Evolutionary Implications of Telomere and Centromere associated Chromosomal Rearrangements in Mammals

Sunandini Ramnarayanan

A dissertation submitted for the partial fulfilment of BS-MS dual degree in Science

Indian Institute of Science Education and Research Mohali
May 2020

Certificate of Examination

This is to certify that the dissertation titled "Evolutionary Implications of Telomere and Centromere associated Chromosomal Rearrangements in Mammals" submitted by Ms. Sunandini Ramnarayanan (Reg. No. MS15143) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr Shashi Bhushan Pandit Dr Rajesh Ramachandran Dr Kuljeet Singh Sandhu

(Supervisor)

Dated: May 4, 2020

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Kuljeet Singh Sandhu at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Sunandini Ramnarayanan
(Candidate)
May 4, 2020

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Kuljeet Singh Sandhu
(Supervisor)

Acknowledgements

I'd like to extend my sincere thanks to my advisor, Dr. Kuljeet Singh Sandhu, for his guidance, support and patience.

I'd like to thank Lipika for being the support system that I needed and all the members of Genome Biology Lab: Yachna, Ken, Meenakshi, Mohan, Jui and Smriti for always being ready to help and for their insightful advice.

To my dearest friends: Ashish for dealing with my inability to understand statistics and for always looking out for me. Prathibha and Gitanjali for all their timely advice and for being there for me when I needed them the most. Thanks to Ahina and Shweta for their emotional support, especially during hard times. Thanks to Megha and Lopa for helping me take a breather when I needed it.

A big thanks to IISER Mohali for implementing a course structure that mandates all students to experience the perils and joys of research through a thesis project, and for being open-minded to allow floating courses such as HSS645, a course for which I'm very grateful to Dr. Adrene F. D'Cruz.

I'd also like to thank my thesis committee members Dr. Shashi Bhushan Pandit, Dr. Rajesh Ramachandran and my thesis advisor, Dr. Kuljeet Singh Sandhu.

I'd like to express my sincere gratitude to Dr. SK Arun Murthi, who had been assigned to me as a faculty mentor in my first year. I'll always be grateful for all the fantastic discussions we had that helped me shape the way I think today and for all the excellent advice that he has given me. I'd also like to thank Dr. N. G. Prasad in the same line for being very supportive.

I'd like to thank my parents for being incredibly supportive. They have always believed that I should carve my own path and have worked hard to ensure that I grow up with all the comforts of life. I would like to sincerely thank them for giving me the freedom to choose and the confidence to believe in myself no matter what the outcome. I'd like to thank my sister for her aid in creating some beautiful diagrams on Adobe Illustrator and for her support!

I would like to express my gratitude to KVPY for monetary support.

And lastly, biggest thanks to all the public databases out there — UCSC Genome Browser, Ensembl, BGEE, NCBI — for without these, my project wouldn't even have taken off.

— Sunandini Ramnarayanan

List of Figures

1 Illustration describing the hierarchical organization of the genome. . . 2
2 Illustration describing rearrangements that change domains 4
3 Human-Mouse Chromosomal Rearrangement Line-plots. 7
3 Human-Mouse Chromosomal Rearrangement Line-plots (cont.). . . . 8
3 Human-Mouse Chromosomal Rearrangement Line-plots (cont.). . . . 9
3 Human-Mouse Chromosomal Rearrangement Line-plots (cont.). . . . 10
4 Line plot for Human Chromosome 11. 11
5 Line plot for Human Chromsome 12. 13
7 Illustration describing rearrangements that change domains and those that do not change domains. 17
8 Criteria for classification of genes as subtelomeric or pericentromeric in a species. 18
9 Gene Expression Divergence calculated using Spearman Correlation. . 20
9 Gene Expression Divergence calculated using Spearman Correlation (cont). 21
10 Barplots comparing significance values of gene expression divergence calculated using Spearman and Pearson correlation. 22
11 Status of CNE-Gene pair Split v/s Gene Expression Divergence . . . 23
12 Comparing significance values of difference in Gene Expression Divergence (Spearman) pre and post removing CNE effect 25
13 Comparing significance values of difference in Gene Expression Divergence (Pearson) pre and post removing CNE effect 26
14 Gene Ontology Terms for genes constrained within 500KB from the Mouse Telomere 28
15 GO Molecular Terms for genes constrained within 1MB from the Mouse Telomere 29
16 Log Raw Expression values at various distances from Human Telomere. 31
17 Log Raw Expression values at various distances from Mouse Telomere. 32
18 Average Expression Divergence v/s Distance from Telomere (control: red, domain change: blue) 33
19 Opossum Chromosome 5 37
20 Some More Examples of Predicting Centromeres using Gene, Gap & Satellite Density for Opossum Chromosomes. 38
21 Some Examples of Predicting Centromeres using Gene, Gap & Satellite Density for Rabbit Chromosomes. 39

List of Tables

1 GO Molecular Terms for genes constrained near centromere of Human Chr 11 12
2 GO Biological Terms for genes constrained near centromere of Human Chr 11 12
3 GO Molecular Terms for genes constrained near telomere of Mouse Chr 10. 13
4 GO Biological Terms for genes constrained within 1MB of Mouse Telomere 29
5 GO Molecular Terms for genes constrained within 250KB of Human Telomere 30
6 GO Molecular Terms for genes constrained within 2.5MB of Human Telomere 30
7 GO Human Phenotype Terms for genes constrained within 10MB of Human Telomere 30
8 Predicted Opossum Centromeres 40
9 Predicted Rabbit Centromeres 40

Notation

1. Chr: Chromosome
2. KB: Kilo Base Pairs (1000 Base Pairs)
3. MB: Mega Base Pairs (10^6 Base Pairs)
4. Hum: Human

Databases used

1. Ensembl Biomart (Ensembl98): *https://asia.ensembl.org/biomart/martview/*
2. UCSC Genome Browser: *https://genome.ucsc.edu/cgi-bin/hgTables*
3. ToppGene: *https://toppgene.cchmc.org/*
4. Gene Ontology Resource: *http://geneontology.org/*
5. NCBI GEO: *https://www.ncbi.nlm.nih.gov/geo/*
6. BGEE: *https://bgee.org/*

Contents

1 Abstract **1**

2 Introduction **2**

3 Investigating Position Effect of Linear Rearrangements **6**
3.1 Using Line Plots 6
3.1.1 Materials & Methodology 6
3.1.2 Results 11
3.1.3 Conclusion 15
3.2 Determining Expression Divergence 17
3.2.1 Introduction 17
3.2.2 Materials & Methodology 19
3.2.3 Results 20
3.2.4 Controlling for CNEs-Gene Split effect 23
3.2.5 Gene Ontologies 28
3.2.6 Raw Expression Comparison 31
3.2.7 Conclusion 33
3.3 Predicting Centromeres 35
3.3.1 Introduction 35
3.3.2 Materials & Methodology 36
3.3.3 Results 37
3.3.4 Conclusion 41

4 Discussion & Future Directions **42**

References **43**

1 Abstract

The hierarchical organisation of the genome allows for it to be regulated at several levels. At the linear level, it involves gene order and neighbourhood, at the chromatin level it involves histone modifications, and at the 3D level it brings in the dynamics of large domains: TADs, cLADs, ciLADs, etc. In this study, we explore the linear ordering of genes and the rearrangements that disrupt them. Chromosomal rearrangements that lead to changes in the linear order of genes are implicated in changes in gene expression. Change in gene expression is, in turn, linked to phenotypic change that is subject to Darwinian selection. Therefore, chromosomal rearrangements play an important role in the course of evolution. Here we explore a subset of chromosomal rearrangements that constrain genes near telomere or centromere in a species-specific manner. We hypothesised that genes constrained near or away from heterochromatin (centromere/telomere) via rearrangements experience differential gene expression compared to those genes that maintain their distance from such domains. We studied the rearrangements that shuffle orthologous genes around in human and mouse chromosomes via two methods: using line plots to visualise rearrangements and using distance cut-offs to filter genes constrained within certain distances of the telomere/centromere in a species-specific manner. Through the second approach, we observed that genes constrained near telomeres show high expression divergence as compared to those genes that maintain their distance from heterochromatin in both human and mouse. The genes constrained near telomere in mouse are associated with pathways in immune response and cancer. Since immune responses are just as important as adaptations for the survival of a species, this result warrants further analysis of the hypothesis in other mammalian species. However, a lack of centromere positions for several mammalian chromosomal assemblies is a hindrance. Centromere position prediction was attempted for mammals with complete chromosomal assemblies, with some success for the assemblies of Opossum and Rabbit. We also come to an understanding, through this analysis, that linear rearrangements that disrupt the 3D organisation of the genome (constitutive LAD to constitutive inter-LAD or *vice versa*) may provide a more comprehensive outlook in elucidating the evolutionary role of rearrangements that displace genes from repressed to expressed compartments or *vice versa* in a species-specific manner — an outlook that studying linear rearrangements with heterochromatin as a repressed compartment alone may not provide.

Keywords: Chromosomal Rearrangements, Heterochromatin, Telomere, Centromere, Expression Divergence, Human, Mouse, Mammals.

2 Introduction

The genome is organized hierarchically. The sheer length of the DNA versus the 3D space available to it within the eukaryotic nucleus necessitates such organization. Given that the genome is hierarchically organized, this allows it to be regulated at several levels (Finn & Misteli, 2019; Nguyen & Bosco, 2015).

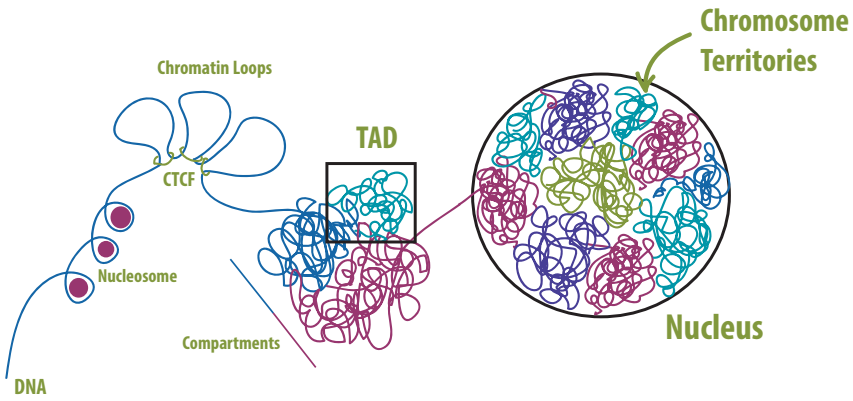


Figure 1: Illustration describing the hierarchical organization of the genome.

At the linear level, genes involved in development are often clustered together, allowing for greater control over its expression. For instance, the HOX genes have maintained their collinearity throughout evolution. The order in which they are present is the same as the order in which they are expressed along the anteroposterior axis. One of the theories that attempt to explain this maintained collinearity is sharing of enhancer (regulatory elements) among the genes (Mann, 1997; Mallo & Alonso, 2013). At the chromatin level, different histone marks can allow for the expression or repression of a gene. For instance, H3K27Me3 is associated with gene silencing, and H3K4Me3 is associated with gene expression. Bivalent marks (having both H3K27Me3 and H3K4Me3) on genes indicates a poised state that is often observed in early developmental stages and is associated with sharper control of gene expression during development: enabling expression when the timing is right and upon receiving proper cues (Voigt, Tee, & Reinberg, 2013; Dong & Weng, 2013). At the 3D level, the chromatin loops over itself and forms domains: Topologically Associated Domains (TADs), which have been implicated in the tight regulation of gene expression in highly coordinated events like that of random inactivation of X-chromosome (Nora et al., 2012).

These 3D domains can be mapped using advanced genomic techniques such as Hi-C. This, in turn, facilitates the study of the dynamics of these domains and their effect on gene expression. For instance, 3D reorganization of TADs has been recently implicated in inter-species differences in Humans and Chimpanzees (Eres, Luo, Hsiao, Blake, & Gilad, 2019).

Therefore, it is evident that alteration in genome regulation at any of these levels can lead to a modification in its transcriptional environment (Shapiro, 1982). This change in expression is viewed as a medium of bringing out phenotypic novelty, as it has been suggested that gene expression divergence is subject to purifying selection and to positive Darwinian selection (Jordan, Mariño-Ramírez, & Koonin, 2005). An elegant study of the linear rearrangement of genes in humans and chimpanzees revealed that genes which change genic neighbourhood show elevated expression divergence compared to those that have conserved genic neighbourhood (De & Babu, 2010; De, Teichmann, & Babu, 2009). These genes show higher expression divergence mostly in brain tissues, possibly reflecting the functional differences of the brain in Humans and Chimpanzees. A similar study, but concerning a different set of mammalian species shows that loss of linear proximity to conserved non-coding elements (owing to rearrangements) specifically in rats is associated with down-regulation of neurogenesis, which ties up neatly in the light of the fact that rats have lissencephalic (smooth) brains in comparison to humans and several other mammals with gyrated brains (Bagadia et al., 2019). As mentioned earlier, 3D reorganization of the genome in Humans and Chimpanzees is hypothesized to give rise to inter-species differences through differential expression (Eres et al., 2019). Therefore, the phenomenon of rearrangements and 3D organization dynamics is rightly put into spotlight with reference to evolution.

To elucidate what causes a change in gene expression and ultimately culminates in the evolution of a new species is an enrapturing problem and is the main theme of this thesis. Gene orders are non-random in eukaryotes. The orders that are maintained in evolution often have a functional or transcriptional relationship (Lopez, Guerra, & Samuelsson, 2010). Therefore, any rearrangement that breaks such conserved order of genes and thrusts several genes into different environments is then naturally expected to bring about changes in gene expression. As it turns out, eukaryotic gene expression is most certainly affected by duplications, loss and rearrangements (Shapiro, 1982; Brown, 1981; De & Babu, 2010; De et al., 2009). It has also been shown that chromosomes in Humans and Chimpanzees that show a greater number of rearrangements show a greater change in expression pattern (Marques-Bonet et al., 2007).

Chromosomal aberrations that modify gene order are often implicated in diseases and cancer (Hasty & Montagna, 2014). Given all this information, it is not a stretch to think that such naturally occurring chromosomal aberrations have a role to play in evolution, as a means of bringing about diversity via changed expression pattern of genes.

Since linear rearrangements have been established as an agent that brings about change in gene expression divergence, this thesis aims to explore the impact of a subset of linear genomic rearrangements in mammals on gene expression divergence. Here we explore the possibility of chromosomal rearrangements, which constrain genes either towards or away from heterochromatin in particular lineages, as agents of gene expression change that culminate in lineage-specific adaptations.

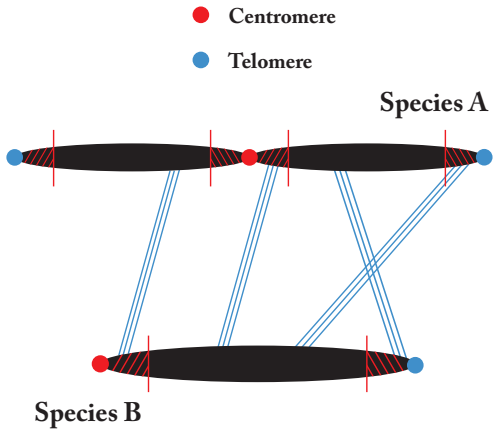


Figure 2: Illustration describing rearrangements that change domains

Eukaryotic chromatin needs to open to allow for gene expression. Telomeres and Centromeres are tightly packed, allowing for little to no gene expression (Jost, Bertulat, & Cardoso, 2012). Therefore, the effect of heterochromatin may pervade the environment of the genes proximal to these regions and affect gene expression. Here, we hypothesize that genes constrained near or away from heterochromatin domains such as the Centromere or Telomere via rearrangements experience differential expression compared to those genes that maintain their distance from such domains. This is called position effect: where the expression of a gene is affected due to its genomic neighbourhood or chromosomal environment and not due to a change in the sequence of the gene or a change in its transcription machinery (Elgin & Reuter, 2013; De & Babu, 2010).

For instance: the mottled eye of Drosophila is caused by a rearrangement that displaces the white gene to a region very close to the centromere of X-chromosome. Therefore, in some ommatidia of the Drosophila compound eye, this gene is silenced, whereas in others it expresses wild type red colour. The gene itself is not mutated, but its changed expression is attributed to its changed position (Elgin & Reuter, 2013; Muller, 1941). Such rearrangements have also been observed in several human diseases (D. A. Kleinjan & van Heyningen, 2005; D.-J. Kleinjan & Van Heyningen, 1998). Other examples of position effect include transgenes which when inserted near the telomere in mouse embryonic stem cells show repressed activity, indicating that gene expression is hampered in an environment such as those near telomeres (Pedram et al., 2006). Since such rearrangements have been observed in different species (De & Babu, 2010; De et al., 2009), we hypothesize that they might result in tweaked phenotypes through gene expression divergence, which may be advantageous to a species' survival.

Therefore, it all boils down to understanding if rearrangements that constrain genes near heterochromatin in a species-specific manner affect gene expression and if this changed gene expression divergence can be linked to lineage-specific adaptations.

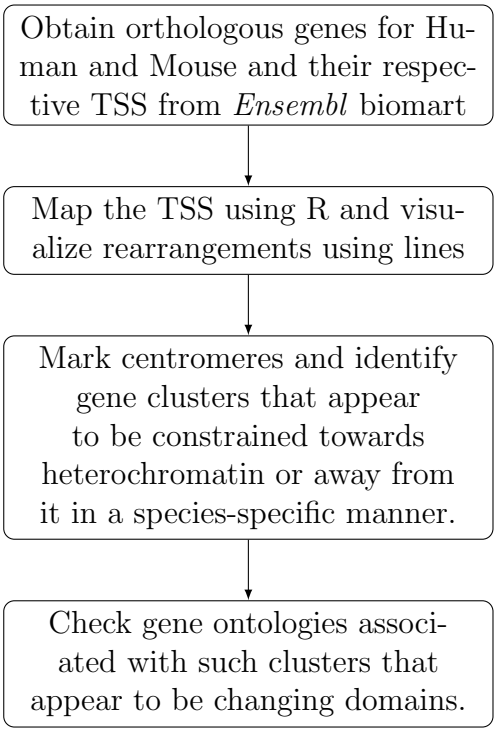
3 Investigating Position Effect of Linear Rearrangements

3.1 Using Line Plots

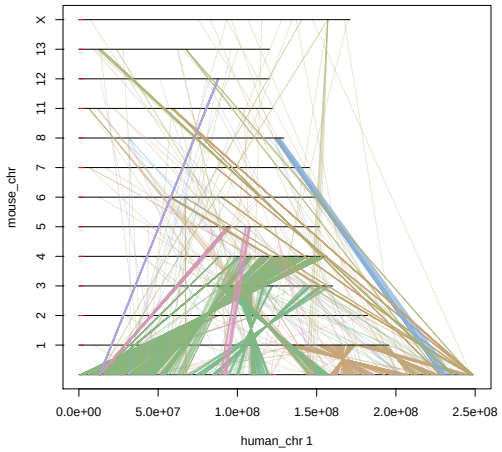
Line plots are simple yet fantastic tools for visualizing rearrangements. Each line joins the TSS (Transcription Start Site) of a human gene to the TSS of its orthologous gene in mouse or any other species. This allows us to visualize synteny breakpoints and also approximate visually the rearrangements that are constraining genes near the telomere or centromere or away from heterochromatin in a species-specific manner.

The following line plots map the TSS of each gene on every human chromosome to the TSS of its respective orthologous gene on mouse chromosomes. The line plots are drawn for 23 (22 autosomes + X) human chromosomes.

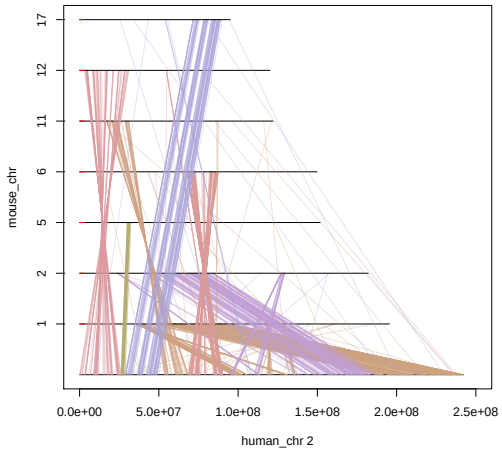
3.1.1 Materials & Methodology



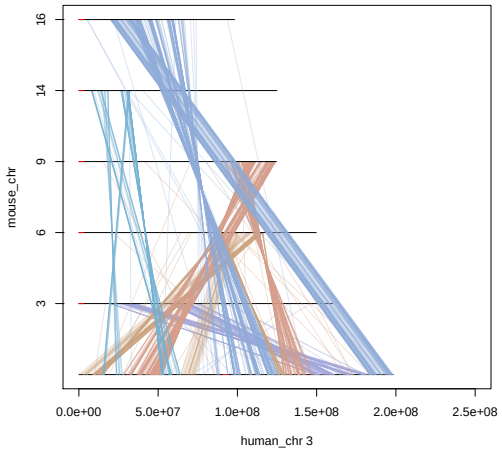
Orthologous genes were obtained from *Ensembl Biomart v98*.



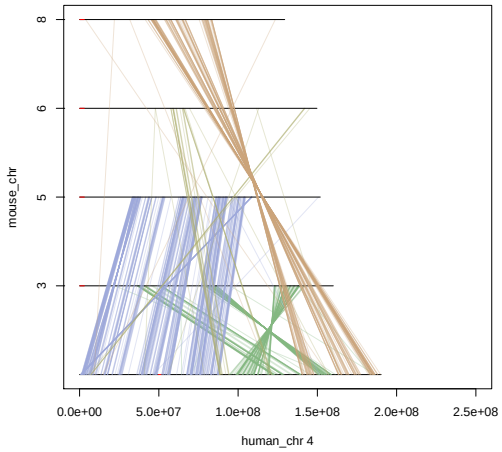
(a) Human Chr 1



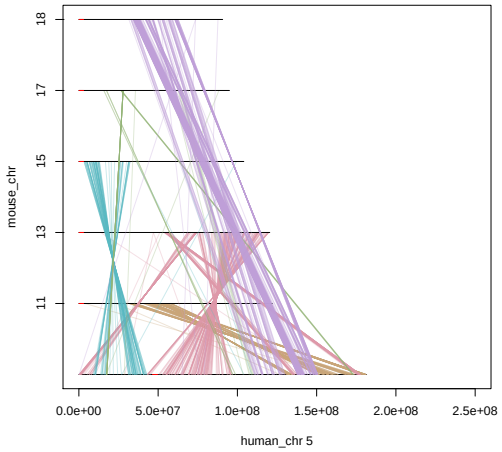
(b) Human Chr 2



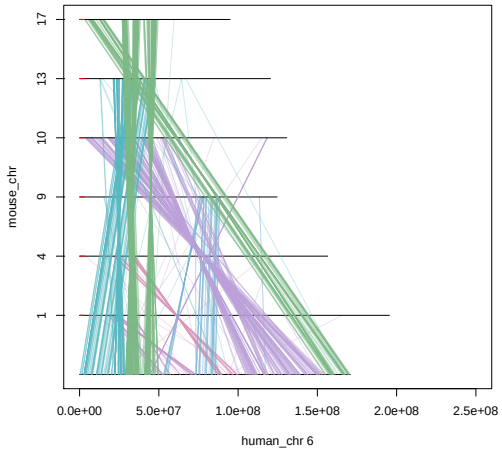
(c) Human Chr 3



(d) Human Chr 4

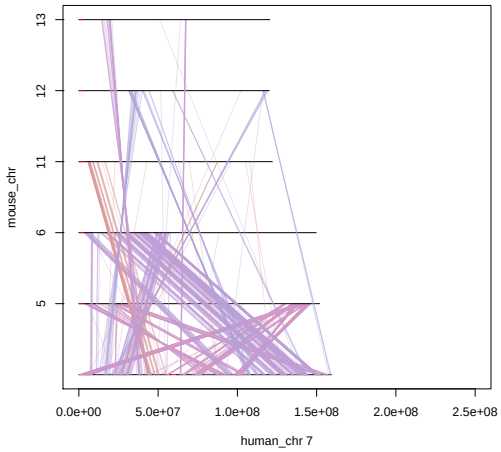


(e) Human Chr 5

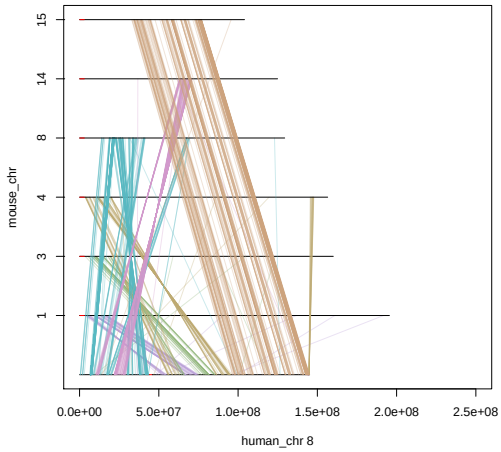


(f) Human Chr 6

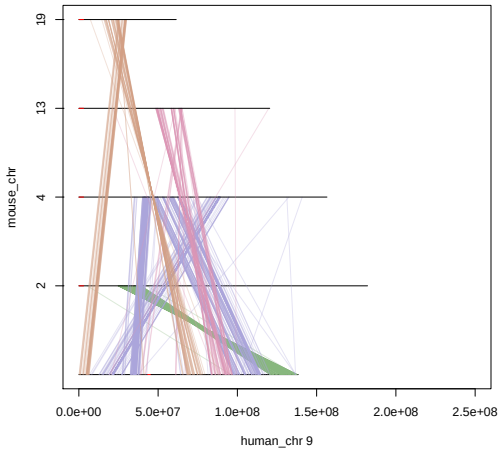
Figure 3: Human-Mouse Chromosomal Rearrangement Line-plots.



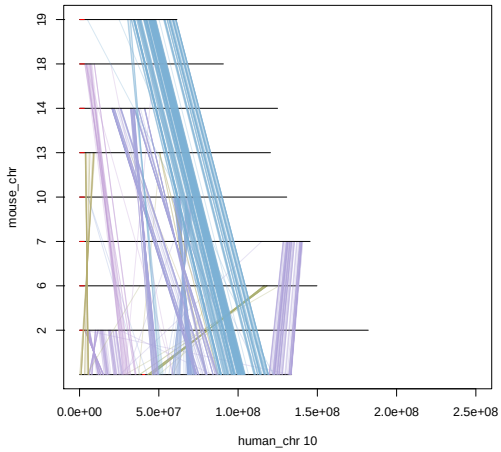
(g) Human Chr 7



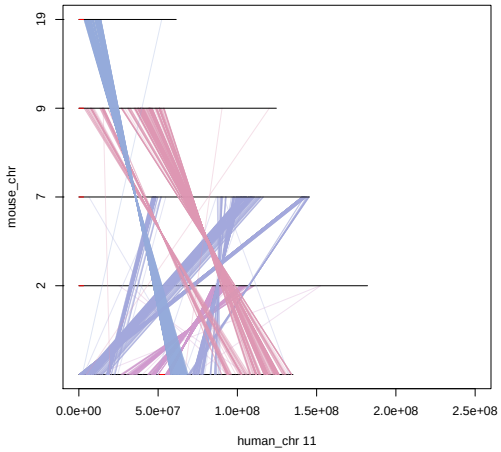
(h) Human Chr 8



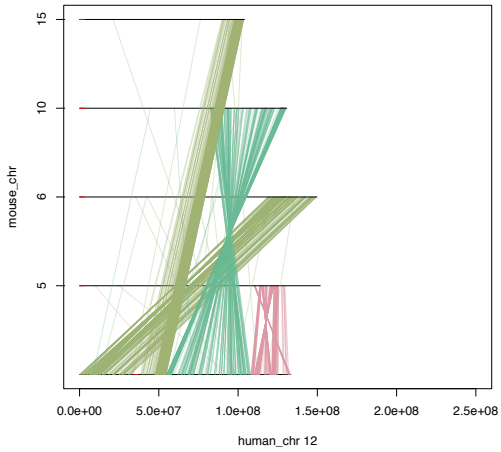
(i) Human Chr 9



(j) Human Chr 10

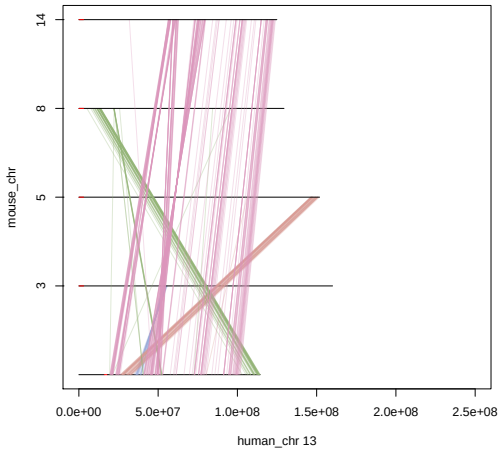


(k) Human Chr 11

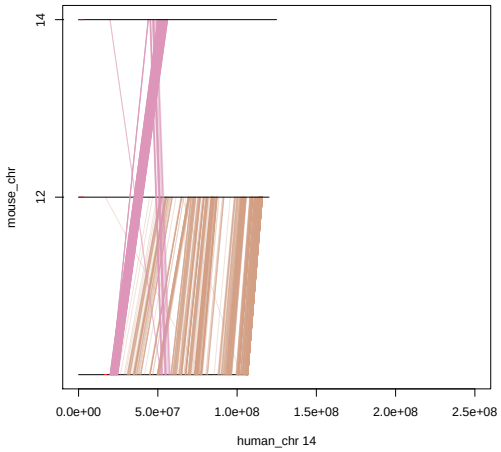


(l) Human Chr 12

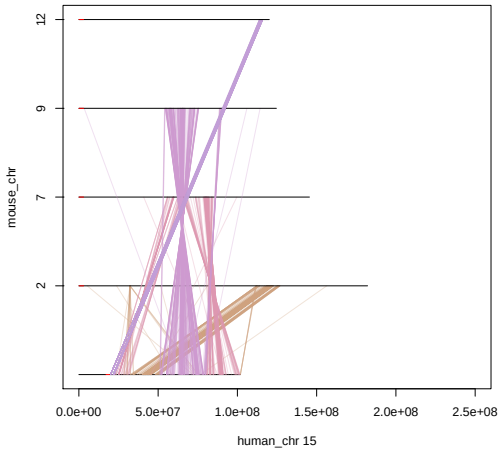
Figure 3: Human-Mouse Chromosomal Rearrangement Line-plots (cont.).



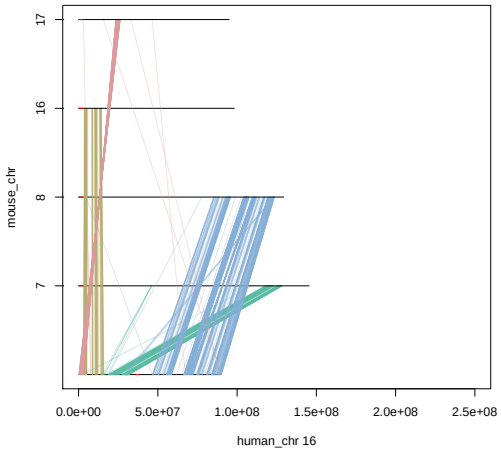
(m) Human Chr 13



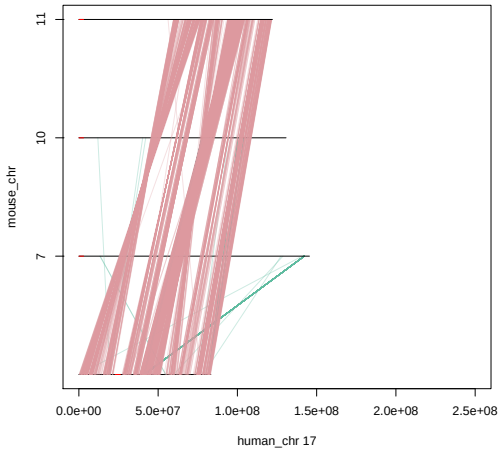
(n) Human Chr 14



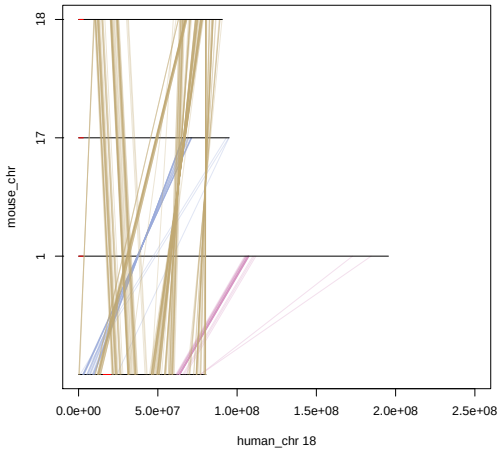
(o) Human Chr 15



(p) Human Chr 16

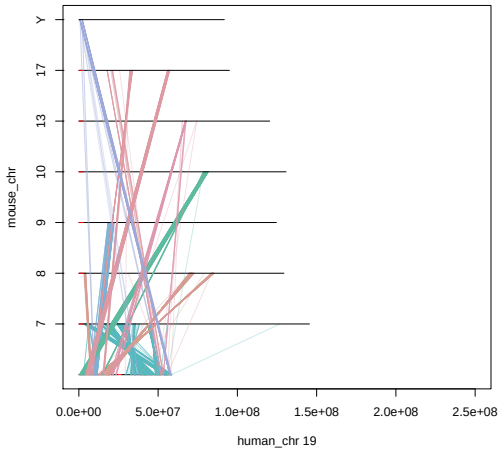


(q) Human Chr 17

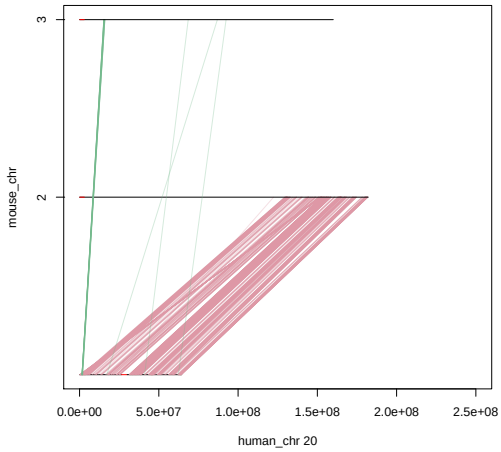


(r) Human Chr 18

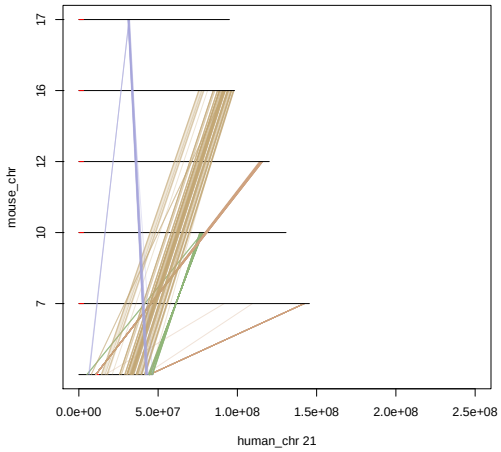
Figure 3: Human-Mouse Chromosomal Rearrangement Line-plots (cont.).



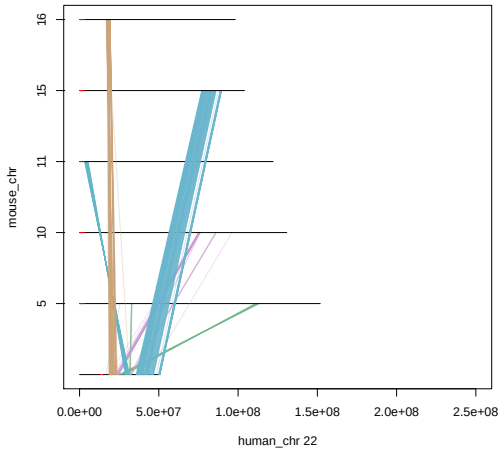
(s) Human Chr 19



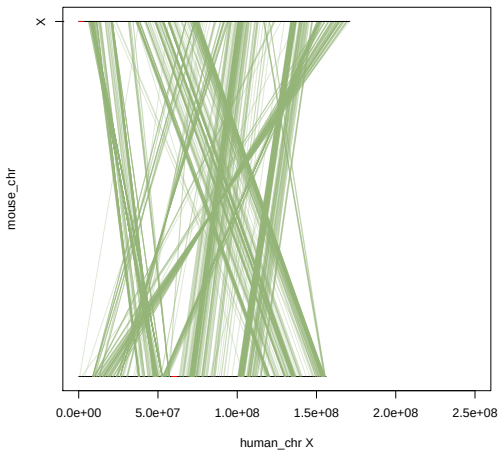
(t) Human Chr 20



(u) Human Chr 21



(v) Human Chr 22



(w) Human Chr X

Figure 3: Human-Mouse Chromosomal Rearrangement Line-plots (cont.).

3.1.2 Results

$$-log_{10}(0.05) = 1.301 \,\&\, log_{10}(0.05) = -1.301$$

Any value more positive or more negative, *respectively*, is significant.

Example 1: Rearrangement of genes near centromere of Human Chromosome 11 to central region of Mouse Chromosome 2.

By central region, here we mean regions that appear far enough from centromere or telomere on a chromosome visually.

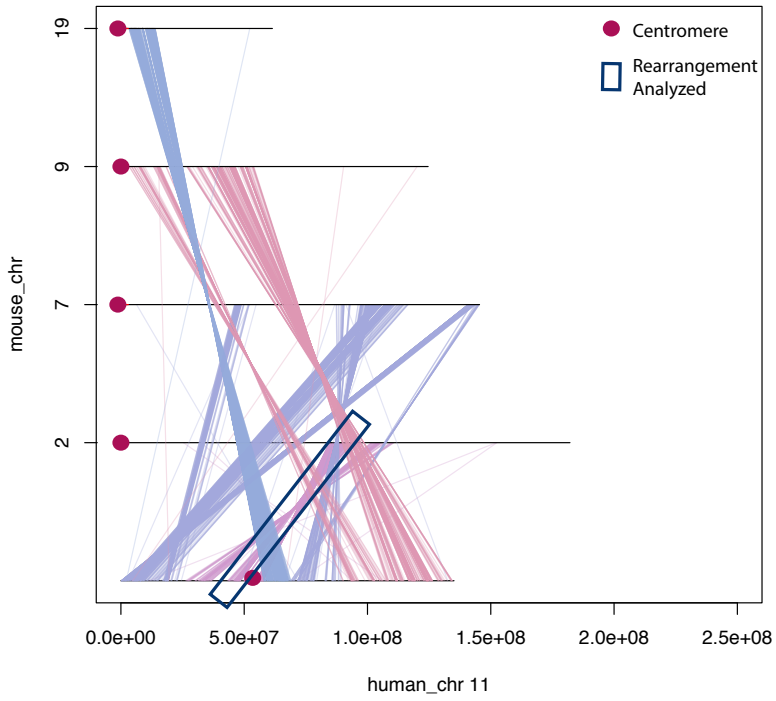


Figure 4: Line plot for Human Chromosome 11.

Table 1: GO Molecular Terms for genes constrained near centromere of Human Chr 11

ID	GO Term	log p-value
GO:0004984	olfactory receptor activity	-15.6234
GO:0005549	odorant binding	-12.2832
GO:0060089	molecular transducer activity	-9.0155

Table 2: GO Biological Terms for genes constrained near centromere of Human Chr 11

ID	GO Term	log p-value
GO:0035608	protein deglutamylation	-3.6326
GO:0050911	detection of chem. stimulus in sensory perception of smell	-15.8182
GO:0051606	detection of stimulus	-11.8386
GO:0007186	G-protein coupled receptor signaling pathway	-8.1158

Example 2: Rearrangement of genes near central region of Human Chromosome 12 to telomere of Mouse Chromosome 10.

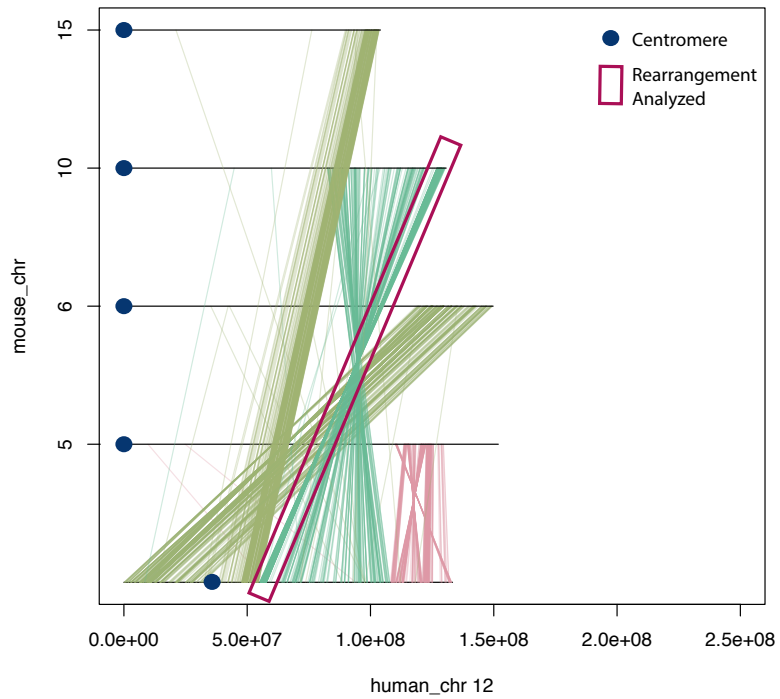
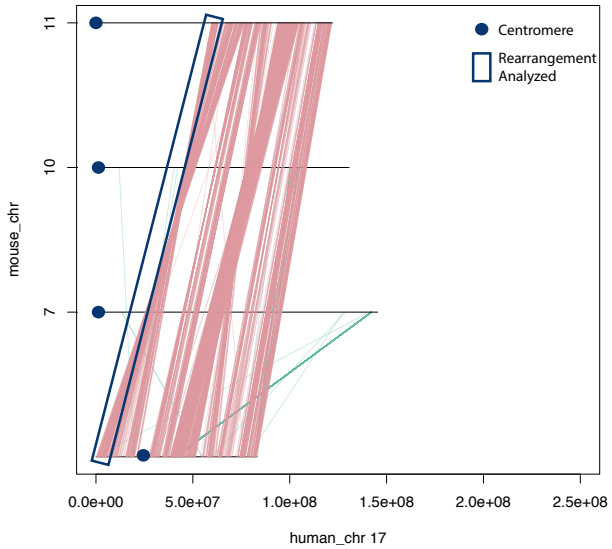


Figure 5: Line plot for Human Chromsome 12.

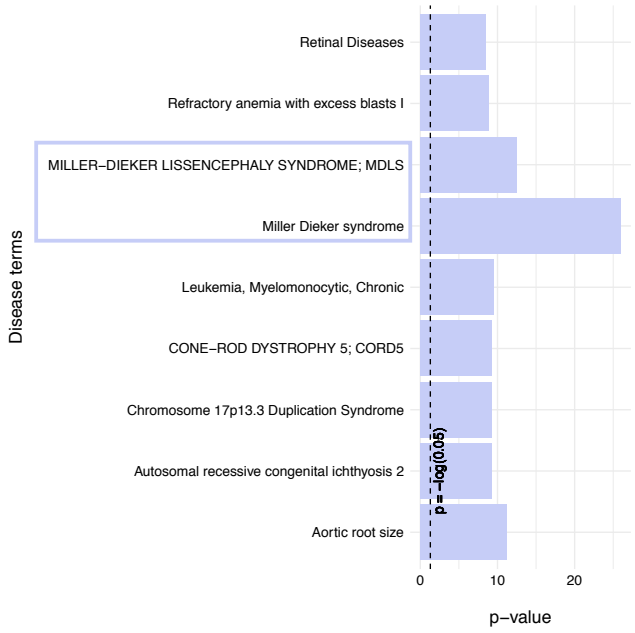
Table 3: GO Molecular Terms for genes constrained near telomere of Mouse Chr 10.

ID	GO Term	log p-value
GO:0004984	olfactory receptor activity	-15.6234

Example 3: Rearrangement of genes near telomere of Human Chromosome 17 to central region of Mouse Chromosome 11.



(a) Line plot for Human Chromosome 17.



(b) Diseases associated with genes constrained near telomere of Human Chr 17.

3.1.3 Conclusion

Of all such rearrangements, I've presented the following three to drive home the fact that some rearrangements are explicable by our hypothesis and some are not. The rearrangement constraining olfactory genes towards the human centromere in "example 1" can be explained by the fact that humans and mice are known to have diverged with respect to the sense of smell (olfaction), indicated by excessive loss of olfactory receptors in humans. Adding to it, about 50% of the olfactory genes in humans are pseudogenes (Gilad, Man, Pääbo, & Lancet, 2003). Therefore, it is quite likely that rearrangements that constrain these olfactory genes near the human centromere, which would lead to reduced expression of these receptors, may play a role in further reducing the reliance on sensory perception of smell in humans compared to apes and other mammalian species.

However, the rearrangement constraining olfactory receptors towards mouse telomere have also been identified (example 2), and this does not fit into our hypothesis very well with the previous reasoning. Additionally, if the same logic of reduced expression near telomeric regions were to be applied to the genes constrained at human telomere (example 3) that play a role in brain development (the loss of which leads to lissencephaly), it is humans that should show lissencephalic brains and not mice!

This discrepancy leads to two ideas. One: the hypothesis needs to be polished and refined. Regions linearly proximal to the telomere may not be as close to the telomere in the light of the 3D organization of the genome (Naumova & Dekker, 2010). Two: a more methodological approach may be needed to elucidate the role of such rearrangements.

The pros of the line-plot approach are: Entire blocks of genes can be classified as constrained towards centromere/telomere or constrained away from telomere/centromere. The advantage of such a method of classification is that since there is no limit on the size of the rearrangements that can be studied, arbitrary loss of genes from gene ontology analysis is prevented (such as those that would occur when giving a distance cut-off). This, in turn, gives stronger and clearer results when analyzing said block of genes for gene ontologies. The additional positive aspect of this approach is that this analysis is done chromosome-wise unlike the method of analysis proposed in the latter part of this conclusion that pools, regardless of chromosome, all rearrangements falling into the category of domain change (euchromatin to heterochromatin or *vice versa*), the disadvantage being that pooling rearrangements will subdue representation of smaller sets of genes with different functions. In a chromosome-wise approach, this disadvantage is mitigated.

There are some shortcomings of the line-plot approach: looking at line-plots depicting rearrangements between two species alone is not enough to truly determine if a block of genes is species-specifically constrained towards the telomere/centromere. Outgroup studies are a must.

The other shortcoming is that one cannot ascertain, in a methodological fashion, if indeed these blocks of genes that are constrained towards centromere/telomere or are constrained away from it in one species do diverge in expression with respect to the reference species. This approach also undermines the fact that the effect of heterochromatin may pervade only up to a certain linear distance from the telomere/centromere. Therefore, a distance-wise classification of rearrangements will eliminate those genes that are too far away linearly from the centromere or telomere to experience the heterochromatin effect.

Hence, the following methodology was devised to consider the factor of distance and to be able to calculate expression divergence associated with each distance. It also brings in the advantage of adding an outgroup to truly elucidate the species-specific nature of the rearrangements that constrain genes towards or away from heterochromatin.

3.2 Determining Expression Divergence

3.2.1 Introduction

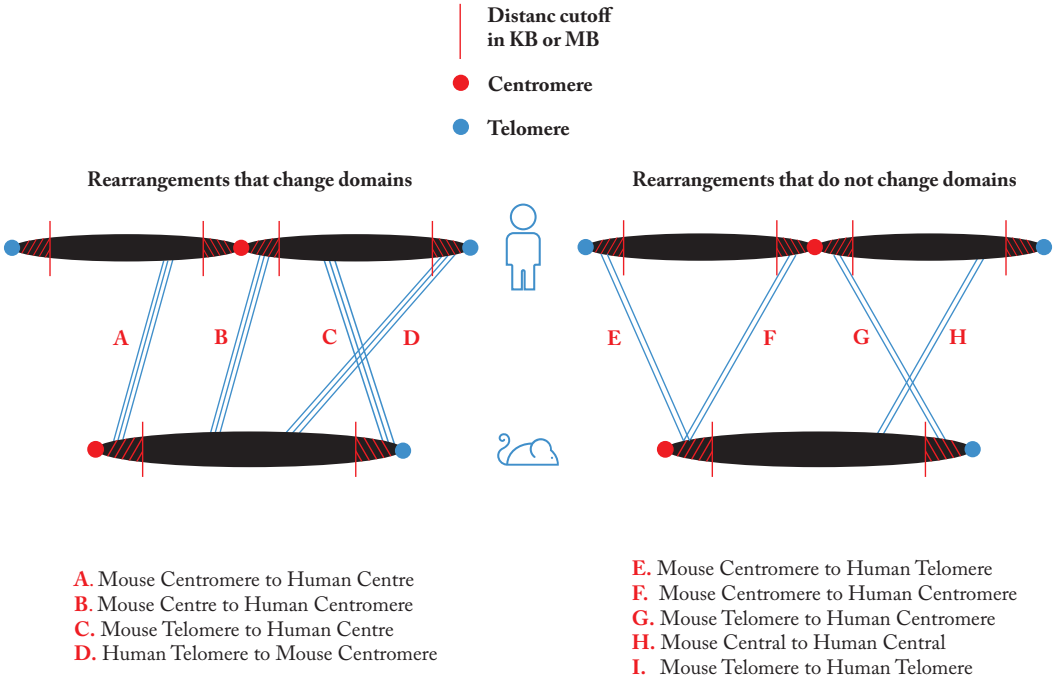
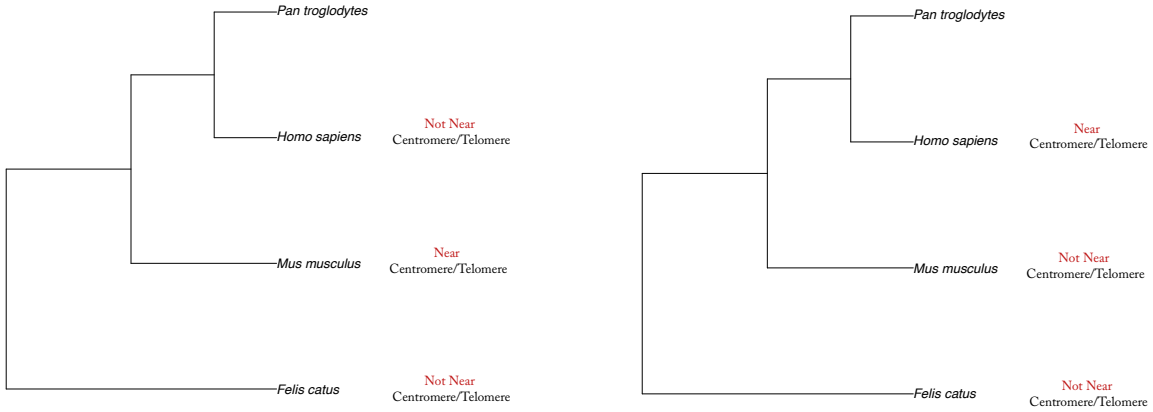


Figure 7: Illustration describing rearrangements that change domains and those that do not change domains.

Since a preliminary analysis using line-plots reflected interesting yet inconsistent gene ontologies, a further methodological approach using distance cut-offs in increasing order (50KB, 100KB, 500KB, 1MB and so on till 10MB) from the telomere and centromere to capture rearrangements that specifically constrained genes to these regions was devised.

Therefore, Human-Mouse rearrangements were run through a program that filtered genes specifically constrained within distance cut-offs from Mouse Telomere, Human Telomere, Mouse Centromere and Human Centromere. Cat was used as an outgroup as it was one of the few organisms for which centromeric positions were available and also because Cat as a species is sufficiently distant from both Human and Mouse.

To define the control set against which the genes constrained towards heterochromatin were to be tested for differential expression, it was decided that distances greater than 10MB from telomere and centromere was far enough. Genes that constitute the control set are those that maintain their distance from heterochromatin in mouse and human depending on the distance cut-offs. That is, those genes that remain within the distance cut-off from the telomere/centromere and those that remain farther than 10MB from telomere/centromere in both species. In contrast, genes that change their distance from heterochromatin (either towards or away) depending on distance cut-offs were to be put in a set termed as "domain change".



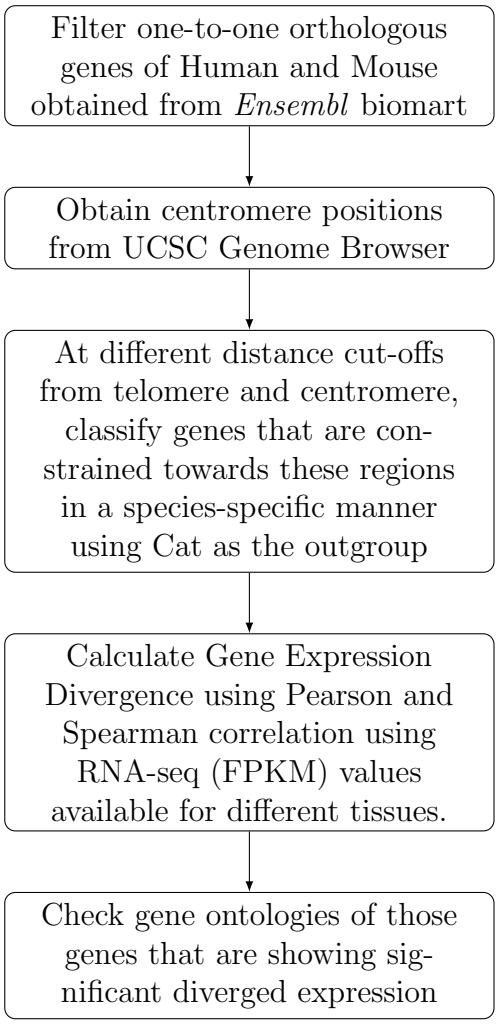
(a) Mouse specific (b) Human specific

Figure 8: Criteria for classification of genes as subtelomeric or pericentromeric in a species.

Therefore, in the boxplots ahead, "mouse telomere" would compare two sets of genes: "Domain change", which contains genes that are constrained towards the mouse telomere within xKB/MB but are far away from heterochromatin (10 MB or farther from telomere/centromere) in Human and Cat, and "Control" set, which contains those genes that are either within xKB/MB of heterochromatin or are father than 10MB from heterochromatin in both Human and Mouse.

To ascertain significance values of difference in gene expression divergence between these two sets, one-sided Wilcoxon-test was used (that is, "domain change" set has higher expression divergence compared to "control" set).

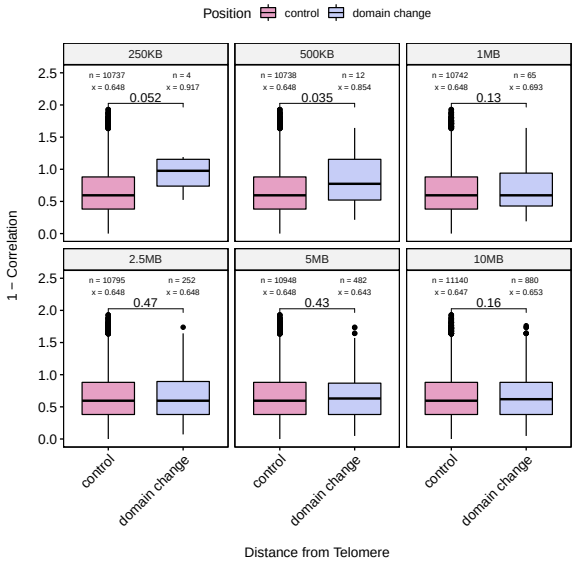
3.2.2 Materials & Methodology



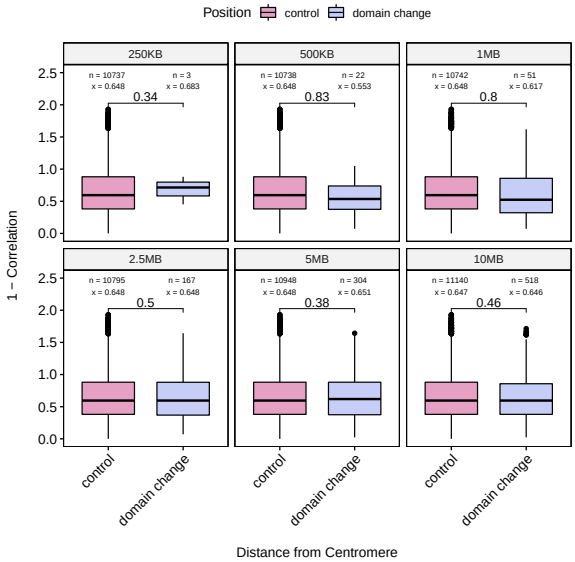
To calculate Gene expression divergence, the following tissues were used: Brain, Colon, Heart, Kidney, Liver, Lung, Skeletal Muscle, Testis. These were obtained from *BGEE* database. Human Tissue expression data was obtained from the following GSE IDs: GSE30611. Mouse Tissue expression data was obtained from the following IDs: GSE41637 and GSE41338 (Only Heart). Before calculating gene expression divergence, the tissues were quantile normalized. The correlation for each vector containing Human and Mouse tissues in order was calculated using Spearman and Pearson correlation. Subtracting the correlation from 1 gives the value of gene expression divergence. Higher the value (between 0 and 2), greater the divergence.

3.2.3 Results

Gene Expression Divergence calculated using ***Spearman*** *Correlation:*

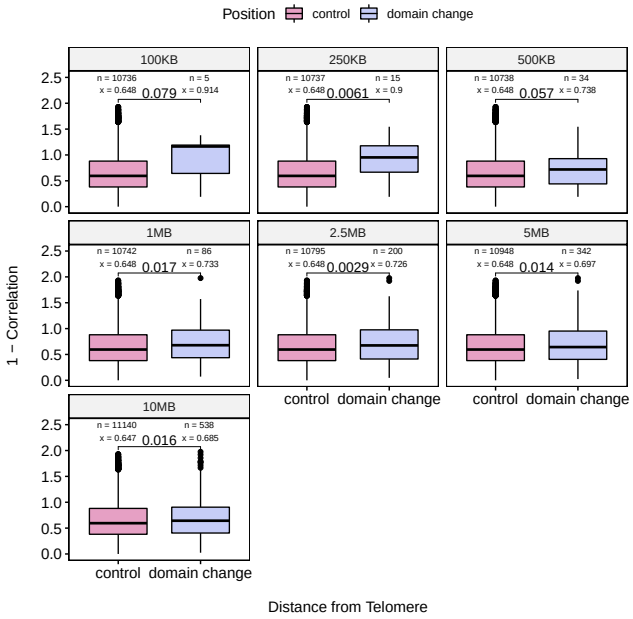


(a) Gene Expression Divergence at various distances from Mouse Telomere

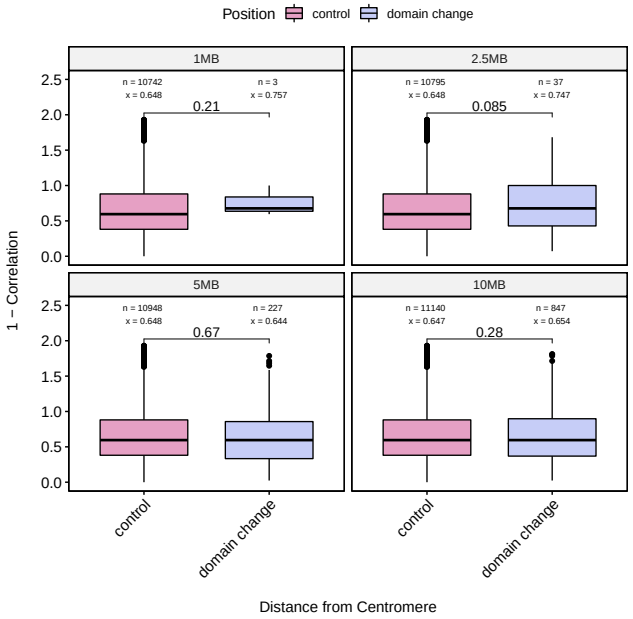


(b) Gene Expression Divergence at various distances from Mouse Centromere

Figure 9: Gene Expression Divergence calculated using Spearman Correlation.



(c) Gene Expression Divergence at various distances from Human Telomere

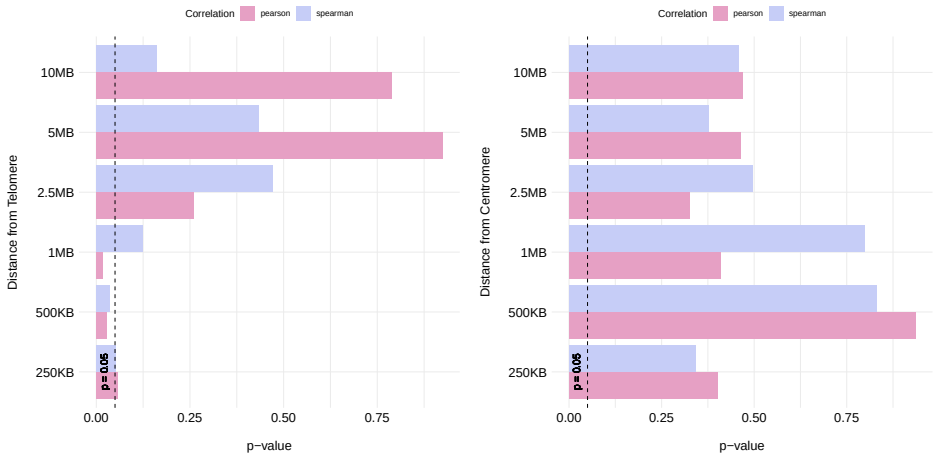


(d) Gene Expression Divergence at various distances from Human Centromere

Figure 9: Gene Expression Divergence calculated using Spearman Correlation (cont).

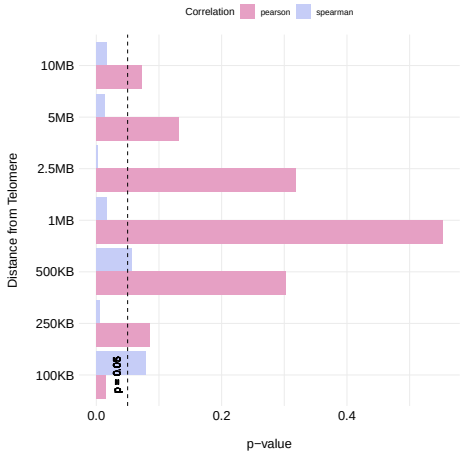
Calculating the same value of gene expression divergence using **Pearson** correlation gives slight variation when it comes to which distances from telomere show significant expression divergence for mouse and human. To make the comparison easier, the following bar plots depict the p-values obtained when comparing the expression divergence (calculated using different methods) of control set v/s the "genes that change domains" (domain change) set.

Comparing results obtained using different methods of calculating expression divergence:

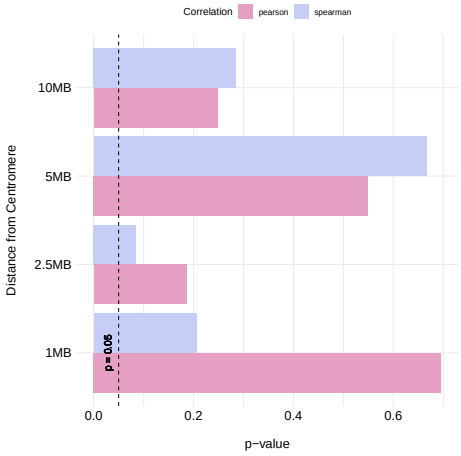


(a) Mouse Telomere

(b) Mouse Centromere



(c) Human Telomere



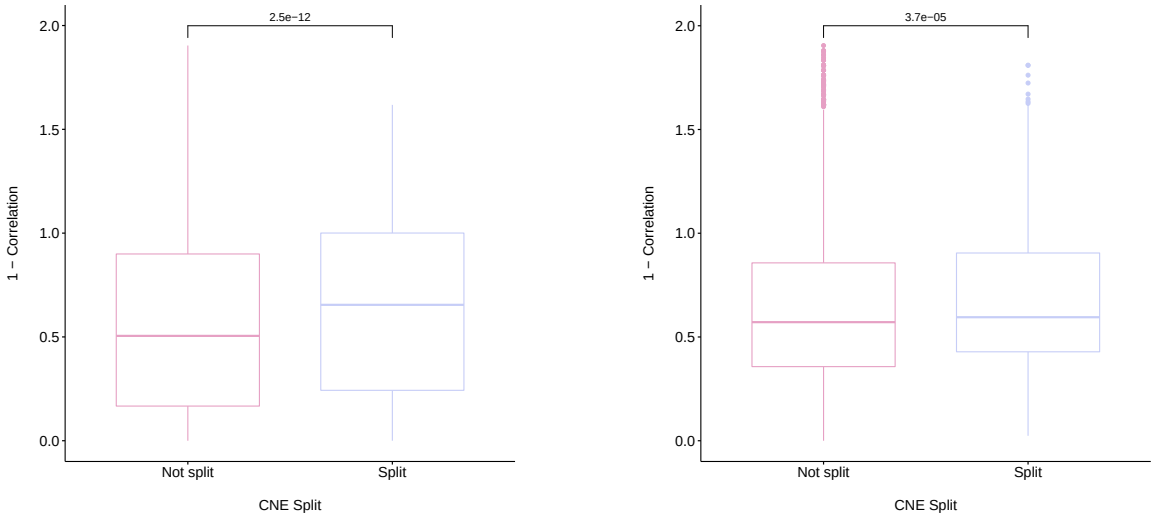
(d) Human Centromere

Figure 10: Barplots comparing significance values of gene expression divergence calculated using Spearman and Pearson correlation.

3.2.4 Controlling for CNEs-Gene Split effect

Linear rearrangements have been the focus of research for quite some time now. Several aspects of linear rearrangements have been studied, one example is the effect of the altered genic environment on gene expression (De & Babu, 2010; De et al., 2009). As mentioned in the introduction, Human and Chimpanzee chromosomes with a greater number of rearrangements show greater differential expression (Marques-Bonet et al., 2007). Subhajyoti De's paper (De et al., 2009) on linear rearrangements and altered genic neighbourhood briefly comments that regions near the centromere and telomere show low CGN score (lower CGN score indicating greater altered genic environment). However, they did not include regions near the telomere and centromere in their analysis due to fear of incorrect sequencing around these segments owing to excessive repeats. Ten years from their paper, one can see that continuous efforts have improved the quality of genome assemblies and annotations. Hopefully, this thesis serves as an attempt to study these excluded regions.

Another effect that has been studied when it comes to linear rearrangements is the role of the position of regulatory elements with respect to genes that they control. A split that is a large distance between a conserved non-coding element and the gene it controls, due to rearrangements, in several gene-CNE pairs specific to a species, has been shown to play a role during its development (Bagadia et al., 2019)



(a) Pearson correlation

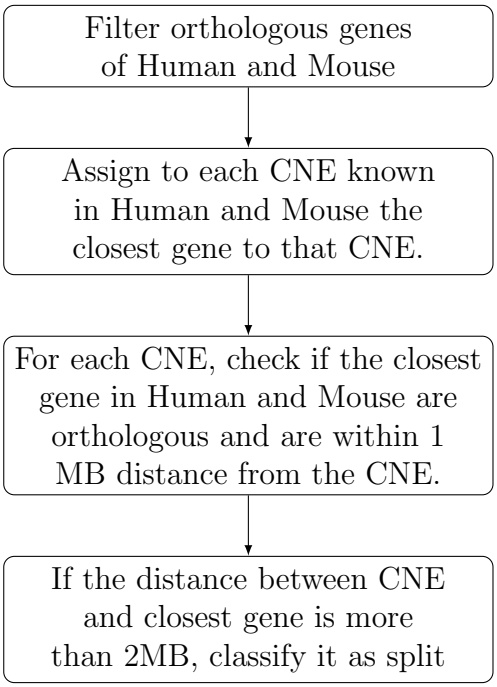
(b) Spearman correlation

Figure 11: Status of CNE-Gene pair Split v/s Gene Expression Divergence

Such gene-CNE split pairs often show differential gene expression in comparison to those gene-CNE pairs that remain intact throughout evolution.

Therefore, it is one of the effects that can be controlled for in the study of telomeric/centromeric effect. It is essential to isolate, as far as possible, the impact of telomere/centromere on the expression of genes that have been constrained towards it in a species-specific manner from effects of other known factors that arise due to linear rearrangements (such as the CNE effect).

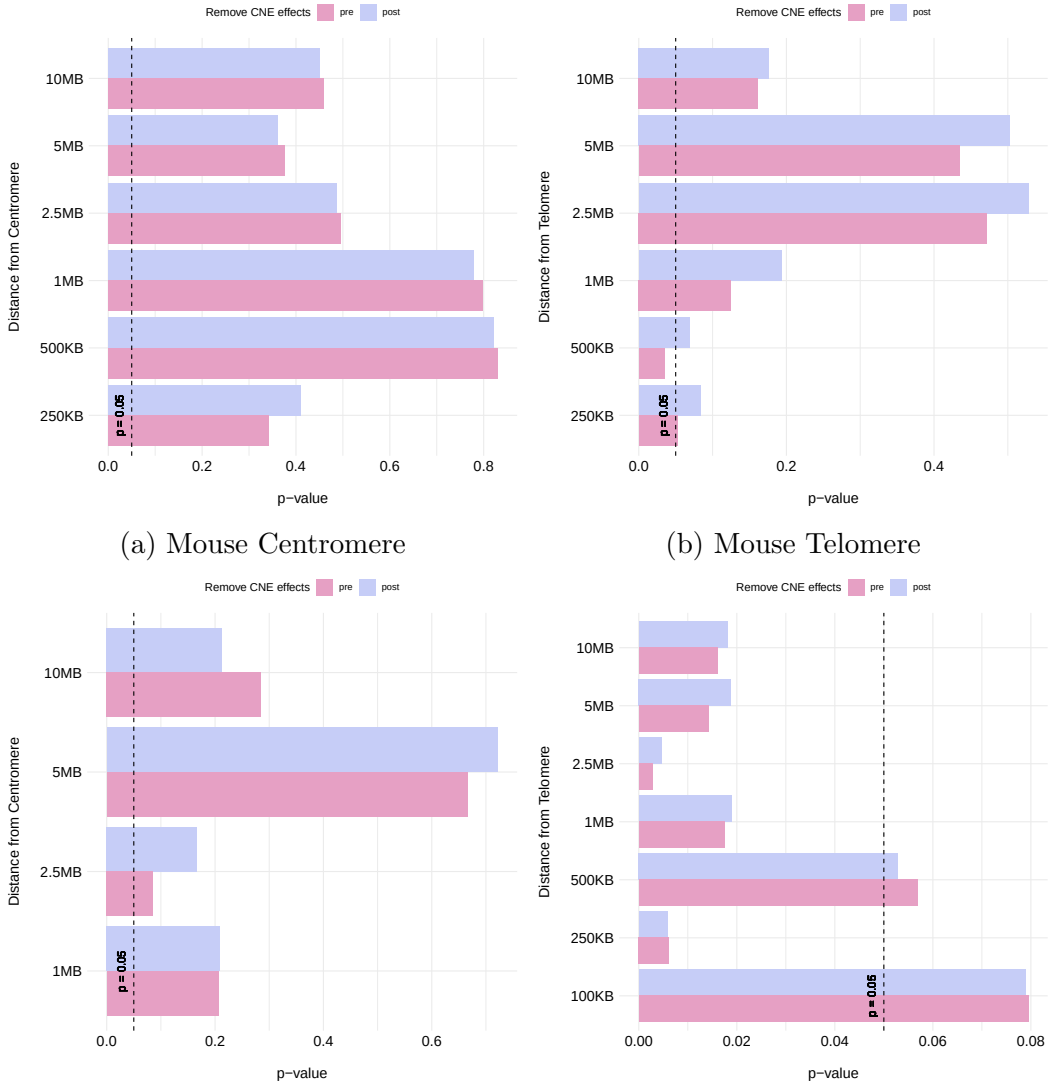
Materials & Methodology



Comparing Gene Expression Divergence pre and post removing CNE effect:

The following bar plots depict the p-values obtained on comparing Gene Expression Divergence of the control set v/s the domain change set pre and post removing CNE effect.

For Expression Divergence calculated using ***Spearman*** *correlation:*



(a) Mouse Centromere

(b) Mouse Telomere

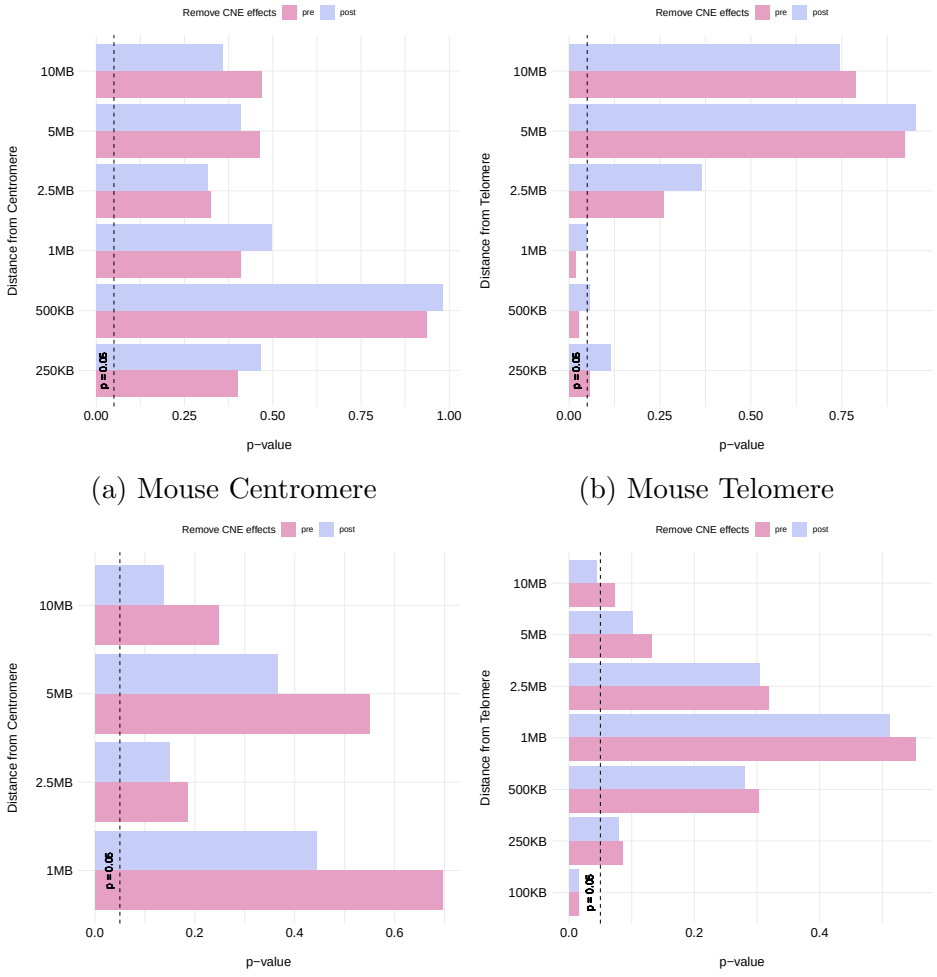
(c) Human Centromere

(d) Human Telomere

Figure 12: Comparing significance values of difference in Gene Expression Divergence (Spearman) pre and post removing CNE effect

On comparing the differential expression pre and post removing CNE effect for mouse telomere: 500KB from the mouse telomere initially contained a list of genes in "domain change" set that were significantly differentially expressed compared to the control set. However, upon controlling for the CNE effect, this significance has dampened into a non-significant value. In the case of genes constrained near Human Telomere, however, there is no change pre and post controlling for CNE effect in the distances that show significant differential expression on comparing "domain change" set with control set.

For Expression Divergence calculated using ***Pearson*** *correlation:*



(a) Mouse Centromere

(b) Mouse Telomere

(c) Human Centromere

(d) Human Telomere

Figure 13: Comparing significance values of difference in Gene Expression Divergence (Pearson) pre and post removing CNE effect

On comparing the differential expression pre and post removing CNE effect for mouse telomere:

1. For 1MB from the mouse telomere, gene expression divergence is significant both pre and post controlling for CNE effect.

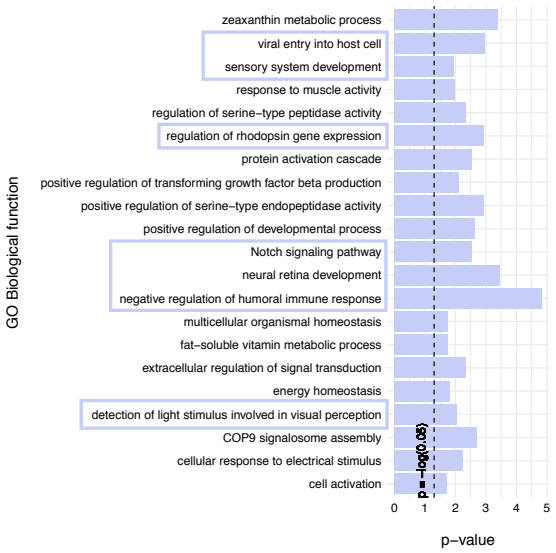
2. For 500KB from the mouse telomere, gene expression divergence is no longer significant after controlling for CNE effect.

No distances show significant expression divergence near Human Telomere (except 100KB from telomere), Human Centromere and Mouse Centromere both pre and post controlling for CNE effect when calculating expression divergence using Pearson correlation.

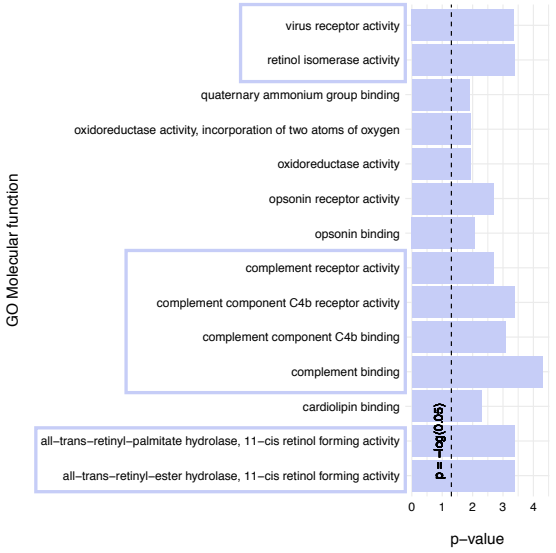
3.2.5 Gene Ontologies

Gene ontologies were obtained using *ToppGene*.

Within 500KB of Mouse Telomere:



(a) GO Biological Terms



(b) GO Molecular Terms

Figure 14: Gene Ontology Terms for genes constrained within 500KB from the Mouse Telomere

Within 1MB of Mouse Telomere

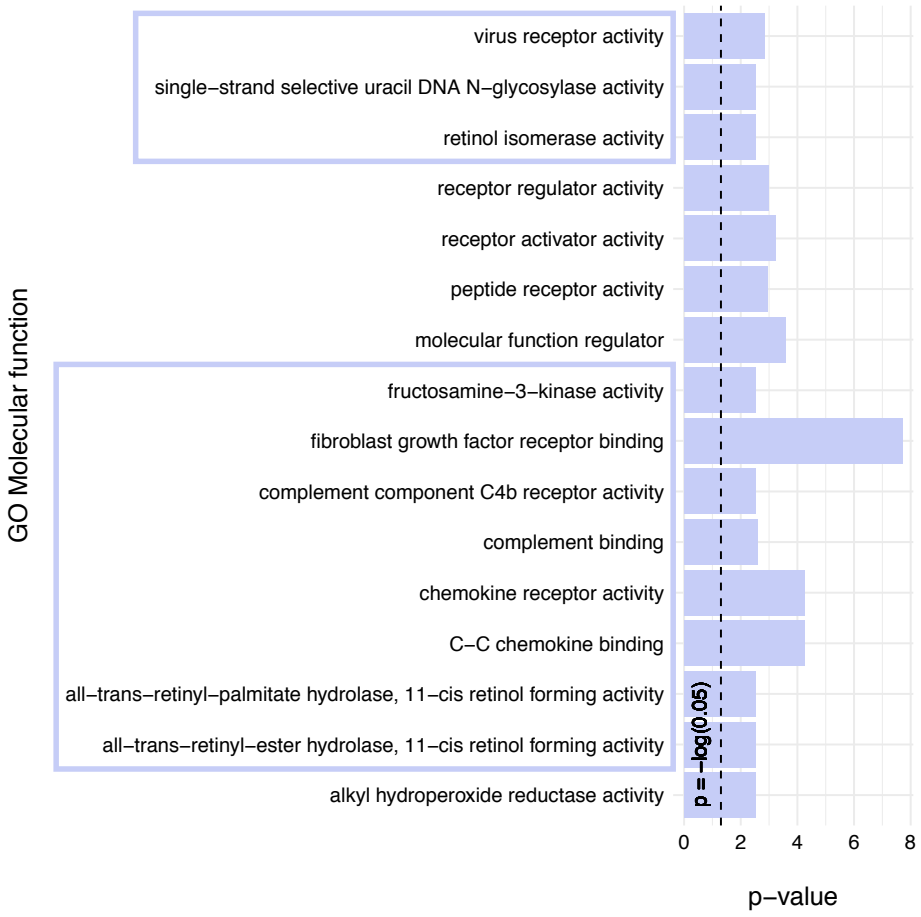


Figure 15: GO Molecular Terms for genes constrained within 1MB from the Mouse Telomere

Table 4: GO Biological Terms for genes constrained within 1MB of Mouse Telomere

ID	GO Term	log p-value
GO:0008543	fibroblast growth factor receptor signaling pathway	-4.6498
GO:0016477	cell migration	-3.6126
GO:0071634	regulation of transforming growth factor beta production	-3.8297
GO:0051094	positive regulation of developmental process	-4.5686

Human Telomere:

Within 250KB from the Human Telomere

Table 5: GO Molecular Terms for genes constrained within 250KB of Human Telomere

ID	GO Term	log p-value
GO:0003905	alkylbase DNA N-glycosylase activity	-3.1079

Within 2.5MB from the Human Telomere

Table 6: GO Molecular Terms for genes constrained within 2.5MB of Human Telomere

ID	GO Term	log p-value
GO:0010498	proteasomal protein catabolic process	-4.6517
GO:0038202	TORC1 signaling	-4.2118
GO:0016567	protein ubiquitination	-4.1427

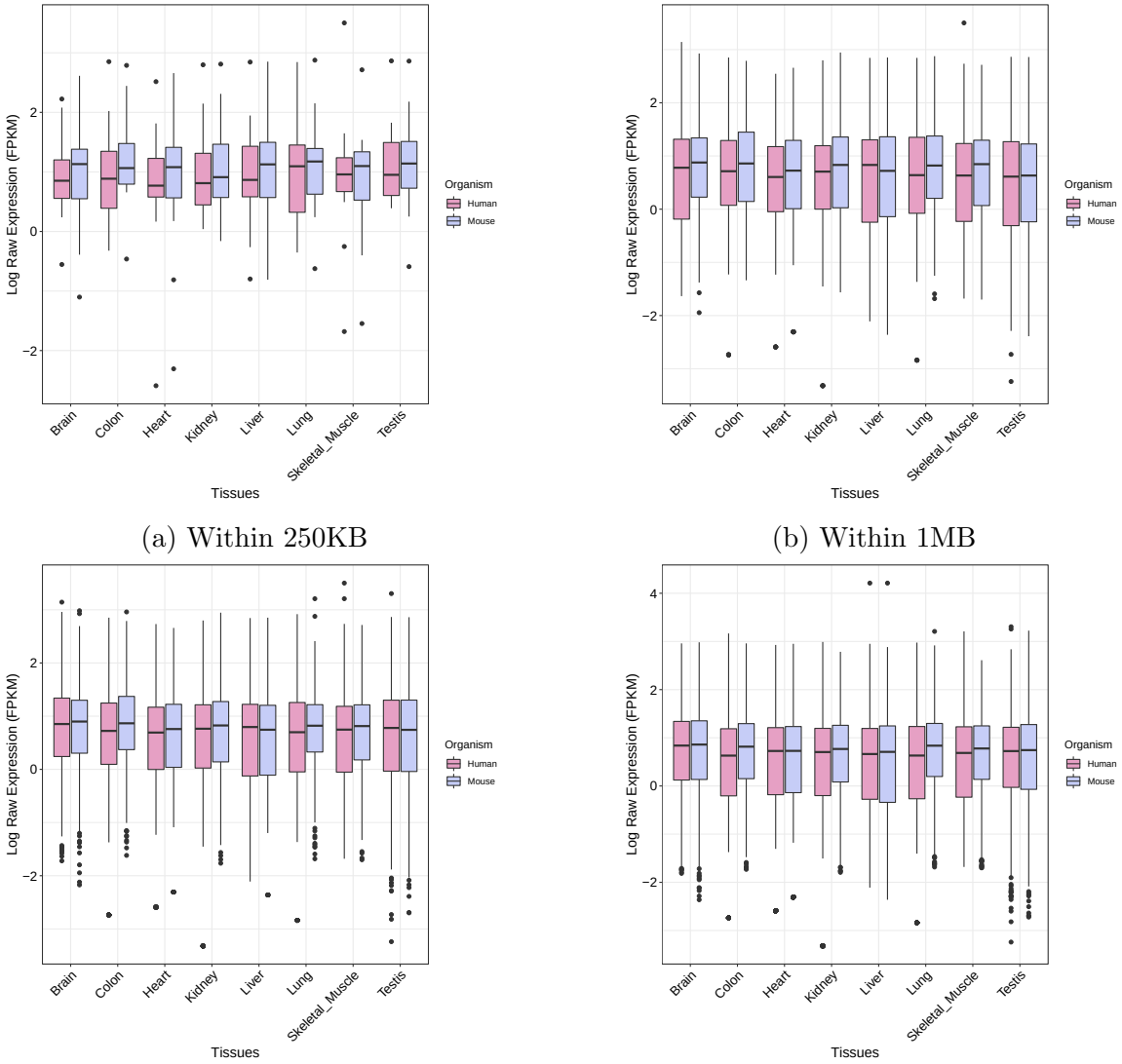
Within 10MB from the Human Telomere

Table 7: GO Human Phenotype Terms for genes constrained within 10MB of Human Telomere

ID	GO Term	log p-value
HP:0002977	Aplasia/Hypoplasia involving the central nervous system	-4.529
HP:0001104	Macular hypoplasia	-4.228
HP:0007364	Aplasia/Hypoplasia of the cerebrum	-4.212
HP:0010767	Sacrococcygeal pilonidal abnormality	-4.180
HP:0000960	Sacral dimple	-4.242

3.2.6 Raw Expression Comparison

Raw Expression Comparison across Human and Mouse tissues for genes constrained towards Human and Mouse Telomere:



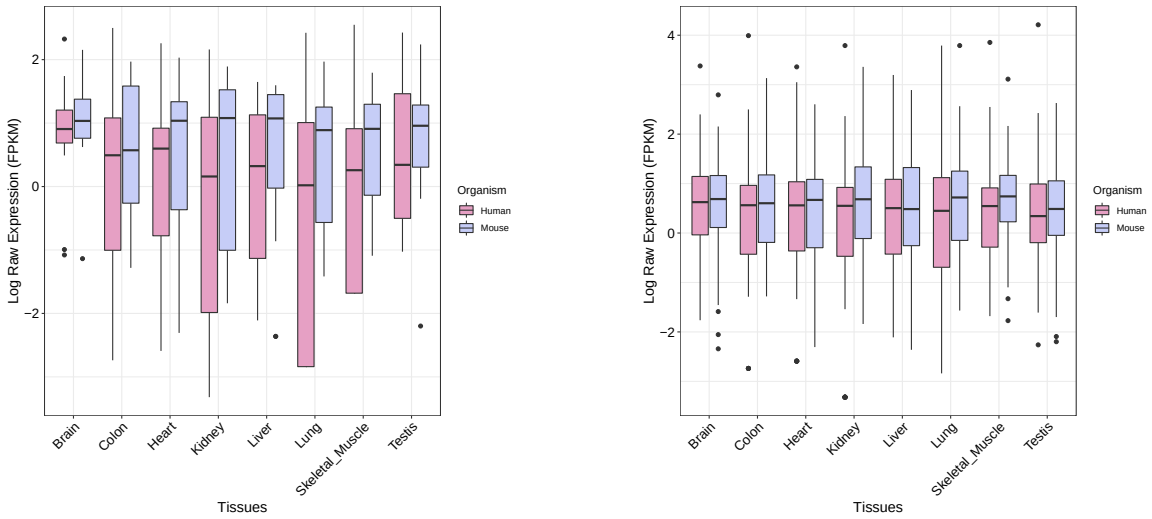
(a) Within 250KB

(b) Within 1MB

(c) Within 2.5MB

(d) Within 10MB

Figure 16: Log Raw Expression values at various distances from Human Telomere.



(a) Within 500KB

(b) Within 1MB

Figure 17: Log Raw Expression values at various distances from Mouse Telomere.

By plotting the Raw Expression of genes near telomeric regions, we expected to see the following trend:

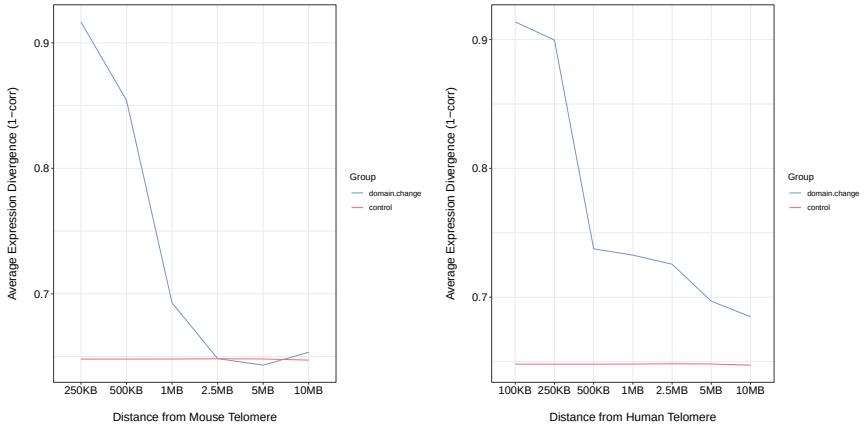
1. If genes are constrained specifically near the Human telomere, then we expected the log raw expression in human tissues for these genes to be lower than their mouse counterparts.

2. If genes are constrained specifically near the Mouse telomere, then we expected the log raw expression in mouse tissues for these genes to be lower than their human counterparts.

Although human tissues show lower expression for genes constrained within various distances from the Human telomere, it appears that so is not the case for genes constrained towards the Mouse Telomere, which show higher expression in mouse tissues compared to human tissues.

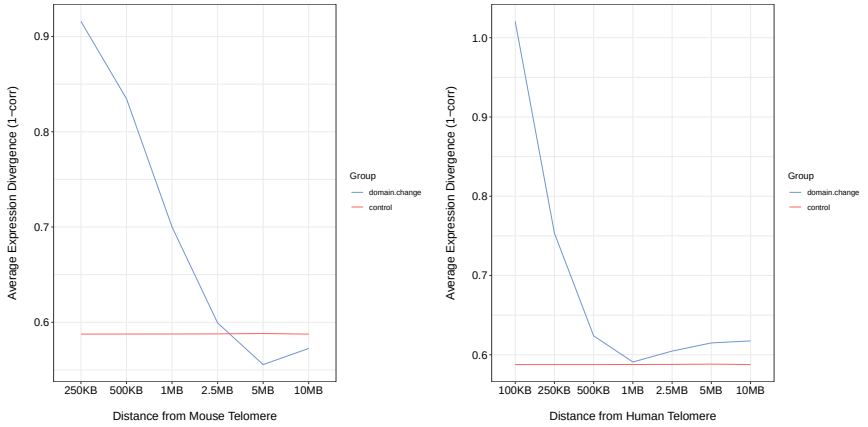
3.2.7 Conclusion

There appears to be a pattern in which the expression divergence decreases along the increasing distances from the telomere in both species.



(a) Mouse Telomere (Spearman)

(b) Hum Telomere (Spearman)



(c) Mouse Telomere (Pearson)

(d) Human Telomere (Pearson)

Figure 18: Average Expression Divergence v/s Distance from Telomere (control: red, domain change: blue)

Differential expression divergence does not appear to be significantly higher for the "domain change" set at any distance from the centromere in both Human and Mouse. Genes constrained within 500KB and 1 MB of the mouse telomere show significant gene expression divergence as compared to the control set (with some discrepancy in Pearson and Spearman correlation derived gene expression divergence). Genes constrained within 100KB, 250KB, 1MB, 2.5MB, 5MB and 10MB of the human telomere show significantly higher gene expression divergence as compared to the control set (Spearman correlation).

On controlling for CNE effect, it appears some gene expression divergence can be attributed to CNE effect. In the case of mouse telomere, for genes constrained within distances of 500KB of it (for both Spearman and Pearson correlation), the domain change set does not remain significantly divergent after controlling for CNE-effect. However, no such change is observed for genes constrained within 1MB of the mouse telomere. It remains significant before and after controlling for CNE-effect (Pearson correlation). These genes are involved in functions related to chemokine receptors and pathways in cancer and immune responses. Humans and mice have clearly shown some differences in terms of immune responses and in terms of their ability to ward off cancer. For instance, naturally occurring ebolaviruses do not seem to infect mice and cancers rarely develop in mice over their short lifespan (Bjornson-Hooper et al., 2019). The differences in human-mouse immune system have been extensively studied, often leading to the conclusion that mice are not a good substitute for studying a (human) disease because of the far too many differences in the responses launched (Zschaler, Schlorke, & Arnhold, 2014). Since survival of a species includes a strong immune system as a necessity apart from adaptations to its environment, it will be very interesting to understand how the genes that are constrained at the mouse telomere, which seem to have a role in viral response, cancer and complement cascade, shape the immune system of mice in comparison to humans.

In case of genes constrained within various distances from the human telomere, post removing CNE effect, all the distance cutoffs that showed significant divergence, remain significant. Gene ontologies for genes constrained within 250KB indicate a role in immune response and DNA repair. Ontologies for genes constrained within 2.5MB of human telomere indicate towards TORC-1 signalling, which has a role in controlling cell growth and metabolism. Partial inhibition of mTORC-1 leads to a significant increase in lifespan in worms and flies, possibly via mechanisms similar to calorie-restriction. Such a link between longevity and reduced mTORC-1 expression has not been established in mammals. Moreover, mTORC-1 also has implications in regulating immune response, and its deletion leads to tumorigenesis (Wullschleger, Loewith, & Hall, 2006). The Human phenotype terms for genes constrained within 10MB of the telomere have implications in underdeveloped cerebrum, which does not abide very well by our hypothesis.

Although some really interesting genes are constrained specifically towards mouse telomere, the inconsistencies in raw expression patterns and in genes constrained at human telomere, suggest that a look at heterochromatin effect by considering 3D genome organization would prove more comprehensive.

3.3 Predicting Centromeres

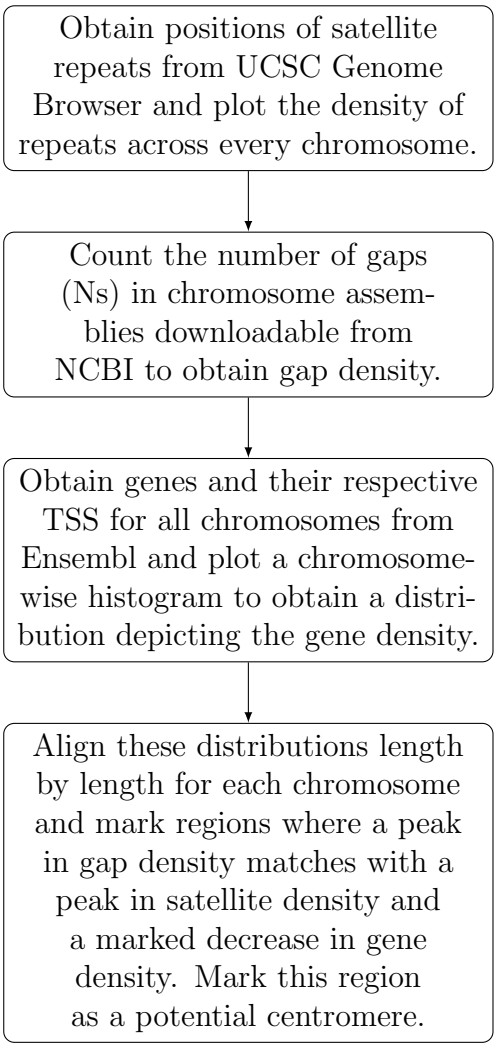
3.3.1 Introduction

Human-Mouse rearrangements that constrain genes near heterochromatin show some interesting phenotypes. Albeit the fact that the theory is not consistent, it provides some direction to carry the study forward and test the same hypothesis with some other species that have distinctly diverged from Humans and have interesting adaptations to their environments. Moreover, if a greater number of species can be involved in the study, the analysis becomes more robust. However, the problem lies in the fact that although many mammals have chromosome level assemblies and decently annotated orthologous genes, several of them lack properly annotated centromere positions. Without these centromere positions, it becomes difficult to filter genes into heterochromatin-associated set and control set. Therefore, as an exercise, we tried to see if the prediction of centromeres on the basis of some known centromeric signatures from genome assemblies was possible at all.

One signature that is commonly attributed to centromeres is the high density of satellite repeats that often makes it challenging to sequence these regions. These satellite repeats can range from a few base pairs to nearly several megabases in length (Hartley & O'Neill, 2019). These satellite repeats vary across taxa in terms of both sequence diversity and also the percentage of the genome that these repeats occupy. However, most of satellite DNA is concentrated near the centromere (pericentromeric region) or in the centromere (Garrido-Ramos, 2017; Jagannathan, Cummings, & Yamashita, 2018). Therefore this serves as an excellent marker for elucidating the position of the centromere. Another signature that could be used to predict centromere positions is gene density: heterochromatin is generally gene-poor, and a dip in gene density can be viewed as a potential centromere.

Apart from these markers, it is often noted that chromosome assemblies annotate repetitive regions with "N", which is considered as a gap, as the nucleotide in the said position is unknown or unmapped. Such large gaps have been used to demarcate centromere positions in the Human chromosome assemblies previously (Altemose, Miga, Maggioni, & Willard, 2014). Therefore, large gaps in chromosome assemblies can be aligned with the other two signatures, and a match in the positions of these three factors (gap, gene and satellite density) can help predict centromeric positions.

3.3.2 Materials & Methodology



3.3.3 Results

Opossum centromeres and Rabbit centromeres were predicted via the use of the aforementioned markers.

An example of predicting centromere on the basis of Gap & Gene density for Opossum Chromsome 5:

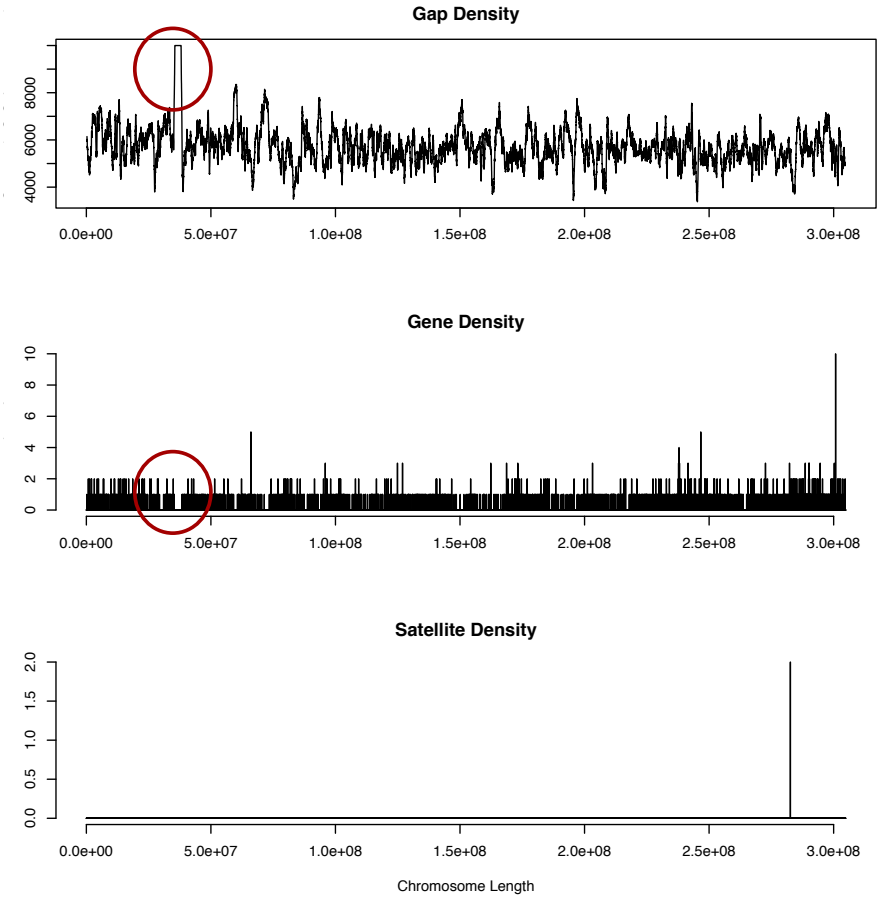
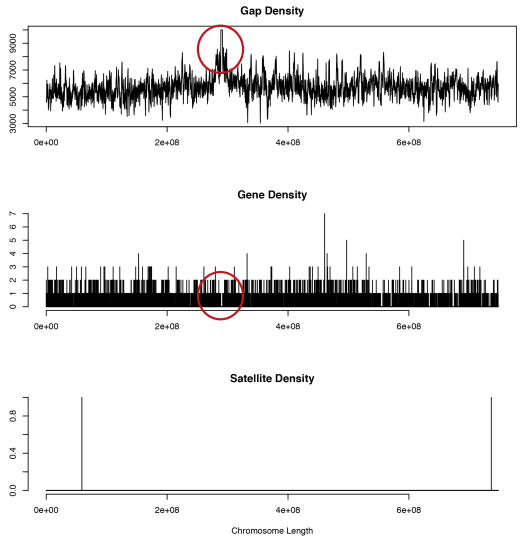
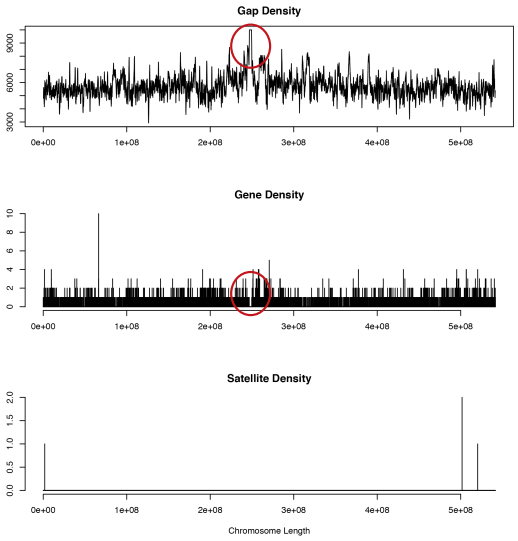


Figure 19: Opossum Chromosome 5

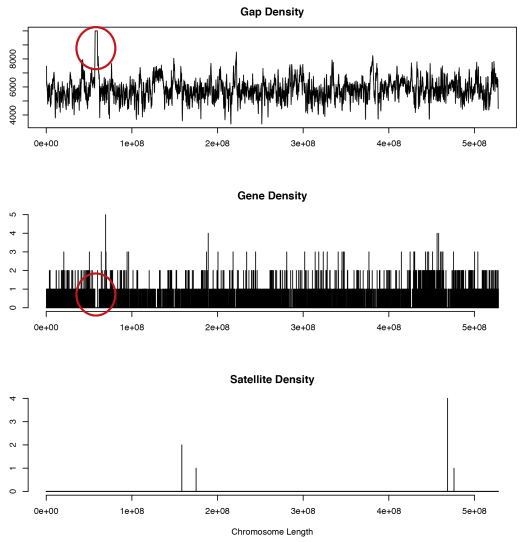
Predicted centromere positions were matched with karyotype/ideogram based predictions of centromere positions available in literature (Pathak, Rønne, Brown, Furlong, & VandeBerg, 1993; Hayes et al., 2002)



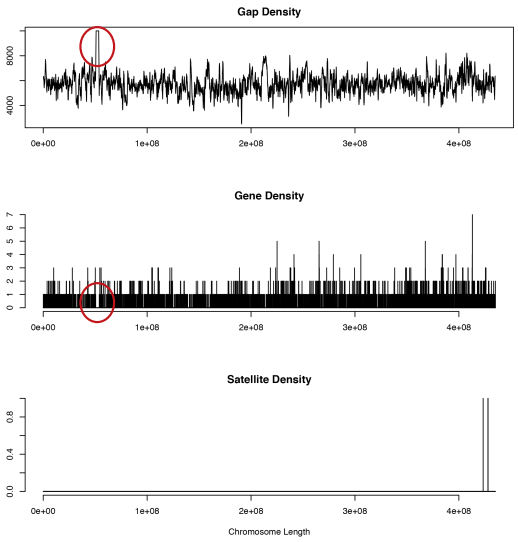
(a) Opossum Chromosome 1



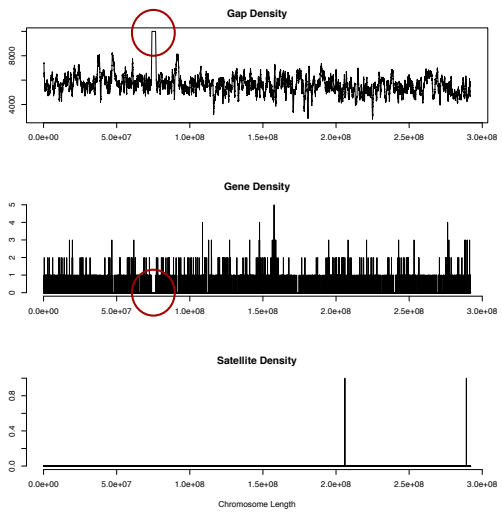
(b) Opossum Chromosome 2



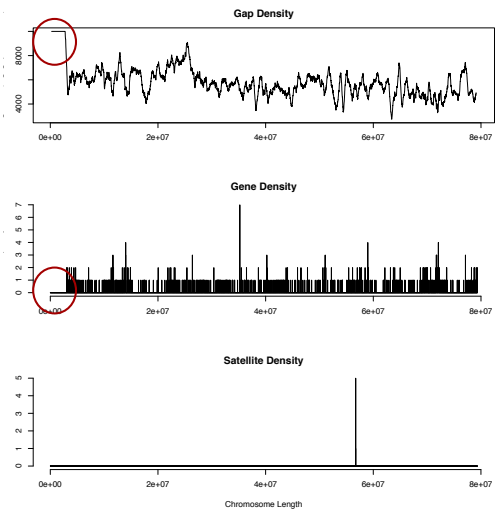
(c) Opossum Chromosome 3



(d) Opossum Chromosome 4

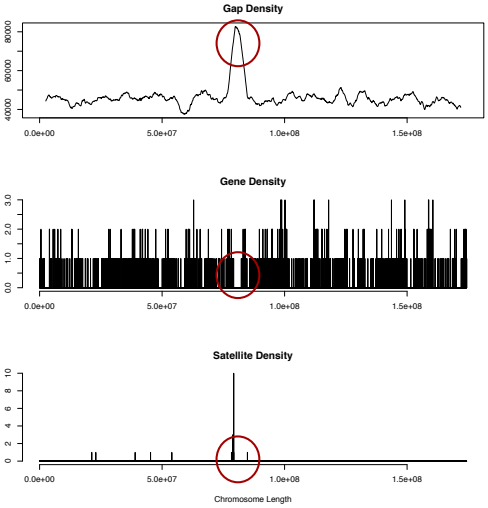


(e) Opossum Chromosome 6

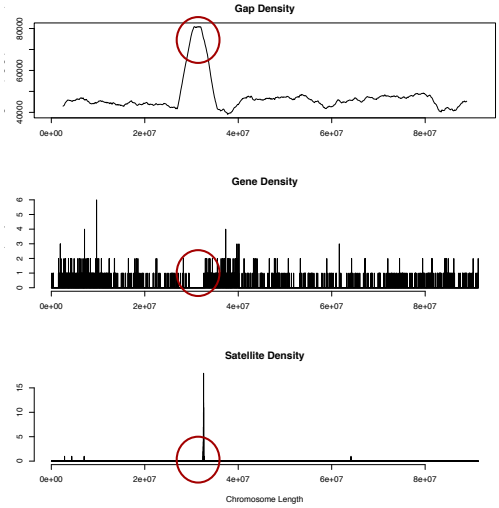


(f) Opossum Chromosome X

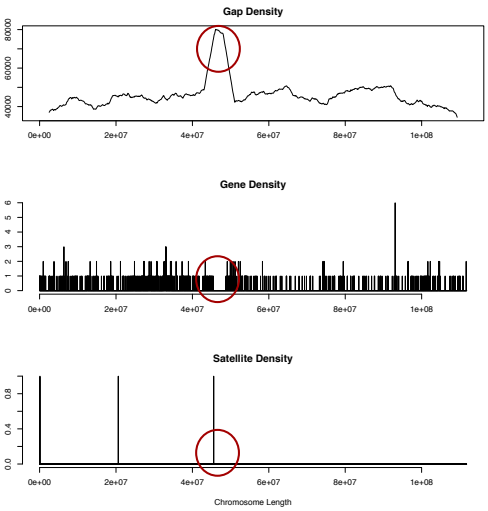
Figure 20: Some More Examples of Predicting Centromeres using Gene, Gap & Satellite Density for Opossum Chromosomes.



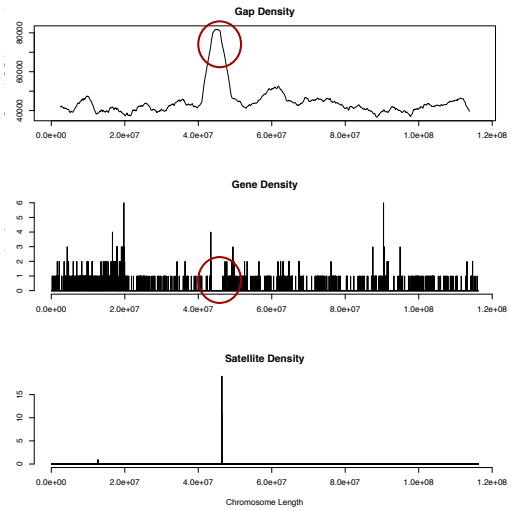
(a) Rabbit Chromosome 2



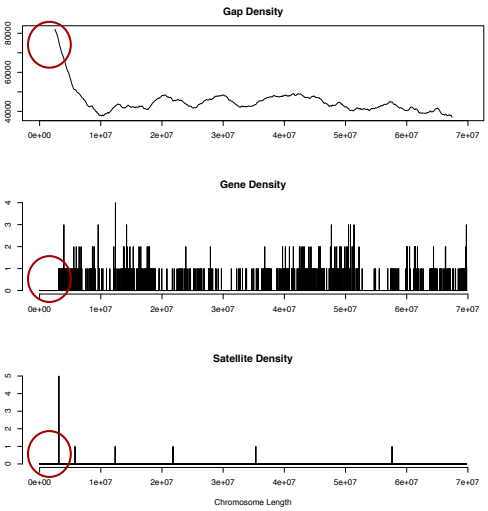
(b) Rabbit Chromosome 4



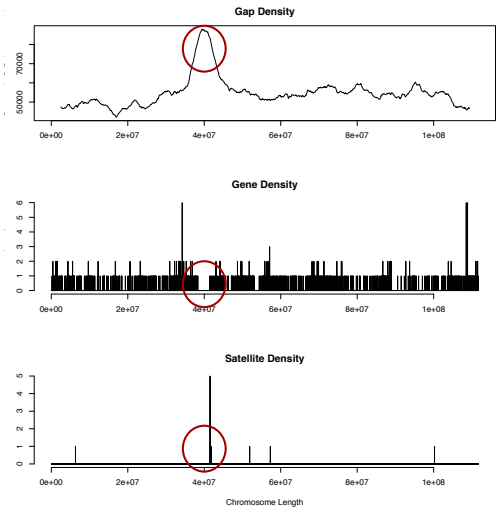
(c) Rabbit Chromosome 8



(d) Rabbit Chromosome 9



(e) Rabbit Chromosome 18



(f) Rabbit Chromosome X

Figure 21: Some Examples of Predicting Centromeres using Gene, Gap & Satellite Density for Rabbit Chromosomes.

Table 8: Predicted Opossum Centromeres

Chromosome	Start	End	Length
1	288708214	291708213	2999999
2	246760696	249761394	3000698
3	56853872	59859752	3005880
4	50630176	53631147	3000971
5	35275970	38275969	2999999
6	73847710	76847709	2999999
7	42498444	45499359	3000915
8	0	2999999	2999999
X	0	3001818	3001818

Table 9: Predicted Rabbit Centromeres

Chromosome	Start	End	Length
1	79861450	82861962	3000512
2	79319264	82319263	999999
3	58762725	61767528	3004803
4	29452887	32455898	3003011
5	4293070	7293069	2999999
6	19455484	22455483	2999999
7	49768112	52769499	3001387
8	45576146	48579106	3002960
9	43425350	46425349	2999999
10	0	3000334	3000334
11	32228397	35229168	3000771
12	17046319	20046318	2999999
13	17056091	20056396	3000305
14	24374029	27374028	2999999
15	0	2999999	2999999
16	7570854	10571144	3000290
17	927914	3927913	2999999
18	0	2999999	2999999
19	0	3000966	3000966
20	0	2999999	2999999
21	0	2999999	2999999
X	38416591	41416714	3000123

3.3.4 Conclusion

The approach of using gap density, satellite density and gene density as markers for centromere position seems to have been fruitful for the chromosome assemblies of Opossum and Rabbit. However, the method has failed largely for some other species such as Rat, Dog, Horse, etc.

It is clear that the factors this approach utilizes is dependent on the quality of the genome assembly and the amount of available information associated with the genome, that is, how far the genome of a species has been studied or has impacted research.

Therefore, this methodology is not robust and needs to be supplemented by factors that are not assembly quality dependent.

It would, however, be very interesting to test our hypothesis once more when more species have maximally annotated orthologous genes and centromere positions.

4 Discussion & Future Directions

Some interesting results obtained via different methods match with our hypothesis and some don't. For instance, we expected regions near telomere and centromere to be enriched with genes that when repressed translate into phenotypes that are specific to that species. However, genes which when down-regulated lead to aplasia of the cerebrum or lissencephaly (smooth brain syndrome) show up near human telomere (albeit within large distances such as 10MB over which heterochromatin may have no influence), which is hard to reconcile with the fact that humans are the ones with gyrated brains (Lewitus, Kelava, Kalinka, Tomancak, & Huttner, 2014). The other problem that arises from studying just the linear order of genes near the telomere or centromere is highlighted by the fact that regions that are near each other in the linear genome may not be close to each other in the 3D genome. The same holds true for the opposite too (Naumova & Dekker, 2010). A region far away from the telomere linearly on a mouse chromosome may be close enough to the telomere to be influenced by the repressive nature of heterochromatin in a 3D environment. Therefore, we come to an understanding that studying the linear rearrangement of genes near the telomere/centromere alone is not enough. For a more refined analysis, we propose to look at linear rearrangements that disrupt the 3D organization of the genome.

It is known that the genome is organized into *cLAD*: constitutive Lamina Associated Domains (that are associated with the nuclear lamina and are always repressed), *ciLADs*: constitutive inter-Lamina Associated Domains (that are always expressed) and *fLADs:* facultative LADs (that change nuclear organization and expression pattern depending upon cell type). These Lamina-Associated Domains are implicated in influencing genome organization (Briand & Collas, 2020). When these domains come into the picture, a generalized view of looking at the regions far from the telomere and centromere linearly on the chromosome as an expressed compartment is transformed into a more specific view where this broad, expressed compartment can now be divided into regions of cLADs and ciLADs. Moreover, it is known that very few rearrangements go from a cLAD to a ciLAD or the other way round in Human-Mouse comparison studies (Meuleman et al., 2013). It is also known that pericentromeric regions are often associated with nuclear lamina (Scherthan et al., 2000; Van Steensel & Belmont, 2017). Since not much was determined about the influence of centromere on gene expression divergence via studies of linear proximity of genes to the centromere, one can look at the effect of rearrangements that disrupt the organization of cLAD and ciLADs on gene expression for a more comprehensive understanding.

References

Altemose, N., Miga, K. H., Maggioni, M., & Willard, H. F. (2014). Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS computational biology*, *10*(5).

Bagadia, M., Chandradoss, K. R., Jain, Y., Singh, H., Lal, M., & Sandhu, K. S. (2019). Evolutionary loss of genomic proximity to conserved noncoding elements impacted the gene expression dynamics during mammalian brain development. *Genetics*, *211*(4), 1239–1254.

Bjornson-Hooper, Z. B., Fragiadakis, G. K., Spitzer, M. H., Madhireddy, D., McIlwain, D., & Nolan, G. P. (2019). A comprehensive atlas of immunological differences between humans, mice and non-human primates. *bioRxiv*, 574160.

Briand, N., & Collas, P. (2020). Lamina-associated domains: peripheral matters and internal affairs. *Genome biology*, *21*(1), 1–25.

Brown, D. D. (1981). Gene expression in eukaryotes. *Science*, *211*(4483), 667–674.

De, S., & Babu, M. M. (2010). Genomic neighbourhood and the regulation of gene expression. *Current opinion in cell biology*, *22*(3), 326–333.

De, S., Teichmann, S. A., & Babu, M. M. (2009). The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome. *Genome research*, *19*(5), 785–794.

Dong, X., & Weng, Z. (2013). The correlation between histone modifications and gene expression. *Epigenomics*, *5*(2), 113–116.

Elgin, S. C., & Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in drosophila. *Cold Spring Harbor perspectives in biology*, *5*(8), a017780.

Eres, I. E., Luo, K., Hsiao, C. J., Blake, L. E., & Gilad, Y. (2019). Reorganization of 3d genome structure may contribute to gene regulatory evolution in primates. *PLoS genetics*, *15*(7), e1008278.

Finn, E. H., & Misteli, T. (2019). Molecular basis and biological function of variability in spatial genome organization. *Science*, *365*(6457), eaaw9498.

Garrido-Ramos, M. A. (2017). Satellite dna: An evolving topic. *Genes*, *8*(9), 230.

Gilad, Y., Man, O., Pääbo, S., & Lancet, D. (2003). Human specific loss of olfactory receptor genes. *Proceedings of the National Academy of Sciences*, *100*(6), 3324–3327.

Hartley, G., & O'Neill, R. J. (2019). Centromere repeats: hidden gems of the genome. *Genes*, *10*(3), 223.

Hasty, P., & Montagna, C. (2014). Chromosomal rearrangements in cancer: detection and potential causal mechanisms. *Molecular & cellular oncology*, *1*(1), e29904.

Hayes, H., Rogel-Gaillard, C., Zijlstra, C., De Haan, N., Urien, C., Bourgeaux, N., ... Bosma, A. (2002). Establishment of an r-banded rabbit karyotype nomenclature

by fish localization of 23 chromosome-specific genes on both g-and r-banded chromosomes. *Cytogenetic and genome research*, *98*(2-3), 199–205.

Jagannathan, M., Cummings, R., & Yamashita, Y. M. (2018). A conserved function for pericentromeric satellite dna. *Elife*, *7*, e34122.

Jordan, I. K., Mariño-Ramírez, L., & Koonin, E. V. (2005). Evolutionary significance of gene expression divergence. *Gene*, *345*(1), 119–126.

Jost, K. L., Bertulat, B., & Cardoso, M. C. (2012). Heterochromatin and gene positioning: inside, outside, any side? *Chromosoma*, *121*(6), 555–563.

Kleinjan, D. A., & van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *The American Journal of Human Genetics*, *76*(1), 8–32.

Kleinjan, D.-J., & Van Heyningen, V. (1998). Position effect in human genetic disease. *Human molecular genetics*, *7*(10), 1611–1618.

Lewitus, E., Kelava, I., Kalinka, A. T., Tomancak, P., & Huttner, W. B. (2014). An adaptive threshold in mammalian neocortical evolution. *PLoS biology*, *12*(11).

Lopez, M. D., Guerra, J. J. M., & Samuelsson, T. (2010). Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. *PloS one*, *5*(5).

Mallo, M., & Alonso, C. R. (2013). The regulation of hox gene expression during animal development. *Development*, *140*(19), 3951–3963.

Mann, R. S. (1997). Why are hox genes clustered? *Bioessays*, *19*(8), 661–664.

Marques-Bonet, T., Sànchez-Ruiz, J., Armengol, L., Khaja, R., Bertranpetit, J., Lopez-Bigas, N., ... Navarro, A. (2007). On the association between chromosomal rearrangements and genic evolution in humans and chimpanzees. *Genome biology*, *8*(10), R230.

Meuleman, W., Peric-Hupkes, D., Kind, J., Beaudry, J.-B., Pagie, L., Kellis, M., ... van Steensel, B. (2013). Constitutive nuclear lamina–genome interactions are highly conserved and associated with a/t-rich sequence. *Genome research*, *23*(2), 270–280.

Muller, H. J. (1941). Induced mutations in drosophila. In *Cold spring harbor symposia on quantitative biology* (Vol. 9, pp. 151–167).

Naumova, N., & Dekker, J. (2010). Integrating one-dimensional and three-dimensional maps of genomes. *Journal of cell science*, *123*(12), 1979–1988.

Nguyen, H. Q., & Bosco, G. (2015). Gene positioning effects on expression in eukaryotes. *Annual review of genetics*, *49*, 627–646.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., ... others (2012). Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, *485*(7398), 381–385.

Pathak, S., Rønne, M., Brown, N., Furlong, C., & VandeBerg, J. (1993). A high-resolution banding pattern idiogram of monodelphis domestica chromosomes

(marsupialia, mammalia). *Cytogenetic and Genome Research*, *63*(3), 181–184.

Pedram, M., Sprung, C. N., Gao, Q., Lo, A. W., Reynolds, G. E., & Murnane, J. P. (2006). Telomere position effect and silencing of transgenes near telomeres in the mouse. *Molecular and cellular biology*, *26*(5), 1865–1878.

Scherthan, H., Jerratsch, M., Li, B., Smith, S., Hultén, M., Lock, T., & de Lange, T. (2000). Mammalian meiotic telomeres: protein composition and redistribution in relation to nuclear pores. *Molecular biology of the cell*, *11*(12), 4189–4203.

Shapiro, J. A. (1982). Changes in gene order and gene expression. *Natl. Cancer Inst. Monograph*, *60*, 87–110.

Van Steensel, B., & Belmont, A. S. (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell*, *169*(5), 780–791.

Voigt, P., Tee, W.-W., & Reinberg, D. (2013). A double take on bivalent promoters. *Genes & development*, *27*(12), 1318–1338.

Wullschleger, S., Loewith, R., & Hall, M. N. (2006). Tor signaling in growth and metabolism. *Cell*, *124*(3), 471–484.

Zschaler, J., Schlorke, D., & Arnhold, J. (2014). Differences in innate immune response between man and mouse. *Critical Reviews^{TM} in Immunology*, *34*(5).