522. PubMed PMID: 10228166; PMCID: 1171334.
32. Koering CE, Pollice A, Zibella MP, Bauwens S, Puisieux A, Brunori M, Brun C, Martins L, Sabatier L, Pulitzer JF, Gilson E. Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. *EMBO Rep*. 2002;3(11):1055-61. doi: 10.1093/embo-reports/kvf215. PubMed PMID: 12393752; PMCID: 1307600.
33. Tennen RI, Bua DJ, Wright WE, Chua KF. SIRT6 is required for maintenance of telomere position effect in human cells. *Nat Commun*. 2011;2:433. doi: 10.1038/ncomms1443. PubMed PMID: 21847107; PMCID: 3528101.
34. Lou Z, Wei J, Riethman H, Baur JA, Voglauer R, Shay JW, Wright WE. Telomere length regulates ISG15 expression in human cells. *Aging (Albany NY)*. 2009;1(7):608-21. PubMed PMID: 20157543; PMCID: 2806043.
35. Robin JD, Ludlow AT, Batten K, Magdinier F, Stadler G, Wagner KR, Shay JW, Wright WE. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev*. 2014;28(22):2464-76. doi: 10.1101/gad.251041.114. PubMed PMID: 25403178; PMCID: 4233240.
36. van Deutekom JC, Wijmenga C, van Tienhoven EA, Gruter AM, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet*. 1993;2(12):2037-42. PubMed PMID: 8111371.
37. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet*. 1992;2(1):26-30. doi: 10.1038/ng0992-26. PubMed PMID: 1363881.
38. Stadler G, Rahimov F, King OD, Chen JC, Robin JD, Wagner KR, Shay JW, Emerson CP, Jr., Wright WE. Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat Struct Mol Biol*. 2013;20(6):671-8. doi: 10.1038/nsmb.2571. PubMed PMID: 23644600; PMCID: 3711615.
39. Ravnan JB, Tepperberg JH, Papenhausen P, Lamb AN, Hedrick J, Eash D, Ledbetter DH, Martin CL. Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *J Med Genet*. 2006;43(6):478-89. doi: 10.1136/jmg.2005.036350. PubMed PMID: 16199540; PMCID: 2564531.
40. Kleefstra T, Smidt M, Banning MJ, Oudakker AR, Van Esch H, de Brouwer AP, Nillesen W, Sistermans EA, Hamel BC, de Bruijn D, Fryns JP, Yntema HG, Brunner HG, de Vries BB, van Bokhoven H. Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. *J Med Genet*. 2005;42(4):299-306. doi: 10.1136/jmg.2004.028464. PubMed PMID: 15805155; PMCID: 1736026.
41. Wilson HL, Wong AC, Shaw SR, Tse WY, Stapleton GA, Phelan MC, Hu S, Marshall J, McDermid HE. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major

- neurological symptoms. *J Med Genet.* 2003;40(8):575-84. PubMed PMID: 12920066; PMCID: 1735560.
42. Descipio C, Schneider L, Young TL, Wasserman N, Yaeger D, Lu F, Wheeler PG, Williams MS, Bason L, Jukofsky L, Menon A, Geschwindt R, Chudley AE, Saraiva J, Schinzel AA, Guichet A, Dobyns WE, Toutain A, Spinner NB, Krantz ID. Subtelomeric deletions of chromosome 6p: molecular and cytogenetic characterization of three new cases with phenotypic overlap with Ritscher-Schinzel (3C) syndrome. *Am J Med Genet A.* 2005;134A(1):3-11. doi: 10.1002/ajmg.a.30573. PubMed PMID: 15704124.
43. DeScipio C, Conlin L, Rosenfeld J, Tepperberg J, Pasion R, Patel A, McDonald MT, Aradhya S, Ho D, Goldstein J, McGuire M, Mulchandani S, Medne L, Rupps R, Serrano AH, Thorland EC, Tsai AC, Hilhorst-Hofstee Y, Ruivenkamp CA, Van Esch H, Addor MC, Martinet D, Mason TB, Clark D, Spinner NB, Krantz ID. Subtelomeric deletion of chromosome 10p15.3: clinical findings and molecular cytogenetic characterization. *Am J Med Genet A.* 2012;158A(9):2152-61. doi: 10.1002/ajmg.a.35574. PubMed PMID: 22847950; PMCID: 3429713.
44. Ofir R, Wong AC, McDermid HE, Skorecki KL, Selig S. Position effect of human telomeric repeats on replication timing. *Proc Natl Acad Sci U S A.* 1999;96(20):11434-9. PubMed PMID: 10500194; PMCID: 18051.
45. Bonaglia MC, Giorda R, Beri S, De Agostini C, Novara F, Fichera M, Grillo L, Galesi O, Vetro A, Ciccone R, Bonati MT, Giglio S, Guerrini R, Osimani S, Marelli S, Zucca C, Grasso R, Borgatti R, Mani E, Motta C, Molteni M, Romano C, Greco D, Reitano S, Baroncini A, Lapi E, Cecconi A, Arrigo G, Patricelli MG, Pantaleoni C, D'Arrigo S, Riva D, Sciacca F, Dalla Bernardina B, Zoccante L, Darra F, Termine C, Maserati E, Bigoni S, Priolo E, Bottani A, Gimelli S, Bena F, Brusco A, di Gregorio E, Bagnasco I, Giussani U, Nitsch L, Politi P, Martinez-Frias ML, Martinez-Fernandez ML, Martinez Guardia N, Bremer A, Anderlid BM, Zuffardi O. Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genet.* 2011;7(7):e1002173. doi: 10.1371/journal.pgen.1002173. PubMed PMID: 21779178; PMCID: 3136441.
46. Luo Y, Hermetz KE, Jackson JM, Mulle JG, Dodd A, Tsuchiya KD, Ballif BC, Shaffer LG, Cody JD, Ledbetter DH, Martin CL, Rudd MK. Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. *Hum Mol Genet.* 2011;20(19):3769-78. doi: 10.1093/hmg/ddr293. PubMed PMID: 21729882; PMCID: 3168286.
47. Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics.* 2004;83(3):349-60. PubMed PMID: 14986705.
48. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PMCID: 2705234.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PMCID: 2723002.
50. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number

- alteration discovery in cancer by exome sequencing. *Genome Res.* 2012;22(3):568-76. doi: 10.1101/gr.129684.111. PubMed PMID: 22300766; PMCID: 3290792.
51. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25(9):1105-11. doi: 10.1093/bioinformatics/btp120. PubMed PMID: 19289445; PMCID: 2672628.
52. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* 2011;12(3):R22. doi: 10.1186/gb-2011-12-3-r22. PubMed PMID: 21410973; PMCID: 3129672.
53. Ruiz-Perez VL, Ide SE, Strom TM, Lorenz B, Wilson D, Woods K, King L, Francomano C, Freisinger P, Spranger S, Marino B, Dallapiccola B, Wright M, Meitinger T, Polymeropoulos MH, Goodship J. Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodistal dysostosis. *Nat Genet.* 2000;24(3):283-6. doi: 10.1038/73508. PubMed PMID: 10700184.
54. Ruiz-Perez VL, Tompson SW, Blair HJ, Espinoza-Valdez C, Lapunzina P, Silva EO, Hamel B, Gibbs JL, Young ID, Wright MJ, Goodship JA. Mutations in two nonhomologous genes in a head-to-head configuration cause Ellis-van Creveld syndrome. *Am J Hum Genet.* 2003;72(3):728-32. PubMed PMID: 12571802; PMCID: 1180248.
55. McKusick VA, Egeland JA, Eldridge R, Krusen DE. Dwarfism in the Amish I. The Ellis-Van Creveld Syndrome. *Bulletin of the Johns Hopkins Hospital.* 1964;115:306-36. Epub 1964/10/01. PubMed PMID: 14217223.
56. Tompson SW, Ruiz-Perez VL, Blair HJ, Barton S, Navarro V, Robson JL, Wright MJ, Goodship JA. Sequencing EVC and EVC2 identifies mutations in two-thirds of Ellis-van Creveld syndrome patients. *Human genetics.* 2007;120(5):663-70. Epub 2006/10/07. doi: 10.1007/s00439-006-0237-7. PubMed PMID: 17024374.
57. Conlin LK, Thiel BD, Bonnemann CG, Medne L, Ernst LM, Zackai EH, Deardorff MA, Krantz ID, Hakonarson H, Spinner NB. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. *Hum Mol Genet.* 2010;19(7):1263-75. doi: 10.1093/hmg/ddq003. PubMed PMID: 20053666; PMCID: 3146011.
58. Kitzman JO, Mackenzie AP, Adey A, Hiatt JB, Patwardhan RP, Sudmant PH, Ng SB, Alkan C, Qiu R, Eichler EE, Shendure J. Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nat Biotechnol.* 2011;29(1):59-63. doi: 10.1038/nbt.1740. PubMed PMID: 21170042; PMCID: 3116788.
59. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012;337(6096):816-21. doi: 10.1126/science.1225829. PubMed PMID: 22745249.
60. Sugimoto M, Furuichi Y, Ide T, Goto M. Incorrect use of "immortalization" for B-lymphoblastoid cell lines transformed by Epstein-Barr virus. *J Virol.* 1999;73(11):9690-1. PubMed PMID: 10577062; PMCID: 113012.
61. Montanaro L, Tazzari PL, Derenzini M. Enhanced telomere shortening in transformed lymphoblasts from patients with X linked dyskeratosis. *J Clin Pathol.* 2003;56(8):583-6. PubMed PMID: 12890806; PMCID: 1770038.

62. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*. 1999;402(6761):551-5. doi: 10.1038/990141. PubMed PMID: 10591218.
63. Majocchi S, Artonovska E, Mermoud N. Epigenetic regulatory elements associate with specific histone modifications to prevent silencing of telomeric genes. *Nucleic Acids Res*. 2014;42(1):193-204. doi: 10.1093/nar/gkt880. PubMed PMID: 24071586; PMCID: 3874193.
64. de Lange T. Human telomeres are attached to the nuclear matrix. *EMBO J*. 1992;11(2):717-24. PubMed PMID: 1537344; PMCID: 556504.
65. Towbin BD, Meister P, Gasser SM. The nuclear envelope--a scaffold for silencing? *Curr Opin Genet Dev*. 2009;19(2):180-6. doi: 10.1016/j.gde.2009.01.006. PubMed PMID: 19303765.
66. Zheng X, Kim Y, Zheng Y. Identification of lamin B-regulated chromatin regions based on chromatin landscapes. *Mol Biol Cell*. 2015;26(14):2685-97. doi: 10.1091/mbc.E15-04-0210. PubMed PMID: 25995381; PMCID: 4501365.
67. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008;453(7197):948-51. doi: 10.1038/nature06947. PubMed PMID: 18463634.