Comparative Genomic Analysis of Human HADHB and WNT5A Gene Containing Loci to Elucidate their Gene Regulatory Networks

ومن يؤت الحكمة
فقد أوتي خيرا كثيرا

QUAID-I-AZAM UNIVERSITY
ISLAMABAD

By

Arshiya Qayyum

Reg # 04281613023

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-I-Azam, University

Islamabad- Pakistan

2020

Comparative Genomic Analysis of Human HADHB and WNT5A Gene Containing Loci to Elucidate their Gene Regulatory Networks

QUAID-I-AZAM UNIVERSITY
ISLAMABAD

By

Arshiya Qayyum

A thesis submitted in the fulfillment of the requirements for the degree
of
BACHELOR OF SCIENCE

IN
BIOINFORMATICS

National Center for Bioinformatics

Faculty of Biological Sciences

Quaid-I-Azam, University

Islamabad- Pakistan

2020

Declaration

The work reported in this thesis was carried out by **Arshiya Qayyum** and I hereby declare that the title of the thesis, **"Comparative Genomic Analysis of Human HADHB and WNT5A Gene Containing Loci to Elucidate their Gene Regulatory Networks**" and the contents of thesis are the product of my own research and no part has been copied from any published source (except the references, standard mathematical or genetic models /equations /formulas /protocols, etc.). I further declare that this work has not been submitted for the award of any other degree /diploma. The University may take action if the information provided is found inaccurate at any stage.

Scholar Signature __________________

CERTIFICATE

This report, submitted by **Miss Arshiya Qayyum** from National Centre for Bioinformatics the ***Undergraduate Program)***, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan, is accepted in its present form as satisfying the requirement of **Research Project (BIF222)** for the Degree of **BS-Bioinformatics.**

Supervisor: ______________________

Dr. Amir Ali Abbasi
Associate Professor
National Centre for Bioinformatics
Quaid-i-Azam University, Islamabad

Internal Examiner: ______________________

Dr. Syed Sikander Azam
Associate Professor
National Centre for Bioinformatics
Quaid-i-Azam University, Islamabad

Chairman: ______________________

Dr. Amir Ali Abbasi
Associate Professor
National Centre for Bioinformatics
Quaid-i-Azam University, Islamabad

Dated: 27 August,2020

DEDICATED

To

MY PARENTS

All that I am, or hope to be, I owe it to my mother.

ACKNOWLEDGMENTS

In the name of Allah, the Most Beneficent and the most Merciful, Alhamdulillah. First and foremost, sincere thanks to **Allah** for all the things He has both given and denied me for only He knows what's best for me. Gratefully, joyously, and relentlessly, I commend His splendor and generosity, my **Al-Basir**; beyond all beautiful dreams and hopes.

I would like to express my deep gratitude to ***Dr. Amir Ali Abbasi,*** my research supervisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work. His support has been a great source of motivation for me.

I wish to express my sincere appreciation to my teacher, ***Miss Huma Shireen*** who has the substance of a genius: she convincingly guided and encouraged me to be professional and do the right thing even when the road got tough. Without her persistent help, the goal of this project would not have been achieved.

Bundle of thanks to my MPhil lab seniors ***Washaakh Ahmad*** and ***Hadia Haider Abbasi*** for their continuous guidance throughout all the thick and thins of my research project and who were always ready to help by keeping their work aside.

I thank my friend and research colleague ***Haseeb Manzoor*** for supporting me throughout all my studies. He always kept me going on my path to success, assisting me as per his abilities and always ensuring me good times keep flowing. One simply could not wish for a better companion.

Last but not the least, my deepest gratitude goes to ***beloved family and friends*** for their endless support, prayers and encouragements all the way since the beginning of my studies.

Arshiya Qayyum

LIST OF ABBREVIATIONS

CREs	Cis Regulatory Elements
TF	Transcription Factor
TFBS	Transcription Factor Binding Sites
TSS	Transcription Start Site
INR	Initiator element
DPE	Downstream Promoter Element
MTE	Motif ten element
BRE	TFIIB Recognition Element
MGI	Mouse Genome Informatics
GXD	Gene eXpression Database
GWAS	Genome-wide Association Study
CMT	Charcot-Marie-Tooth
SNPs	Single Nucleotide Polymorphisms
BLAST	Basic Local Alignment Search Tool
ETG	Enhancer-target genes
CNE	Conserved Non-coding Elements
UCSC	University of California, Santa Cruz
HSA	Homo Sapiens

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In developmental biology, the limb is an extensively used model organ. Prominent variances in the morphological structure of limbs among different species are observed, even though the signaling paths involved during the limb development are mainly conserved. This reveals that gene regulation plays a major role in limb patterning. The process of genome-wide identification of the genes and their regulatory elements, which are involved in the development and growth of limbs has been alleviated by the latest advancements in the field of genomics.

Variations in gene expression that can direct to crucial limb morphological alterations in evolution, for example fin-to-limb reformation, loss of limbs in snakes and acquisition of wing in bats, are triggered by the modifications in the gene regulatory elements (mainly Cis-acting elements). Additionally, accumulation of evidence demonstrates that the disturbance of certain gene regulatory elements particularly enhancers leads to the production of unique limb deformities. Thus, it is evident from latest researches that the remote cis-regulatory elements mediate the spatiotemporal control of gene expression, which forms the basis of regular limb development.

Though enhancers perform a fundamental part in evolution and minor changes in the enhancer sequence or dosage may lead to functional or structural limb malformation, their detection continues to be elusive mainly due to complexities in accomplishing an extensive pairing of enhancers and target genes. Several different comparative genomics techniques, both wet and computational, have been in use for enhanced interpretation and understanding of the enhancer structure and function. Conclusively, one significant comparative genomics technique developed and used to analyze the conservation pattern of limb specific enhancers of human chromosome 2 and 3 and to link them to their target genes, is synteny mapping, which cleared route for better systematic study of their roles in evolution and diseases.

Chapter 1

INTRODUCTION

1. INTRODUCTION

The human skeletal framework provides the elementary sustenance and support to the internal structure of the body (Iscan et al.,2013). Normal adult axial skeleton comprises of 80 bones that takes the shape of head and body trunk. Joined to this axial skeleton is the appendicular skeleton, whose 126 bones constitute the girdles and the body limbs (Iscan et al., 2013). Limbs are the paired jointed appendages located on the body, categorized into two divisions, the upper limbs, attached to the upper body, and the lower limbs, appended to the lower body (Iscan et al.,2013).

Originally, all the animals including primates used both the upper and the lower limbs for locomotion and movement (Amis et al., 1990). Later, evolution brought about variations that allowed different environmental adaptabilities (Young et al., 2010). For instance, in man, due to its upright posture, several discrete functional necessities are reliant on the upper and lower limbs (Iscan et al.,2013). The lower limbs are fitted for mass-bearing support and stability, as well as for body locomotion like walking or running (Young et al., 2010). The upper limbs, which were formerly used for locomotion, are now extremely mobile and can be used for a widespread variety of activities and motion (Amis et al., 1990). The extensive range of upper limb actions, coupled with the capability to effortlessly maneuver things with our hands and opposable thumbs, has consented human beings to create the state of the art, modern world in which we live. Amongst the remaining animal kingdom, limbs are still principally used for movements, for instance walking, running, hopping, and climbing. Nevertheless, in some, the forelimbs are used for holding, mining, and manipulation. In birds, wings are the forelimbs that aid in flying, whereas in sea animals like the whale, they form the flippers that facilitate swimming. (Amis et al., 1990)

The developmental and functional pattern of limbs is significantly controlled by an increasing quantity of distinct genes (Tabin et al., 1991). This involvement of different but specific genes in limbs' functioning and growth is particularly implicated by their expression pattern (Tabin et al., 1991). For example, the spatial and temporal expression patterns of homeobox-genes, genes encoding receptors for developing anatomical signaling particles and genes encoding growth factors during normal limb development, with regard to numerous experimental embryological operations convincingly present the role of these specific genes in controlling limb morphogenesis (Tabin et al., 1991).

The human genome, reckoned as the blueprint of life, comprises of all the DNA sequences, both coding and non-coding, which regulate every single function in each and every cell type existing in our body (Shabalina et al., 2004). It is approximated that the human genome contains about 20,000 to 25,000 genes, that characterize only 2% of genomic sequence, while 98% of the genome is non-coding (Shabalina et al., 2004). A gene is the smaller, functional unit of genome, heritable in nature. They differ in size from a few hundred DNA nucleotide bases to more than 2 million bases, each occupying a specific position on one of the twenty-three pairs of chromosomes present inside the nucleus (Shabalina et al., 2004).

Genes encode proteins and these proteins dictate all the cell functions which regulate different biological activities in addition to controlling ribosomal RNAs and proteins (Alberts et al., 2013). The non-coding part of DNA comprise mainly of maintenance elements, for instance telomeres, centromeres and replication origins which are essential for DNA replication and repair control, and elements like promoters, insulators, enhancers, repressors, and regulatory RNAs which influence the spatial-temporal expression regulation of the coding genes (Alberts et al., 2013).

1.1 Gene Expression Regulation

Eukaryotic gene expression is a firmly regulated process. During cell differentiation and development, the expression of appropriate concentration of genes in the right cell type and in the exact stretch of time, in response to various internal and external stimuli is extremely crucial for normal functioning (Barrett et al., 2012).

With the progress in both genome-wide experimental methods and computational approaches in the post-sequencing genomics period, it is significant to study the interaction between various regulatory elements and the associated proteins, which control the pattern of gene expression at a single gene locus in addition to the global gene expression control across the genome within complex biological and transcriptional programmes (Barrett et al., 2012).

Regulatory control of gene expression can take place in several ways i.e. during transcription, processing of mRNA, and protein translation and at the protein stability level (Barrett et al., 2012). However, it is thought that gene regulation happens mostly at the transcriptional level. The eukaryotic transcriptional structure comprises of the cis-acting elements and the trans-acting elements, which are the two complimentary regulatory parts playing a vital role in gene expression regulation: (Venter et al., 2001).

1.2 Trans-acting and Cis-acting Regulatory Elements

The trans regulatory elements are basically transcription factors or other proteins which bind to DNA by identifying and then joining to certain DNA sequences in the cis regulatory elements to start, boost or repress the transcription of DNA (Hu et al., 2010). A single transcription factor may control many different genes, or many transcription factors may perform in a dense, combinatorial connected method, binding to the cis-acting elements at various binding sites for transcription factors, producing a huge range of specific and unique regulatory motifs (Hu et al., 2010). Approximately 1800 different transcription factors are encoded by the human genome. (Venter et al., 2001)

The cis-acting elements are regions of DNA sequences containing epigenetic information present either in the coding or non-coding region of the genome (Wittkopp et al., 2012). They create an accessible region in the DNA usually by chromatin remodeling and DNA or Histone modifications so that trans-factors can easily bind to the DNA and initiate transcription (Wittkopp et al., 2012). On the contrary, some of these Cis-acting elements generate inaccessible chromatin environments to prevent trans-factors from binding to the DNA, hence preventing transcription initiation (Wittkopp et al., 2012).

Cis-acting DNA sequences comprise of two discrete elements: the proximal elements or promoters and the distal regulatory sequences consisting of enhancers, silencers or repressors, insulators and locus control regions (LCRs) (Maston et al., 2006). All regulatory elements work in collaboration with each other to control the synchronized expression patterns of the gene (Maston et al., 2006). Cis regulatory elements are summarized below in Figure 1.1.

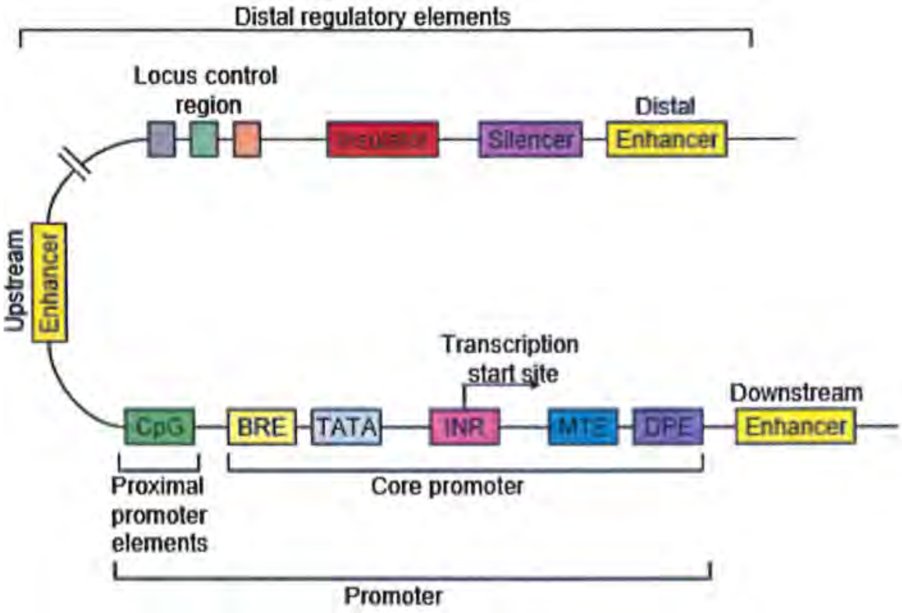

Figure 1.1: Cis-Regulatory Elements. A diagrammatic representation of the types of cis-acting regulatory elements i.e. promoters, enhancers, silencers, insulators and locus control regions participating in gene expression regulation.

1.2.1 Promoters

A typical promoter comprises of a core promoter and proximal promoter elements (Mastonet al., 2006). The core promoter is situated around 35 base pairs (bp) upstream or downstream of the transcription start site (TSS) and functions as the transcription factors binding site. The core promoter contains several elements i.e. a TATA box (TATA), an initiator element (INR), a downstream promoter element (DPE), a motif ten element (MTE) and a TFIIB recognition element (BRE). All these core elements commence the recruitment of TF to the promoter for transcription of the gene to take place (Mastonet al., 2006). The proximal promoter elements denote the DNA bases that lie upstream to the core promoter e.g. CpG islands (extending over about 1 kb around the TSS (transcription start site)) which are involved in altering the rate of transcription (Smale and Kadonaga, 2003).

1.2.2 Silencers

Silencing elements decrease the activities of the promoters, therefore inhibiting the transcription of associated genes in certain distinct cell types during stages in development (Mastonet al., 2006). Silencers are orientation and distance independent of the target gene. They may be present in the proximal promoter region, as an element of distal enhancer, or appear self-reliantly in distant areas, upstream or downstream of the gene which they are controlling (Mastonet al., 2006). Silencers facilitate repressions by combining with different repressor proteins. These repressors can function self-sufficiently, in collaboration with other repressor proteins (Sertil et al., 2003), or via the binding of a co-repressor (Chen and Evans, 1995).

1.2.3 Insulators

Insulators function in the genome to restrict genes from being wrongly transcribed by the regulatory elements of the adjacent genes (Maston et al., 2006). They are naturally 500 bp to 3 kb in size and usually function in two main ways. First, they may be a part of the genome and stop enhancer activity by inhibiting the promoter-enhancer interaction (Zhao and Dean, 2004). Second, they may function by blocking the distribution of suppressive chromatin marks in areas comprising transcriptionally active genes (West et al., 2002). Sometimes, insulators are fused with trans-acting proteins to mediate their function e.g. CTCF binds to insulators at the β-globin locus and to all known vertebrate insulators. (Bell et al., 1999)

1.2.4 Locus Control Region

LCRs are bunches of cis-acting regulatory elements such as enhancers, insulators and silencers where the combined operation of these elements, outcomes in the overall control of gene expression regulation (Maston et al., 2006). Like other cis-regulatory elements, LCRs can be positioned at upstream regions, downstream sequences or inside the intronic region of the gene that they regulate. However, contrasting normal enhancers or silencers, LCRs act in a copy number dependent manner and produce an open chromatin structure for associated genes (Li et al., 2002).

1.2.5 Enhancers

Enhancers are Cis-regulatory elements which - unlike silencers - increase the activities of promoters, therefore facilitating the target genes transcription in specific cell types during particular stages in growth and development (Mastonet al., 2006). Usually promoters are stimulated by a broad range of enhancers in various spatial and temporal situations or in response to different stimuli (Visel et al., 2009). A regular enhancer is nearly 50 bp to 1.5 kb in size and comprises of several different transcription factors binding sites (TFBS) which are generally conserved sequences with a precise degree of degeneracy, which transcription factors identify and bind (Mastonet al., 2006). The specificity of the enhancer is controlled by the orientation of the different TFBS. Despite this, enhancer elements are orientation and distance independent (Visel et al., 2009). They can be situated several kb upstream of the promoter, downstream of the promoter in intronic regions, or at/distal to the 3' end of the gene (Visel et al., 2009).

An enhancer can mediate the activation of its corresponding promoter in several different ways i.e. the enhancer bound proteins and promoters may cooperate with one another by looping out the DNA sequence in between, initiating the development of a multi-protein complex for transcription to start. (Ptashne and Gann, 1997; Rippe et al., 1995; Vilar and Saiz, 2005). Likewise, the enhancer and promoter may not encounter one another. Instead, the enhancer may instruct the DNA element to recognize and then bind to certain regions in the nucleus, where high concentrations of transcription factors facilitate transcription (Lamond and Earnshaw, 1998).

In contrast, enhancers may also function via supercoiling of DNA, nucleosome remodeling and altering chromatin structure to produce an available structure for attachment of regulatory proteins to initiate transcriptions (Freeman and Garrard, 1992).

Serious genetic diseases can result in response to malfunctioning of these regulatory elements which leads to faulty gene expression regulation. Understanding the basics of regulatory elements as well as regulation of gene expression is essential to completely outline the function of our genome as well as to look for therapeutic treatments for different genetic diseases. (Barrett et al., 2012). Accumulation of evidence demonstrates that unique limb deformities like Charcot-Marie-Tooth disease (CMT) and Robinow syndrome are produced by the disturbance of certain gene regulatory elements particularly enhancers.

1.3 Charcot-Marie-Tooth Disease

Charcot-Marie-Tooth disease (CMT) is a collection of disorder that mainly damages the peripheral nerves. the nerves that convey sensory information and neural signals from the brain and spinal cord, to and from the other parts of the body (Pareyson et al., 2009). CMT can likewise explicitly influence the muscle controlling nerves. Gradually developing muscle feebleness characteristically distinctively gets visible in puberty or early middle age, however the sickness can begin at any age (Pareyson et al., 2009). Since the longer nerves are majorly harmed, symptoms typically start in the feet and lower legs and afterwards can influence the fingers, hands, and arms. Mostly those suffering from CMT are physically handicap, though a few people may never perceive that they have the sickness (Pareyson et al., 2009).

CMT additionally perceived as genetic motor and sensory neuropathy, is one of the most well-known hereditary nervous issues, influencing approximately 2.6 million people globally (Pareyson et al., 2009). Almost CMT is inherited in all cases. It is likely to have more than one type of CMT if the individual has alterations in two or more genes, each of which triggers a type of the disease. CMT is a heterogenous hereditary condition, indicating alterations in numerous distinct genes can generate comparable clinical indications. CMT is termed after the three doctors who identified and then explained it in 1886 (Pareyson et al., 2009).

There are several different types of CMT disease (Table 1.1), each type exhibiting some common symptoms but differ by their inheritance pattern, beginning age, and whether the axon or myelin sheath of the peripheral nerves is affected.

Table 1.1: Types of CMT.

Type of CMT	Responsible Gene	Chromosomal Position
CMT1A	PMP22	17p12
CMT1B	MPZ	1q22
CMT2	MPZ or GJB1	1q22 or Xq13.1
CMT3	PMP22, MPZ or EGR2	17p12, 1q22 or 10q21.3
CMT4	GDAP1	8q21
CMTX1	GJB1	Xq13.1
***Axonal CMT**	*HADHB	*2p23

This table describes the different types of CMT, the genes responsible for causing the disease and the chromosomal positions of the CMT causing genes. * sign highlights Axonal CMT which is caused by mutation in HADHB gene, lying on the chromosome 2 at position p23.

Distinctive early symptoms of CMT include loss of motion or paralysis of the foot and lower leg muscles, causing trouble in lifting the foot (foot drop) and a fast-treaded pace with recurrent stumbling or falling (Pareyson et al., 2009). Patients may also suffer from balance problem. Foot disfigurements, for example, high curves and twisted toes (hammertoes), are additionally normal in CMT (Pareyson et al., 2009). The lower legs may form an "inverted champagne bottle" silhouette attributable to the loss of muscle mass. As the illness intensifies, weakness, atrophy and muscle decay may also arise in the hands, causing trouble with fine motor skills. Nerve torment can differ from minor to serious, and a few people may require foot or leg supports or other orthopedic gadgets to preserve motion (Pareyson et al., 2009).

1.4 Chromosome 2

Chromosome 2 turned out to be appropriate for our research study since the enhancer region isolated from human chromosome 2 contains a distinct intronic disease variant and this SNP is further found to be responsible for CMT disease that is a limb affecting disorder.

Chromosome 2 is the second largest of the 46 chromosomes existing in human cells (Ijdo et al., 1991). It composes about 8% of the total DNA found inside our cells and spans a total of 243 million base pairs (Ijdo et al., 1991). It is believed that two ancestral chromosomes fused head to head to form chromosome 2 as 24 sets of chromosomes are found in all other types of primate family Hominidae, apart from humans, and the banding pattern of chromosome 2 in humans is similar to that of two distinct chromosomes in gorilla, chimpanzee and orangutan. Besides, unlike the standard human chromosome which consists of one centromere in the middle and telomeres at the ends of a chromosome, human chromosome 2 has the fragments of a second centromere and its telomeres are present at both the midpoint and the extremes. (Ijdo et al., 1991)

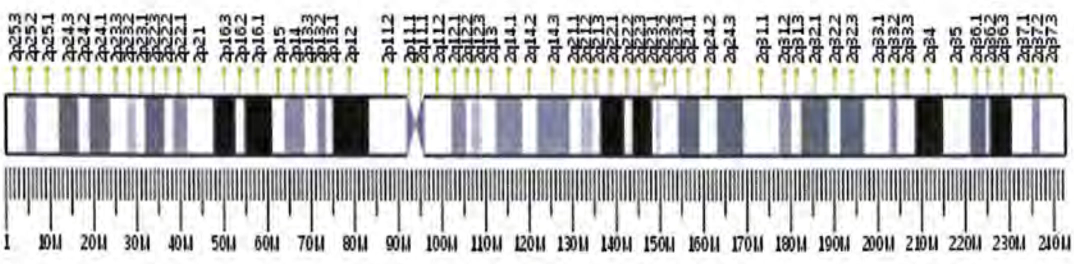

Figure 1.2: G-banding ideogram of human chromosome 2. It is in resolution 850 bphs. Band length in this diagram is proportional to base-pair length. This type of ideogram is usually utilized in genome browsers (e.g. Ensembl, UCSC Genome Browser).

Genetic researchers are working these days to identify and locate the different genes present on human chromosomes and recent evaluations propose that chromosome 2 comprises about 1490 genes that provide instructions for making proteins (Ijdo et al., 1991). Gene irregularities on chromosome 2 have been linked to numerous significant disorders and syndromes, including Charcot-Marie-Tooth (CMT) disease, maturity-onset diabetes, primary pulmonary hypertension and autism (Ijdo et al., 1991).

1.5 Robinow Syndrome

Robinow syndrome is an incredibly uncommon acquired skeletal dysplasia - disorder that influences growth and proper development of the bones and other parts of the body (Patton et al., 2002). It is characterized by short height, mesomelic shortening of limbs, genital hypoplasia, and craniofacial anomalies (Patton et al., 2002).

There are two types of Robinow syndrome (Table 1.2): autosomal recessive Robinow syndrome, and the less severe autosomal dominant Robinow syndrome. They are set apart on the bases of their inheritance pattern, symptoms, and severity.

Table 1.2: Types of Robinow Syndrome.

Types of Robinow Syndrome	Responsible Genes	Chromosomal Positions
Autosomal recessive	ROR2	9q22.31
***Autosomal dominant**	FZD2, *WNT5A, DVL1, or DVL3	17q21.31, *3p14.3, 1p36.33 or 3q27.1

This table describes the different types of Robinow Syndrome, the genes responsible for causing the disease and the chromosomal positions of the Robinow Syndrome causing genes. * sign highlights Autosomal dominant type of Robinow syndrome, which is caused by mutation in WNT5A gene, lying on the chromosome 3 at position p14.3.

Both, the dominant and the recessive form of Robinow syndrome, illustrate numerous common features and physical traits (e.g., craniofacial irregularities, short height, skeletal deformities, and genital hypoplasia). Be that as it may, of course, the manifestations and physical discoveries connected with the passive form are probably going to be more outrageous (Patton et al., 2002). Newborns with the latent type of Robinow syndrome show more extreme rib variations from the norm (e.g., anomalous relocation, combination, as well as nonappearance of specific ribs) and inadequacies influencing bones of the spinal section (vertebrae) than those newborn children with the prevailing type of the disorder (Patton et al., 2002). Moreover, short height, underdevelopment of the lower arm bones (radioulnar hypoplasia), and abnormalities of the fingers are more critical. Affected kids may display displacement of the head of one of the forearm bones (radial head dislocation), an irregularity seldom witnessed in infants with the prevailing type of Robinow syndrome (Patton et al., 2002).

1.6 Chromosome 3

Chromosome 3 also turned out to be appropriate for our research study considering the fact that a specific enhancer region isolated from human chromosome 3 contains seven distinct intronic disease variants and their SNPs are further found to be responsible for another limbs affecting disorder i.e. Robinow syndrome.

Chromosome 3 is the third largest chromosome of the total 23 pairs of chromosomes existing in human cells (Naylor et al., 2001). It extents approximately 200 million nucleotide base pairs which are the fundamental units of DNA, constituting around 6.5% to 7% of the hereditary DNA in the human cells (Naylor et al., 2001).

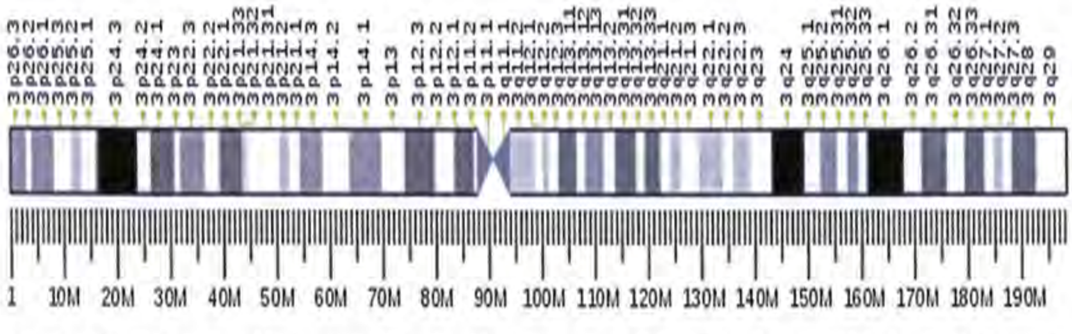

Figure 1.3: G-banding ideogram of human chromosome 3. It is in resolution 850 bphs. Band length in this diagram is proportional to base-pair length. This type of ideogram is usually utilized in genome browsers (e.g. Ensembl, UCSC Genome Browser).

The centromere of chromosome 3 is located someplace near its center, making chromosome 3 a metacentric chromosome (Naylor et al., 2001). There are numerous essential genes positioned on chromosome 3 including some gene clusters that encode the olfactory receptors (associated with feeling of smell) along with chemokine receptors that facilitate inflammatory procedures (Naylor et al., 2001). Genetic study is determined on finding and categorizing the genes on human chromosomes and estimations propose that chromosome 3 comprises around 1150 to 1600 genes. Few significant genes present on chromosome 3 include ALAS1, CCR5, TMIE and WNT5A (Naylor et al., 2001).

1.7 Synteny and Syntenic Conservation

Analyzing the conservation pattern of the enhancers and their surrounding regions in upper primates with the aid of synteny helps to infer the association between the putative enhancers, its disease variants and the associated disease-causing genes.

Genome mapping and sequencing have allowed the comparison of the overall genomic sequence of numerous dissimilar species. The characteristic result shows that relatively recently diverged creatures express similar blocks of genes in the same relative positions in their genomes (Liu et al., 2018). This state which is interpreted generally as having shared chromosome sequences is called synteny (Liu et al., 2018). In general, a block of synteny is described as a stretch of genomic sequence including several orthologous genes that are co-arranged with another genome. Synteny identification is a sorting and arranging method for every single neighborhood likeness between various genome sequences into a clear comprehensive representation (Liu et al., 2018).

Conserved or shared synteny denotes at least two homologous genes that are conserved in two or more different species, irrespective of the gene arrangement on the chromosome (Moreno-Hagelsieb et al.,2001). It is the conservation of the content and order of genes in specified areas of chromosome in associated classes of organisms. Various discoveries on chromosome advancement revealed that preserved synteny occurs among firmly related species as well as over long developmental timescales (Moreno-Hagelsieb et al.,2001). Conserved synteny over a long-time span is a predominantly frequent characteristic about developmentally important genes, representing that chromosome reorganizations are not an unbiased, haphazard development in genome evolution (Moreno-Hagelsieb et al.,2001).

One of the most consistent measure for determining the orthology of genome sequences in various species is shared synteny (Liu et al., 2018). Greater-than-estimated shared synteny can indicate choice for important purposeful associations among conserved genes, for instance grouping of alleles that are beneficial when acquired together, or shared regulatory processes (Moreno-Hagelsieb et al.,2001). Synteny is extensively utilized in examining complex genomes, microbial genomics, phylogenetic relationships, and to deduce the genome organization of extinct ancestral species (Moreno-Hagelsieb et al., 2001).

1.8 Aims of Study

Comparative genomics offers an influential tool for exploring evolutionary transformations amid organisms, serving to recognize conserved or similar genes among species, along with the genomic sequences that impart every organism its exclusive features (Frazer et al., 2003). It facilitates researchers find indicators that locate the positions of genes and the regulatory sequences that control gene expression. The present-day study tries to excerpt and understand the adjacent regions of two distinct limb specific enhancers which are associated to two different limb diseases, to find the conservation pattern of the enhancer-surrounding genes among mammals and to represent the conservation of the limb specific genes surrounding these two enhancer regions in the form of synteny. Synteny allows us to identify and isolate the genes that are conserved throughout the species, and then to link the enhancer with the gene it regulates based on conservation pattern. Owing to disease relevance and some previously performed studies, enhancer regions of chromosome 2 and 3 are selected for our research study. A region of 4mbs upstream and downstream of the limb specific enhancers, is analyzed to study the conservation pattern of the genes and the disease variants with the help of synteny mapping.

Chapter 2

MATERIALS AND METHOD

2.2 Enhancer Conservation

The conservation of limb specific enhancers was seen through UCSC conservation track.

2.2.1 UCSC Browser

The University of California Santa Cruz (UCSC) is a well-known genome Bioinformatics website consisting of a suite of free, open-source, on-line tools and one of the largest online genome sequence repositories that can be utilized to inspect, study, and query genomic information. These tools are accessible by anyone with an Internet browser and an interest in genomics. UCSC conservation track displays multiple sequence alignments (MSA) of over 100 vertebrate species and extent of developmental preservation utilizing two distinct procedures (phastCons and phyloP) from the PHAST bundle, for all species. The MSAs were produced using multiz and other tools in the UCSC. Conservation of selected human limb specific enhancers of chromosome 2 and 3 in other species is observed from UCSC genome browser particularly for GRCh38 genome assembly.

2.3 Sequence similarity

2.3.1 Ensembl-BLAST

The Basic Local Alignment Search Tool (BLAST) locates areas of local similarity between different DNA sequences. The tool matches the given nucleotide or protein sequences to sequence database and computes the numerical importance of matches. BLAST is utilized to gather useful and evolutionary associations between different sequences besides identifying affiliates of various gene families. Genomic sequences of the limb specific enhancers of chromosome 2 and 3 are retrieved from Ensembl database and are then blast against upper primates using the ensembl-blast tool to find the sequence similarity among various species. The resulting top hits are selected and viewed to analyze the adjacent genomic regions of the enhancer in the designated species.

2.4 Variant Conservation

2.4.1 UCSC browser

Conservation pattern of the variants of chromosome 2 and 3, that were retrieved from GWAS, is studied using UCSC genome browser particularly for GRCh38 genome assembly.

Regional Analysis

The hits against the human limb specific enhancers of the chosen species are examined one after the other. The analysis covers a region of 4mbs upstream and 4mbs downstream of the enhancer region. Following are the checkpoints of our research analysis:

i. Enhancer lies within a gene or not.

ii. Enhancer is in a gene desert or not.

iii. Check the expression of the genes within 4mb region of the enhancer's upstream and downstream.

iv. Whether any of the gene within the specified region is reported in literature to have role in the aforementioned disease linked with the enhancers or not.

v. Filter the limb specific genes by checking their expression in expression databases, to make the synteny.

2.6 Gene Expression

2.6.1 MGI-Gene expression database (GXD)

The Gene Expression Database (GXD) is a public reserve for gene expression information from the mouse research center. GXD supplies and incorporates several kinds of expression data and makes the information accessible without any restrictions in presentations suitable for complete study. This expression database emphasizes on endogenous gene expression throughout mouse growth. GXD stores basic data from diverse expression databases. By coordinating the information accumulated, GXD provides complete information about the expression reports of transcripts and proteins in various mouse strains and mutants. Combination with MGD empowers a consolidated examination of genotype, sequence expression, and phenotype. The type of assay selected for viewing the expression is RNA in situ. The check point of the research was to filter and isolate the genes showing expression in the limbs

2.7 Synteny

To associate the enhancers with the target genes, Synteny mapping is done, which diagrammatically represented the limb specific genes within the enhancer's 4mbs upstream and downstream region. It paved a way to analyze the conservation pattern of the genes along with their related enhancers within the selected species.

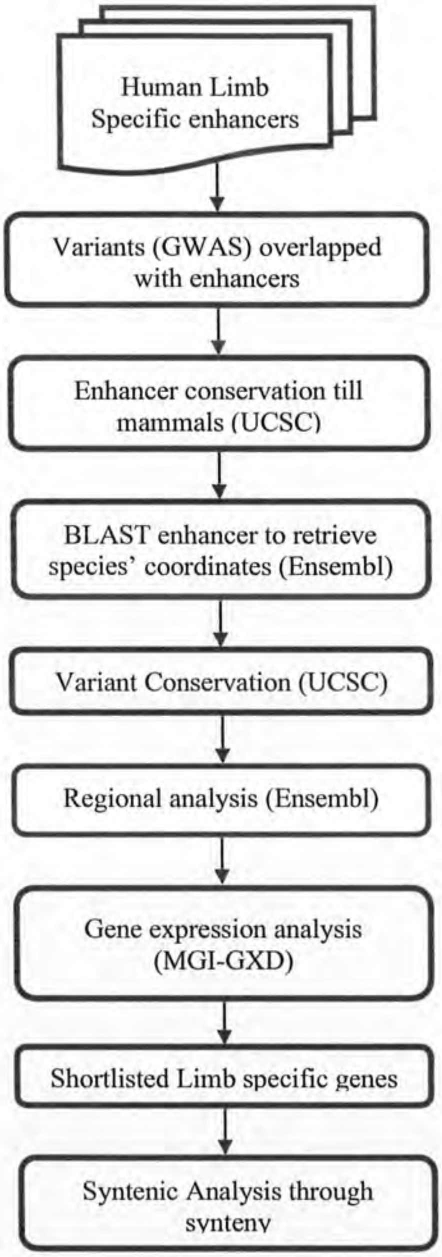
Human Limb
Specific enhancers
Variants (GWAS) overlapped with enhancers
Enhancer conservation till mammals (UCSC)
BLAST enhancer to retrieve species' coordinates (Ensembl)
Variant Conservation (UCSC)
Regional analysis (Ensembl)
Gene expression analysis (MGI-GXD)
Shortlisted Limb specific genes
Syntenic Analysis through synteny

Figure 2.1: Flow chart depicting the devised working pipeline, followed during the research study. Primarily, limb specific enhancers of human chromosome 2 and 3 were overlapped with disease variants to check variants' presence in enhancers. Then, the enhancer conservation pattern was analyzed, and the enhancer region was Blast against different species to get their enhancer coordinates. The conservation pattern of the variants was also studied. Lastly, the regional analysis of the enhancer was done and after checking the gene expression of the neighboring genes, limb specific genes were selected for syntenic conservation analysis through synteny.

Chapter 3

RESULTS

3 RESULTS

The limb specific enhancers of human chromosome 2 and 3 were overlapped with disease variants to check variants' presence in enhancers. Then, the enhancer conservation pattern was analyzed, and the enhancer region was Blast against different species to get their enhancer coordinates. The conservation pattern of the variants was also studied. Lastly, the regional analysis of the enhancer was done and after checking the gene expression of the neighboring genes, limb specific genes were selected for syntenic conservation analysis through synteny.

3.1 Enhancer-Variant Overlap

The disease variants of human chromosome 2 and 3 (taken from GWAS) were overlapped with the selected enhancers using bed tools to check whether the variant lies within the enhancer or not. The variants that overlapped and resided within limb specific enhancers of the chromosome 2 and 3 are shown in Table 3.1 and 3.2 respectively.

Table 3.1: Variant lying in enhancer region of HSA2.

Variant Region	Variant Location
aga	26269951 - 26269953

This table describes the sequence of the disease variant that overlapped with the enhancer region of chromosome 2. Its location is also given in Grch38 genome assembly

Table 3.2: Variants lying in enhancer region of HSA3.

Variant Region	Variant Location
taa	55467213-55467215
cgc	55467280-55467282
acc	55467436-55467438
ata	55467606-55467608

ggg	55468038-55468040
acg	5468080-55468082
cgt	55468081-55468083

This table describes the sequence of the disease variants that overlapped with the enhancer regions of chromosome 3. Its location is also given in Grch38 genome assembly.

3.2 Enhancer Conservation

As a pre-requisite to our research, what species were to be selected for the enhancer region analysis, we tested the conservation status of the enhancer in 100 vertebrates via UCSC multiz alignment. The conservation status of the limb specific enhancers of the chromosome 2 and 3 shown in figure 3.1 and figure 3.2 respectively illustrate that the enhancer region is more conserved in upper primates.

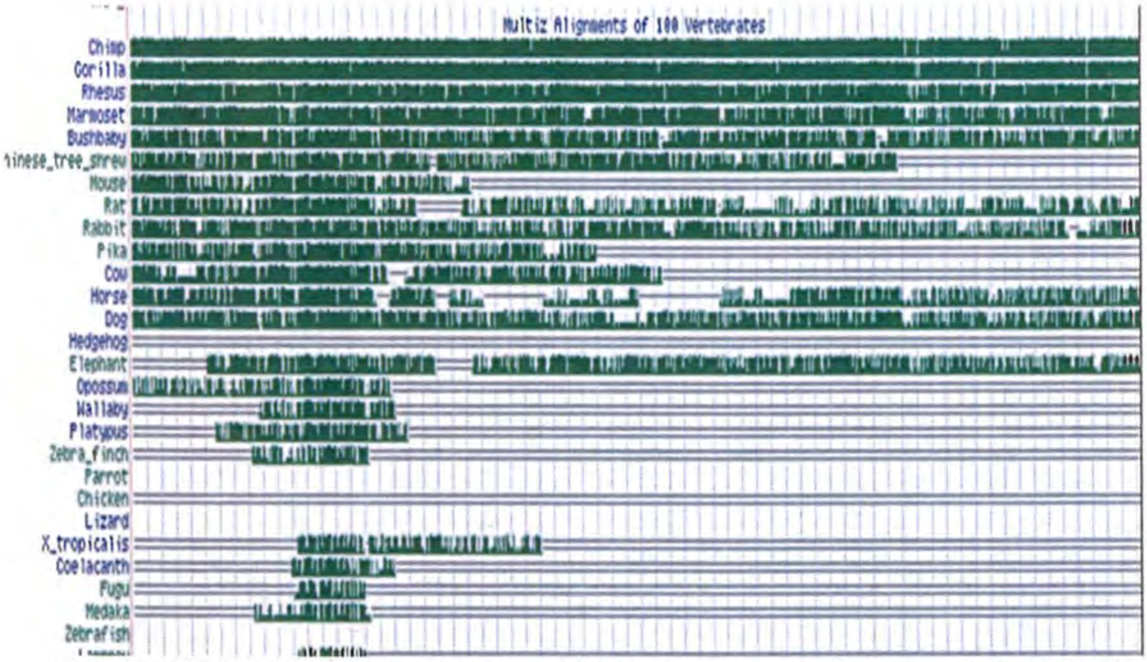

Figure 3.1: Conservation status of enhancer of human chromosome 2. The multiz alignment shown that the enhancer region is more conserved in upper primates. The names of the species are mentioned on the left side, and the green peaks in front of species illustrate the conservation level of the enhancer region in the specie.

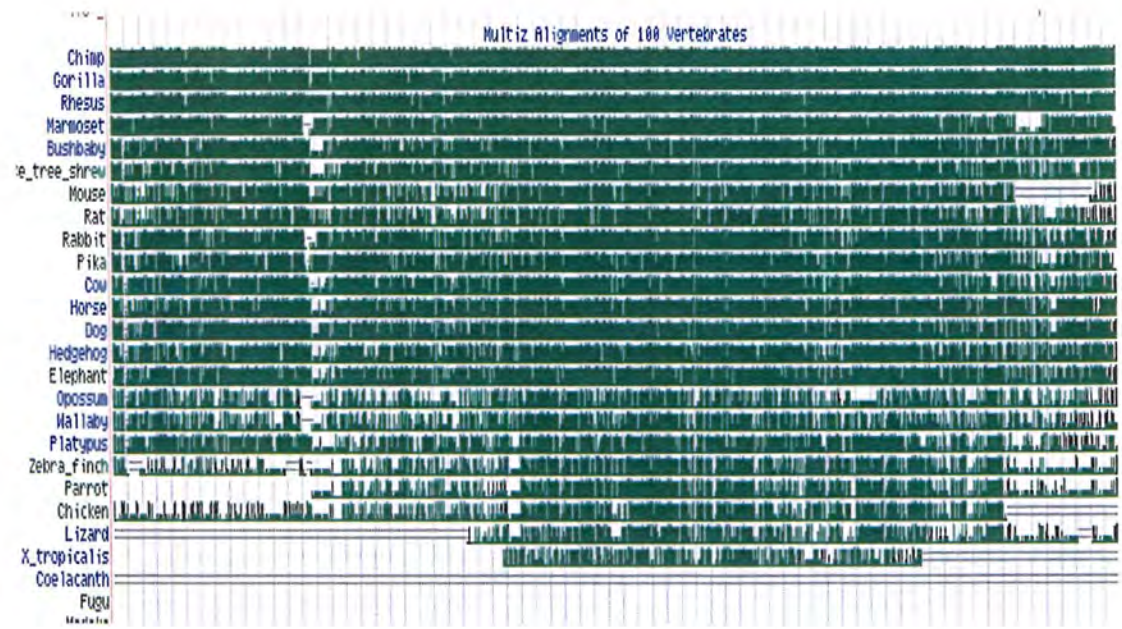

Figure 3.2: Conservation status of enhancer of human chromosome 3 The multiz alignment shown that the enhancer region is more conserved in upper primates. The names of the species are mentioned on the left side, and the green peaks in front of species illustrate the conservation level of the enhancer region in the specie.

3.3 Ensembl-BLAST results

The genomic sequences of the limb specific enhancers of chromosome 2 and 3 are retrieved from Ensembl and are uploaded on the ensembl-blast tool to find the sequence similarity against the selected species. The resulting top hits are selected and viewed to examine the surrounding regions of the enhancer in all species.

Table 3.3: Ensembl-BLAST results of enhancer of HSA2

Species	Genomic Location	Overlapping Genes	Orientation	Query Start	Query End	Length	Score	E-Value	%Id
Human	2:26269774-26270782	HADHB	Forward	1	1009	1009	1994	0	100
Mouse	5:30168478-30168578	HADHB	Forward	132	234	103	113	1E-21	89.32

Chimpanzee	2A:26808436-26809443	HADHB	Forward	1	1009	1009	1945	0	99.41
Elephant	Scaffold_20: 39538702-39538954	HADHB	Forward	487	256	256	117	2E-23	80.86
Rhesus (Macaque)	13:82699741-82700752	HADHB	Reverse	1	1019	1019	1486	0	93.33
Orangutan	2a:85881631-85882641	HADHB	Reverse	1	1013	1013	1785	0	97.33
Opossum	1:507094840-507094896	HADHB	Forward	176	57	57	73.6	2E-10	91.23
Dog	17:20375256-20375377	ENSCAFG 00000029395	Forward	166	1009	122	170	1.00-39	92.62

This table describes the BLAST Results of the human enhancer region of Chromosome 2 against Mouse, Chimpanzee, Elephant, Rhesus, Orangutan, Opossum and Dog. The Genomic Location, Overlapping Genes, Orientation, Query Start, Query End, Length, Score, E-Value and %Id values of the selected species are also stated in the table.

Table 3.4: Ensembl-BLAST results of enhancer of HSA3.

Species	Genomic Location	Overlapping Genes	Orientation	Query Start	Query End	Length	Score	E-Value	%Id
Human	3:55467180 - 55468112	WNT5A	Forward	1	933	933	1844	0	100
Mouse	14:28525256-28525547	WNT5A	Reverse	361	651	293	356	6.00 E-95	90.44
Chimpanzee	3:56611547-56612484	WNT5A	Forward	1	933	938	1781	0	98.93
Elephant	scaffold_12: 11946524 - 11947404	-	Reverse	35	922	895	984	0	88.72
Rhesus (Macaque)	2:111627855-111628763	-	Forward	23	933	911	1650	0	97.91
Orangutan	3:90815951-90816890	ENSPPYG 00000013781	Reverse	1	933	941	1688	0	97.56
Opossum	6:280924669-280924721	-	Reverse	475	527	53	57.8	1.00 E-05	88.68
Dog	20:34788535-34789442	-	Reverse	26	933	919	1196	0	91.62

This table describes the BLAST Results of the human enhancer region of Chromosome 3 against Mouse, Chimpanzee, Elephant, Rhesus, Orangutan, Opossum and Dog. The Genomic Location, Overlapping Genes, Orientation, Query Start, Query End, Length, Score, E-Value and %Id values of the selected species are also stated in the table.

3.4 Variant conservation

Enhancer regions were found conserved in mammals. Next, we checked whether the variants are also conserved in the species under study or not. The disease variants of chromosome 2 and 3 (taken from GWAS) were overlapped with their enhancers for analysis of variant conservation pattern through UCSC multiz alignment in the sėlected species. Disease variants of chromosome 2 and 3 were responsible for CMT and Robinow Syndrome respectively. Table 3.5 and 3.6 shows their conservation status in the species under analysis.

Table 3.5: Variant Conservation of enhancer of HSA2.

Specie	Variant Region 26269951 - 26269953
Human	*aga
Chimp	*aga
Orangutan	*aga
Rhesus	*aga
Dog	*aga
Elephant	*aga
Opossum	*aga
Mouse	agg

This table describes the conservation status of the variant of the limb specific enhancer of the human chromosome 2 in other species **under analysis**. * sign indicates the conserved variant sequence in the given region. The variant sequence aga is conserved in all the selected species except in mouse.

Table 3.6: Variants Conservation of enhancer of HSA3.

Specie	Variant Regions						
	55467213-55467215	55467280-55467282	55467436-55467438	55467606-55467608	55468038-55468040	55468080-55468082	55468081-55468083
Human	*taa	*cgc	*acc	*ata	ggg	*acg	*cgt
Chimp	*taa	*cgc	*acc	*ata	ggg	*acg	*cgt
Orangutan	*taa	cgt	*acc	*ata	*gag	*acg	*cgt
Rhesus	*taa	*cgc	*acc	*ata	*gag	atc	tct
Dog	*taa	*cgc	atc	*ata	*gag	gcg	cgt
Elephant	caa	*cgc	atc	*ata	gaa	atg	tgt
Opossum	tac	tgc	tta	*ata	cag	ggg	ggc
Mouse	tga	cac	ctc	*ata	gaa	gta	tat

This table describes the conservation status of the variant of the limb specific enhancer of the human chromosome 3 in other species under analysis. * sign indicates the conserved variant sequence in the given region. The variant ata in the region 55467606 -55467608 shows conservation in all the selected species.

3.5 Regional Analysis

We performed regional analysis on 4mb upstream and downstream of the enhancer. Enhancers of both the chromosome 2 and 3 were not found in the gene rich areas. All the genes that were residing in this region were directed to the gene expression database for gene expression analysis.

3.6 Gene expression Analysis

With the help of MGI-GXD (Gene eXpression Database) we performed gene expression analysis on all the genes residing in 4mb upstream and downstream of the enhancer region. As our enhancers were limb specific so we shortlisted limb specific genes for further analysis and the data was written down in excel (Table 3.7 and Table 3.8).

Table 3.7: Surrounding genes of enhancer of HSA2.

Specie	Genes residing upstream				CNE region	Genes residing downstream			
	4MB	3MB	2MB	1MB		1MB	2MB	3MB	4MB
Human	-	-	ADCY3	HADHA DNMT3A	HADHB	FAM166C CGREF1 ABHD1 CAD	TRIM54 GTF3C2 PPM1G	FOSL2 PPP1CB WDR43 ALK	LBH
Chimp	-	-	ADCY3	HADHA DNMT3A	HADHB	CGREF1 ABHD1 CAD	TRIM54 GTF3C2 PPM1G	FOSL2 PPP1CB WDR43 ALK	LBH
Orangutan	ALK	FOSL2 PPP1CB WDR43	TRIM54 GTF3C2 PPM1G	FAM166C CAD	HADHB	HADHA DNMT3A	ADCY3	-	-
Macaque	ALK	FOSL2 PPP1CB WDR43	TRIM54 GTF3C2 PPM1G	FAM166C CGREF1 ABHD1	HADHB	HADHA DNMT3A	ADCY3	-	-
Mouse	-	-	-	HADHA	HADHB	CGREF1 ABHD1 CAD TRIM54 GTF3C2	FOSL2	PPP1CB	-
Dog	-	-	ADCY3	HADHA DNMT3A	-	FAM166C ABHD1 CAD TRIM54 GTF3C2 PPM1G	FOSL2	PPP1CB WDR43 ALK	LBH
Elephant	-	-	ADCY3 DNMT3A	HADHA	HADHB	FAM166C ABHD1 CAD	TRIM54 GTF3C2 PPM1G CAD CGREF1	FOSL2	WDR43 PPP1CB ALK
Opossum	-	-	ADCY3 DNMT3A	HADHA	HADHB	FAM166C CAD	TRIM54 GTF3C2 PPM1G CAD	FOSL2	WDR43 PPP1CB

This table shows the region wise distribution of the limb specific genes within 4mb upstream and downstream of limb specific enhancer of human chromosome 2 against other species.

Table 3.8: Surrounding genes of enhancer of HSA3.

Specie	Genes residing upstream				CNE region	Genes residing downstream			
	4MB	**3MB**	**2MB**	**1 MB**		**1 MB**	**2MB**	**3MB**	**4MB**
Human	-	FLNB PDHB ARF4	ARHGE3 IL17RD HESX1	ERC2	WNT5A	-	CHDH	SPCS1 PRKD TKT DCP1A	RAD542 DUSP7 TNNC1
Chimp	PDHB	FLNB ARF4	ARHGE3 IL17RD HESX1	ERC2	WNT5A	-	CHDH	SPCS1 PRKD TKT DCP1A	RAD542 DUSP7 TNNC1
Orangutan	RAD54L2 TNNC1	SPCS1 PRKCD TKT DCP1A	CHDH	-	-	ERC2	ARHGEF3 IL17RD HESX1	FLNB ARF4	PDHB
Macaque	-	FLNB PDHB ARF4	ARHGEF3 IL17RD HESX1	ERC2 WNT5A	-	-	CHDH	SPCS1 PRKCD TKT DCP1A	RAD54L2 DUSP7 TNNC1
Mouse	-	TNNC1 SPCS1 PRKCD TKT	CHDH DCP1A	-	WNT5A	ERC2	ARHGEF3 IL17RD HESX1 ARF4	-	-
Dog	RAD54L2	TNNC1 SPCS1	CHDH DCP1A TKT PRKCD	-	-	ERC2 WNT5A	ARHGEF3 IL17RD HESX1 ARF4	FLNB	-
Elephant	DUSP7 TNNC1	PRKCD TKT DCP1A	CHDH	-	-	WNT5A	ARHGEF3 IL17RD HESX1 FLNB	ARF4	PDHB
Opossum	-	TKT	CHDH	-	-	ERC2 WNT5A	IL17RD HESX1	-	DUSP7

This table shows the region wise distribution of the limb specific genes within 4mb upstream and downstream of limb specific enhancer of human chromosome 3 against other species.

3.7 Synteny

Synteny portrays the physical co-localization of hereditary loci on the similar chromosome inside an individual or species. It is the preservation of gene order blocks in two groups of genetic material that are being assessed with each other. This concept of gene conservation can also be referred to as shared synteny. In our research, we observed synteny in human, chimp, orangutan, macaque, mouse, dog, elephant and opossum. As our enhancers were limb specific, so we only refined limb specific genes within the region of enhancer's 4mbs upstream and downstream for syntenic conservational analysis. Figure 3.3 & 3.4 shows the synteny of the enhancer of HSA2 and HSA3 respectively.

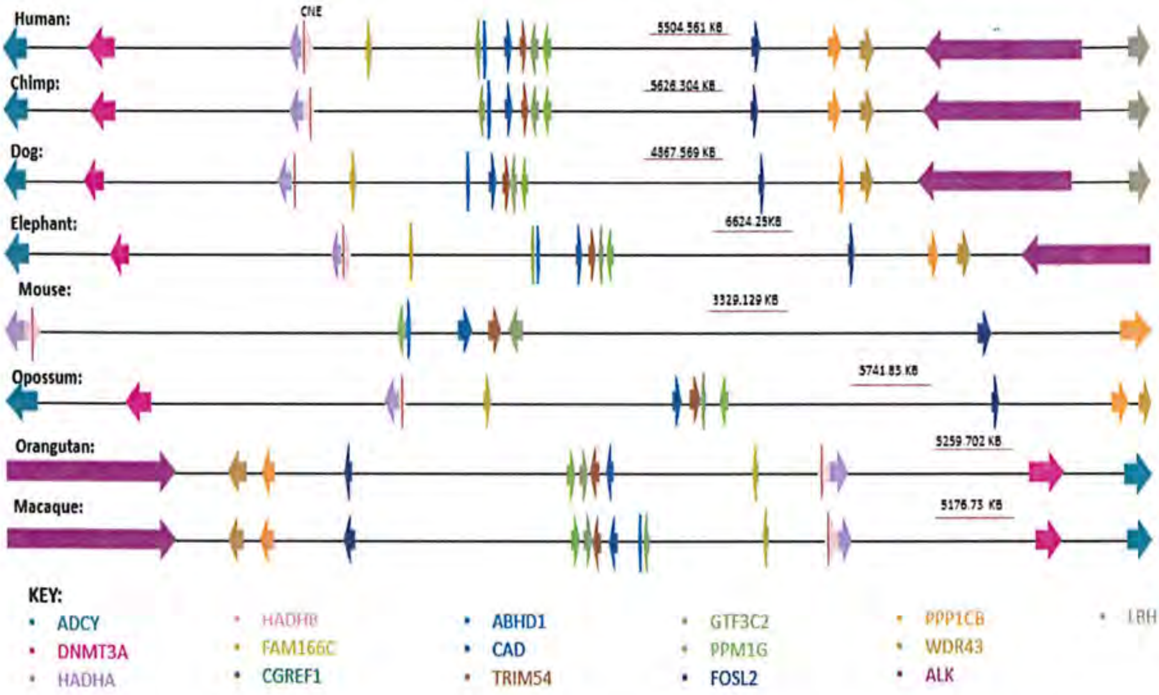

Figure 3.3: Synteny of enhancer of HSA2. Synteny and the syntenic conservation of the limb specific genes surrounding the given enhancer of human chromosome 4 within a region of 4mbs. Colored arrows depict different genes. The direction of gene transcription is represented by the direction of gene arrow. Red vertical line characterizes the position of CNE-enhancer. Horizontal orange line portrays scale.

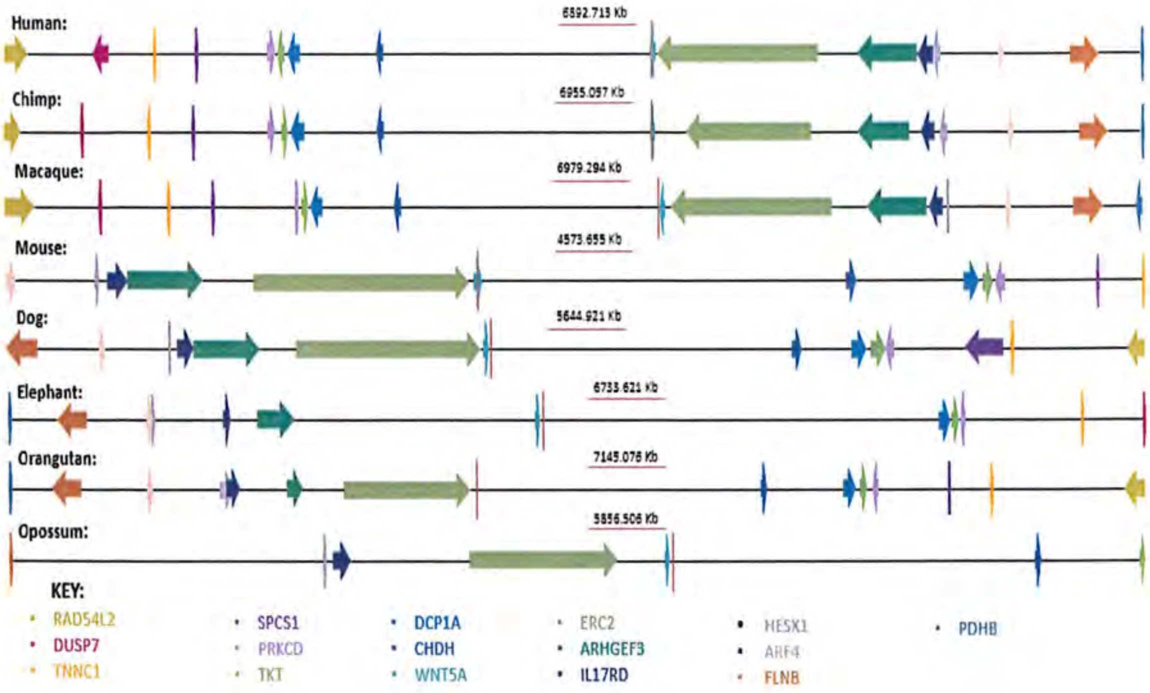

Figure 3.4: Synteny of enhancer of HSA3. Synteny and the syntenic conservation of the limb specific genes surrounding the given enhancer of human chromosome 6 within a region of 4mbs. Colored arrows depict different genes. The direction of gene transcription is represented by the direction of gene arrow. Red vertical line characterizes the position of CNE-enhancer. Horizontal orange line portrays scale.

Chapter 4
DISCUSSION

4. DISCUSSION

DNA, our genetic manual consists of instructions for the proteins, which along with building the structure of our bodies, also provide the power and driving force for our bodies to function well. But less than 2 percent of our DNA actually encodes them. The remaining - 98.5 percent of the human DNA - is hypothesized as "Junk DNA" that scientists assumed useless for a long time (Judith Mary et al., 2019). Nature has an extraordinary, unusual way of writing. Our genetic script uses just four letters: A, G, C, and T. Different arrangements of these four characters constitute our genes, which instruct the formation of proteins. But before arranging the proteins together, DNA transcribes into strands of RNA that are sliced and then reunited into relatively smaller sequences (Judith Mary et al., 2019).

During this chopping of RNA, the non-coding parts of DNA - the junk – are discarded, indicating they never even get used to build proteins. Then why nature holds so much apparently unnecessary information in its guide is a query that scientists keep on pondering. A comparatively rational explanation is that this "junk DNA" might not be so meaningless after all. Modern-day studies reveal that the "junky" fragments of our genome might play some very important and vital roles, exclusively in the control of gene activity (Alberts et al., 2013).

Noncoding DNA is the part of genome that functions as regulatory DNA elements, controlling the spatial-temporal regulation of the genes i.e. monitoring the location and the time for the gene to be switched on and off. Noncoding DNA comprises of numerous kinds of cis and trans regulatory elements like promoters, enhancers, silencers, insulators and many other non-coding elements (Alberts et al., 2013).Such regulatory elements act by either stimulating or suppressing the transcription method used to transform genomic information into proteins by offering positions for the transcription factors to bind to the DNA and initiate transcription (Alberts et al., 2013).

Amid every single regulatory element, enhancers possess a major responsibility. They are distal regulatory elements playing a vital character in governing the gene expression through the binding of a combination of various transcription factors just like promoters (Hari Prakash et al., 2019). But clarifying the activity of enhancers stays more subtle as they do not have an exact sequence motif or structure for their definite genome-wide recognition. Furthermore, the relative position of the enhancer with reference to its target genes can be significantly inconsistent as they may exist in the locality of their target genes but do not essentially determine the nearest gene. Enhancers may be present either downstream or upstream of their target or may function across mediating genes to influence the target gene. The verity that one enhancer can influence several genes, or one gene can be controlled by various enhancers makes the enhancer-target genes (ETG) coupling more complex (Hariprakash et al., 2019). Lastly, the action of enhancers may be limited to a specific tissue or a cell type, a definite time in life, or to a certain biological, pathological or ecological state. Though this dynamic character of enhancers allows their genomic nature to regulate the spaciotemporal expression of the genes, it additionally makes the detection and functional annotation of the enhancers in the genome challenging (Pennacchio et al., 2013).

Even with these challenges, some comparative genomics techniques can be used to identify enhancers on a genome-wide scale. Accumulation of evidence demonstrates that the part of non-coding DNA that is exceedingly conserved throughout evolution between different species is rich with enhancer sequences, especially the DNA part that is active in early developmental stages (Pennacchio et al., 2013). Comparisons of different genomes manifest homologous sequences that reveal their mutual evolutionary origin and succeeding conservation. During evolution, owing to the phenomenon of natural selection, the functional part of DNA segments remains more preserved than the nonfunctional segments. So, the DNA sequences that are conserved between different species are expected to code the same functions. Sequence comparisons between species offer evidence on gene structures and expose regulatory elements. Further, sequence conservation pattern between different species, inside genic and non-genic regions, can be applied for the creation of physical map or synteny. We exploited same property of comparative genomics in our study for the identification of the enhancer target genes.

Limiting the scope to limbs, our lab mates compiled a catalogue of enhancer regions having high regulatory probability of limb specific genes. This catalogue was generated and purified using strong functional and statistical evidence. Two distinct putative enhancers were selected, one from Chromosome 2 and another from Chromosome 3. Both the chromosomes are larger in size comparatively to other chromosomes and are loaded with genes. The enhancers were selected on the basis of their disease relevance, and further the disease variants of both these enhancer regions were obtained using GWAS. Deep literature studying was done to identify the particular disease-causing genes which are regulated by the specifically selected enhancer regions. A region of 4mbs upstream and downstream of our enhancer region was selected as the research space since an enhancer can regulate a gene upto 4mbs distance. All the genes obtained surrounding the enhancer region upto 4 mbs distance went through Gene Expression database for the inspection of their expression. The genes which showed their expression in limbs were filtered out as we previously knew that the enhancer chosen was limb specific, so clearly the target gene should also have limb specific properties. The conservation pattern of the filtered genes (showing expression in limbs) was analyzed with the help of synteny mapping.

The goal was to find the genes that were responsible for causing the limb specific disease and were regulated by the given enhancer regions. We victoriously retrieved the relation of the limb specific enhancer with the disease-causing gene through the literature and then we validated the results with the help of conservation pattern analysis of the enhancer and the genes through synteny.

Through literature review, we found out that Axonal CMT is a heterogeneous syndrome of the peripheral nervous system, caused by complex heterozygous mutation in HADHB gene. The typical symptoms included weakness in legs, bilateral foot drop, gait disturbance, muscle softness and decay prominently in the distal parts of the lower limbs, and absence of deep tendon reflexes in the limbs. Then by using Bioinformatics computational tools and databases, we learnt that HADHB gene is situated on the p arm of chromosome 2 on position 23, the same chromosome on which our enhancer region is located. In the syntenic analysis of 4mb region upstream and downstream of the chromosome 2 enhancer, it was observed throughout the mammals (except dogs) that limb specific enhancer of chromosome 2 lies inside the disease-linked target gene i.e. HADHB gene. This conservation pattern of the enhancer and the gene analyzed through synteny validated the literature information that it was in fact the HADHB gene, which was regulated by the given enhancer region and was involved in causing axonal CMT which is a limb specific disease.

Again, while reviewing the literature about our second disease, we found out that autosomal dominant Robinow syndrome - an incredibly uncommon acquired skeletal dysplasia that affects the development of many parts of the body, particularly the skeleton, can be produced by alterations in numerous genes, including FZD2, WNT5A, DVL1 and DVL3. Among all these disease-causing genes, WNT5A caught our interest as its located on the p arm of chromosome 3 on position 14.3, the same chromosome on which our enhancer region is located. Next, during synteny analysis of 4mb region upstream and downstream of the chromosome 3 enhancer, it was observed throughout the mammals (except orangutan) that limb specific enhancer of chromosome 3 lies either inside the disease-linked target gene i.e. WNT5A gene or either within 1 kb distance of the target gene. This conservation pattern of the enhancer and the gene analyzed through synteny again validated the literature information that it was in fact the WNT5A gene, which was controlled by the given enhancer region and was involved in causing Robinow syndrome.

In conclusion, this research represents a Bioinformatics platform supported by literature and experimental evidence, which can be used for locating disease-causing target genes and other co-localized genes around the given enhancer with the help of syntenic analysis between different species.

Chapter 5
REFERENCES

5. REFERENCES

Iscan, M. Y., & Steyn, M. (2013). The human skeleton in forensic medicine. Charles C Thomas Publisher.

Petit, F., Sears, K. E., & Ahituv, N. (2017). Limb development: a paradigm of gene regulation. Nature Reviews Genetics, 18(4), 245-258.

Ijdo, J. W., Baldini, A., Ward, D. C., Reeders, S. T., & Wells, R. A. (1991). Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proceedings of the National Academy of Sciences, 88(20), 9051-9055.

Pareyson, D., & Marchesi, C. (2009). Diagnosis, natural history, and management of Charcot–Marie–Tooth disease. The Lancet Neurology, 8(7), 654-667.

N Patton, M. A., & Afzal, A. R. (2002). Robinow syndrome. Journal of medical genetics, 39(5), 305-310. aylor, S. L., & Garcia, D. K. (2001). Chromosome 3. e LS.

. Patton, M. A., & Afzal, A. R. (2002). Robinow syndrome. Journal of medical genetics, 39(5), 305-310.

Ehrlich, J., Sankoff, D., & Nadeau, J. H. (1997). Synteny conservation and chromosome rearrangements during mammalian evolution. Genetics, 147(1), 289-296.

Liu, D., Hunt, M., & Tsai, I. J. (2018). Inferring synteny between genome assemblies: a systematic evaluation. BMC bioinformatics, 19(1), 1-13.

Moreno-Hagelsieb, G., Treviño, V., Pérez-Rueda, E., Smith, T. F., & Collado-Vides, J. (2001). Transcription unit conservation in the three domains of life: a perspective from Escherichia coli. TRENDS in Genetics, 17(4), 175-177.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., ... & Gordon, L. (2010). Ensembl 2011. Nucleic acids research, 39(suppl_1), D800-D806.

Hariprakash, J. M., & Ferrari, F. (2019). Computational Biology Solutions to Identify Enhancers-target Gene Pairs. Computational and structural biotechnology journal, 17, 821-831.

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. Nature Reviews Genetics, 14(4), 288-295.

Holzbaur, K. R., Delp, S. L., Gold, G. E., & Murray, W. M. (2007). Moment-generating capacity of upper limb muscles in healthy adults. Journal of biomechanics, 40(11), 2442-2449.

Young, N. M., Wagner, G. P., & Hallgrímsson, B. (2010). Development and the evolvability of human limbs. Proceedings of the National Academy of Sciences, 107(8), 3400-3405.

Tabin, C. J. (1991). Retinoids, homeoboxes, and growth factors: toward molecular models for limb development. Cell, 66(2), 199-217.

Shabalina, S. A., & Spiridonov, N. A. (2004). The mammalian transcriptome and the function of non-coding DNA sequences. Genome biology, 5(4), 105.

Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., ... & Walter, P. (2013). Essential cell biology. Garland Science.

Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cellular and molecular life sciences, 69(21), 3613-3634.

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. Annu. Rev. Genomics Hum. Genet., 7, 29-59.