CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Identifying Genetic Expression of Regulatory Regions in Human and Non-Human Primates Influencing Facial Feature Development

by

Hadiqa Nadeem

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2021

Every challenging work needs self-efforts as well as the guidance of elders. I dedicate this thesis to my parents whose affection makes me able to get such success and to my teachers whose encouragement has always been my source of inspiration.

CERTIFICATE OF APPROVAL

Identifying Genetic Expression of Regulatory Regions in Human and Non-Human Primates Influencing Facial Feature Development

by

Hadiqa Nadeem

(MBS193023)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Bashir Ahmad	IIU, Islamabad
(b)	Internal Examiner	Dr. Sahar Fazal	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
September, 2021

Dr. Sahar Fazal
Head
Dept. of Bioinfo. and Biosciences
September, 2021

Dr. M. Abdul Qadir
Dean
Faculty of Health and Life Sciences
September, 2021

Author's Declaration

I, **Hadiqa Nadeem** hereby state that my MS thesis titled "**Identifying Genetic Expression of Regulatory Regions in Human and Non-Human Primates Influencing Facial Feature Development**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Hadiqa Nadeem)
Registration No: MBS193023

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Identifying Genetic Expression of Regulatory Regions in Human and Non-Human Primates Influencing Facial Feature Development**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Hadiqa Nadeem)
Registration No: MBS193023

Acknowledgement

My deep gratitude first goes to my supervisor: Dr. Marriam Bakhtiar (Assistant Professor, CUST. Islamabad) who expertly guided me throughout my degree program and who shared the excitement of 2 years of discovery which kept me constantly engaged with my research.

My appreciation also extends to my Co-supervisor Dr. Rabail Zehra (Assistant Professor, NUMS. Islamabad), Dr. Babar Jamal (Assistant Professor, NUMS. Islamabad) and Dr. Faheem Khan (Assistant Professor, NUMS. Islamabad) for continuously mentoring and helping me out in every situation. Above ground, I am indebted to my family whose value to me only grows with age.

Thank you so very much to the rest of all for being part of this journey.

(Hadiqa Nadeem)

Abstract

The development of genome technology has opened up new possibilities for comparative primate genomics. Non-Human primates' genomes provide vital information into the genetic similarities and variances among species utilized as disease models. Facial musculature produces facial expression, which is a form of remotely comparable non-vocal communication employed by non-human primates. Primates have some of the most complicated face musculature and generate the most complex facial exhibits of all mammals. This study outlines current understanding about primate genome content and dynamics, as well as a set of short-term objectives to link unique genetic variations to common facial attributes that can help us better to understand normal and aberrant facial feature maturation, generate prospective estimators of evolutionary impact on human facial expressions, and increase our expertise to reconstruct forensic faces from DNA.

Contents

Author's Declaration iv

Plagiarism Undertaking v

Acknowledgement vi

Abstract vii

List of Figures xii

List of Tables xiii

Abbreviations xiv

1 Introduction 1
- 1.1 Aim of study . . . 8
- 1.2 Objectives of study . . . 8

2 Literature Review 10
- 2.1 Chromatin Looping Mechanism . . . 12
- 2.2 Significance of Facial Expressions in Primates . . . 14
- 2.3 Continuity of Phylogenetics of Facial Expressions in Primates . . . 15
- 2.4 Facial Processing in Human and Non-Human Primates . . . 16
  - 2.4.1 Neurological Influence on Facial Movements . . . 17
  - 2.4.2 Dominancy of Emotional Influence on Brain . . . 18
- 2.5 Primate Facial Muscle Identity Recognition and its Evolution . . . 18
  - 2.5.1 Gestating the Primate Facial Musculature . . . 19
  - 2.5.2 Profiling of Primate Facial Musculature . . . 21
- 2.6 Emerging Patterns of Facial Musculature and its Dynamics . . . 23
  - 2.6.1 Genetic Variation Among Primate Species (Facial Musculature) . . . 24
  - 2.6.2 Dysmorphic Morphological Changes (Facial Musculature) . . . 25
- 2.7 Importance of Transcriptomics in Primate Studies . . . 26
- 2.8 Importance of Enhancers in Primate Genomics . . . 27
  - 2.8.1 Enhancers in Eukaryotic Genome . . . 28

2.8.2 Chromatin Template-Closed Compaction . . . 28
2.8.3 Enhancer RNA - eRNA . . . 29
2.8.4 Impact of Chromosomal Translocation – Genetic Variations 30

3 Materials And Methods 32
3.1 Extraction of Genes Involved in Human Facial Morphology . . . 32
3.2 Extraction of Human Facial Enhancers . . . 33
3.3 Filtration of Human Enhancers (NOSE and EYES) . . . 33
3.4 Formation of Ortholog Sequences – Non Human Primates . . . 34
3.5 Identification of Transcription Factors and Binding sites (NOSE and EYES) . . . 34
3.6 Manual Inspection of Human Specific Transcription Factors (Muscle Specific) . . . 35
3.7 Manual Inspection of Human Specific Transcription Factor Binding Sites . . . 35
3.8 Elucidation of Genetic Variations . . . 35
3.9 DNA-Protein Docking for Human Specific Structural Analysis . . . 36
3.10 Overview of Methodology . . . 36

4 Results and Discussion 38
4.1 Extraction of Genes Involved in Human Facial Morphology . . . 38
4.1.1 DCHS2 Gene . . . 39
4.1.2 PDE8A Gene . . . 40
4.1.3 SCHIP Gene . . . 40
4.1.4 ASPM Gene . . . 40
4.1.5 DLX6 Gene . . . 41
4.1.6 EDAR Gene . . . 41
4.1.7 HOXD1 Gene . . . 42
4.1.8 WDR27 Gene . . . 42
4.1.9 PAX3 Gene . . . 42
4.1.10 TP63 Gene . . . 43
4.1.11 PABP1 Gene . . . 43
4.1.12 TMEM163 Gene . . . 44
4.1.13 COL17A1 Gene . . . 44
4.1.14 PARK2 Gene . . . 45
4.1.15 OSR Gene . . . 45
4.1.16 FOXA Gene . . . 46
4.1.17 MAFB Gene . . . 46
4.1.18 TRPC6 Gene . . . 47
4.1.19 TBX15 Gene . . . 47
4.1.20 RPS12 Gene . . . 48
4.1.21 FREM1 Gene . . . 49
4.1.22 HOXD Gene . . . 50
4.1.23 PKDCC Gene . . . 50
4.1.24 C5orf50 Gene . . . 51

4.1.25 EPHB3 Gene . . . 51
4.1.26 ZNF219 Gene . . . 52
4.1.27 SOX9 Gene . . . 52
4.1.28 CASC17 Gene . . . 53
4.1.29 KCTD15 Gene . . . 53
4.1.30 PRDM16 Gene . . . 54
4.1.31 DHX35 Gene . . . 54
4.1.32 SUPT3H Gene . . . 55
4.1.33 ALX3 Gene . . . 55
4.1.34 GL13 Gene . . . 56
4.1.35 RAB7A Gene . . . 57
4.1.36 MBTPS1 Gene . . . 58
4.1.37 PCDH15 Gene . . . 59
4.2 Extraction of Human Facial Enhancers . . . 60
4.2.1 VISTA Enhancer Browser . . . 60
4.2.1.1 Ultra-conservation Properties . . . 61
4.2.1.2 Types of Enhancers . . . 61
4.2.1.3 Experimental Dataset . . . 62
4.2.1.4 Technical Background - Directions . . . 62
4.2.2 Dataset Retrieval . . . 63
4.3 Formation of Ortholog Sequences . . . 64
4.3.1 Ensembl Genome Browser . . . 64
4.3.2 BLAT (Blast Like Alignment Tool) . . . 65
4.3.3 Dataset Retrieval . . . 65
4.4 Transcription Factors and Binding Sites . . . 65
4.4.1 TRANSFAC-gene Xplain Database . . . 66
4.4.2 Match 1.0 Public . . . 67
4.4.3 Dataset Retrieval . . . 67
4.5 Generation of Human Specific Data . . . 68
4.6 Elucidation of Differential Genetic Variations . . . 70
4.7 DNA – Protein Docking . . . 71
4.7.1 HADDOCK 2.4 . . . 72
4.7.1.1 Scoring Function . . . 73
4.7.1.2 Input Parameters . . . 73
4.7.1.3 Output Parameters . . . 74
4.7.1.4 PDBe PISA . . . 74
4.7.2 Dataset Retrieval . . . 76
4.7.2.1 Docked File 141-OCT . . . 76
4.7.2.2 Docked File 141-OCT . . . 77
4.7.2.3 Docked File 156-PAX . . . 78
4.7.2.4 Docked File 1122-Oct . . . 78
4.7.2.5 Docked File 1833-Oct . . . 80
4.7.2.6 Docked File 2328-Oct . . . 81

5 Conclusion and Future Prospects **84**

Bibliography **92**

Appendix A **92**

List of Figures

1.1 Summary of activation of gene transcription. 3

2.1 Reported facial genes that are involved in facial morphology. 11
2.2 Chromatin looping well explained during transcription mechanism. . 13
2.3 A simplified view of primate phylogeny. 20
2.4 Human Facial Musculature System. 22

3.1 Methodology opted for this study. 37

4.1 Docked Complex (DNA-Protein) for Enhancer 141-OCT. 76
4.2 Docked Complex (DNA-Protein) for Enhancer 141-OCT. 77
4.3 Docked Complex (DNA-Protein) for Enhancer 156-PAX. 78
4.4 Docked Complex (DNA-Protein) for Enhancer 1122-Oct. 79
4.5 Docked Complex (DNA-Protein) for Enhancer 156-PAX. 79
4.6 Docked Complex (DNA-Protein) for Enhancer 1833-Oct. 80
4.7 Docked Complex (DNA-Protein) for Enhancer 1833-Oct. 81
4.8 Docked Complex (DNA-Protein) for Enhancer 2328-Oct. 82
4.9 Summary of Research Project. 83

List of Tables

5.1 List of reported genes enrolled in facial morphology along with their associated functions. 92
5.2 Functional description of extracted facial genes 95
5.3 Dataset retrieval of VISTA ENHANCER (5 OUT OF 103) 97
5.4 Dataset of 5 out of 103 Ortholog sequences made by Ensembl . . . 98
5.5 Dataset of 5 out 103 Transcription factors along with binding sites made by TRANSFAC-geneXplan database 100
5.6 27 major Transcription factors involved in Nose and Eyes (Human specific) 107
5.7 Human Specific Transcription factors (definite phenotypic expression) 109
5.8 Summary chart of PDBe PISA analysis parameters 110
5.9 PDBe PISA summary chart for individual docked interfaces 113

Abbreviations

ASP	Abnormal spindle genes
ASMP	Assembly Factor For Spindle Microtubules
bZIP	Basic Leucine Zipper
CDD	Cleidocranial dysostosis
CNCC	Cranial Neural Crest cells
GO	Gene ontology
GWAS	Genome Wide Association studies
HLS	Human Lineage specific
NCC	Neural Crest cells
PDE	Phosphodiesterase
RBH	Reciprocal Best-hit
SHH	Sonic hedge-hogs
TF	Transcription factors
TFB's	Transcription factor binding sites
TRP6	Transient Receptor Potential Cation Channel Subfamily

Chapter 1

Introduction

Facial Morphology is a very complicated assembly that displays a very amazing interaction with genes, as well as extremely varied distinctive anatomical components. The facial surface basically comprises of coordinated complex of skeletal structures which can be later on recognized at gestation period (4th week). The growth, structure and patternation of facial features in under the control of some essential growth factors, integrated proteins and morphogenetic gradients. Developmental facial processes can be characterized by distinctive features like; head shells, nose patterns, chin nodules, lip grooves, eye orientation and much more. These protrusions have decorous relation with genes which made molecular studies to have several important clinical applications related to Facial Morphology.

Human genome contains several regulatory regions which contribute towards the dynamic genomic functions in order to determine how, when, where and at what level the respective gene is to be expressed. Keep in mind that the total skeleton of the facial and the connective tissues "NCC cranial neural crests cells" are tightly coordinated, essentially spatially positioned tissues / structures that influence morphological features of face" that can later be fused and proliferate into face [1].

Facial feature development involves multiple genes that can have influence on more than one particular facial trait, this refers to the existence of facial variations

as per according to associated genetic effects. The DNA has the potential to recognize the candidate facial feature and associate it under the localized gene expression that can affect 2 variant kinds (I) Variants of coding: congenital facial feature abnormality and/or syndrome. (II) Variants that do not code: typical face/ gene expression range; These genetic variations that influence face morphology can influence the brain independently, disrupting maturation and leading to long-lasting disorders in the central nervous system [2].

Facial asymmetry, height, width, protrusion and aperture all fall into the gene orientation, so the variations are the effects of mutations, genetic drifts and the complex phenotypical variants that modify genetic interactions and affect gene function. Now, here is the question whether there is a sequenced facial gene set which has distinct phenotypical effects, then why are there different facial characteristics? The answer to the question above is straightforward: There are various conditions focusing on phenotype and genetic variants encapsulating facial features with respect to genetic additives, environmental (epigenetics) effects and other associated medical histories, giving an insight into this query that besides genetic impacts, there are highly affected environmental factors that contribute to facial shape and characteristics [3].

The process of face development begins in the early stages, given that growing fetuses become more environmentally friendly and other lifestyles (gestation period). The overall face variance depends on the amount of the GM in the whole process (on a scale from normal variations up to the variations that leads to abnormalities). (I) Germline genetic variations: facial phenotypic influences through genetic paths" according to GWAS-genome large association studies. "EWAS-epigenome large association studies" and (II) Epigenetic variations: moments that lead to DNA methylation. The stage at which epigenetics play a key role in the actions of genes (pre-natal growth) and change the process of DNA methylation and historical modification, acetylation, phorylation and chromatin modelling are also converged in facial feature-phénotypic recombinations [2]. Although face morphology varies in a person's development throughout life (adult birth growth). The eukaryotic transcription works are divided into three key steps; I initiation, which

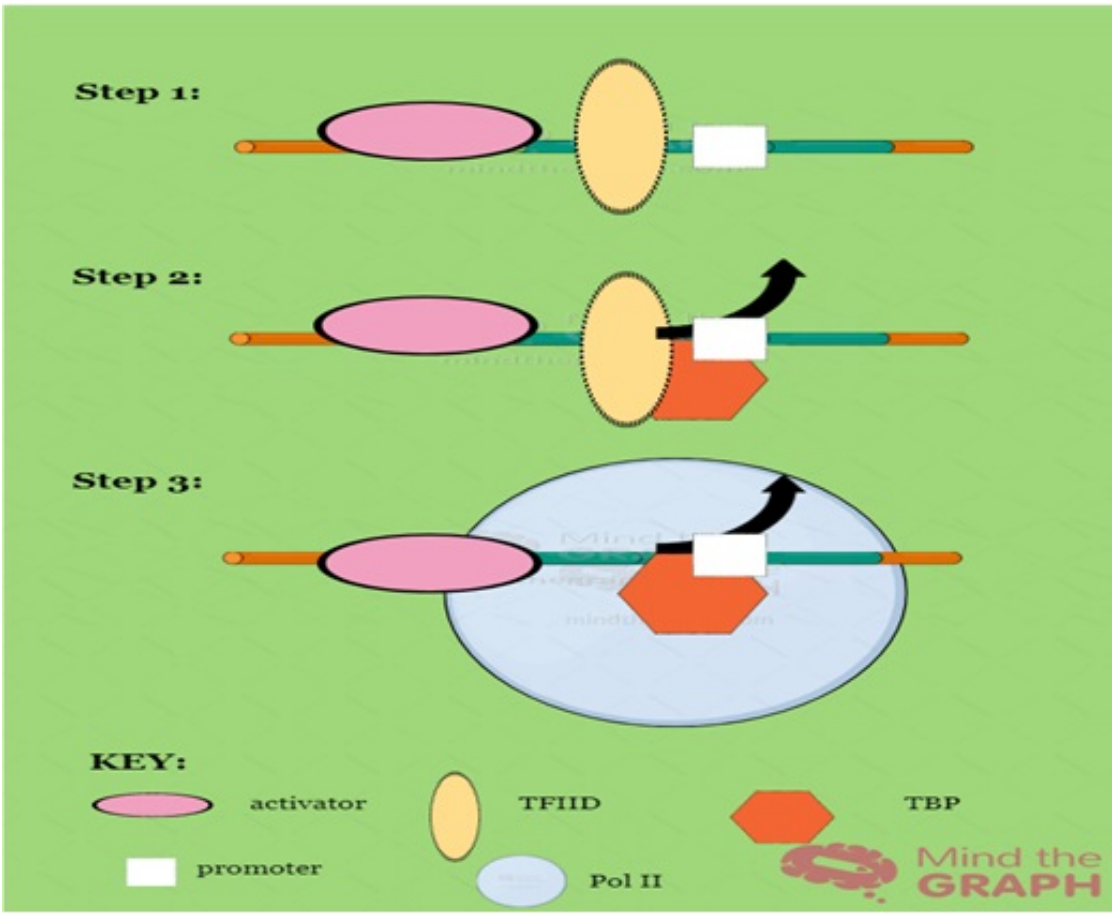

FIGURE 1.1: Summary of activation of gene transcription.

is supported by the end of the elongation of (III) by promoter clearing. In the earliest stage of synthesis and maturation of mRNA transcription factors are necessary. Multiple factors such as TFIIA (the PIC pre-initiation complex), TFIIID, TFIIE, TFIIF and TFIIH which are subsequently transcribed by Polymerase II are included. There is a great deal of emphasis on the statement that the temporal and spatial characteristics of genes are so important in the transcription process. For transcriptional activities, the correct time and defined specificity of each gene are necessary. The promoter comprises DNA fragments, which were later identified by Pol II (binding of general transcription factor TFIID with tightly associated protein complex TBP-transcription binding proteins). TFIID uses mediators to recognise the initiators (Inr) and the promoter elements (activators-Act) and package the DNA into a highly structured chromatin structure which is transferred during cell divisions to daughters' cells. The transcription machines should be linked to the central developer to start the process and in this orderly fashion

gene activation has been done Transcriptional activation of gene in a summarized model has been shown in (Figure: 1.1).

The emphasis on human facial morphology is very favourable; thus it serves the aims and essential information concerning the underlying facial skeleton, the facial feature skeletal dimensions, neurocranium understanding, as well as highly integrated, complicated interrelationships with genetic and phénotypical terms. The quantitative facial characters can lead to thousands of considerably more informed conditions, by which facial features can be thoroughly examined and the resulting genetic variants are significantly interrelated.

When discussing the significance of genetic variants, we need to be aware that precisely what "variation" means: in essence, some environmental inputs as well as the genetic contribution may modify the times and the expression of this particular genetic mechanism. When we state that between closely related individuals there is a greater degree of resemblance, there must be certain features (quantitative or qualitative) that can generate such a circumstance.

The genetic connections of face measures are heritable; this means that every single gene has its own particular heritage aspect (the result in the diversity of subject matter). Consequently, the biological mechanism "finishes" on the genetic and environmental factors simultaneously. Environmental components effect and integrate into the genetic aura and change the genetic make-up, which includes medications, vitamins shortcomings, cigarette smoking, poor nutrition, alcohol consumption, exposure to increased radiation (UV-range). A large amount of the cause of face shape variation initiates, in part, with the total relevant genetic diversity, which in particular causes imagery, phenotyping and confounding factors in people [1].

So far over several years, genetics of human facial development have been explored. Recent developments are progressing in such a way that underlying morphological investigations are now viewed as Hominids specific development studies which also have their foundations in taxonomic orientation. Human facial morphology has recently made progress because to in brain and progress in genetic architecture.

The entire facial feature skeleton face of the same taxonomic group has the same dimension of genetic heritage. There are a number of reports already on considerable amounts of facial measurement data, such as photographic studies, cephalometric (2D-3D) imaging and now changes of focus/dimensional/objective/vision or just vision so that the genetic and phenotyl relationship of all human facial morphology can be found and developed [1].

The large ongoing study discloses some facial feature disorders, which are mainly due to SNP mutations, but are defined as underlying disruption in the skeletal or soft tissue. It has the consequence that the growth, dimensions and form of the face characteristics alter. These genetic changes focus on the Mendelian pattern of transmission to generations; (I) Cousin Syndrome -TBX15: pelviscapular dysplasia with anomalies of the epiphysis, congenital dysfunction and face dysfunction, frontal bossing, hyperteloric, thin palpebral splits, deep set globes, strabism, slightly rotated posterior. (II) The otofaciocervical PAX1 gene: prominent, dysmorphic ears (low-set and cup formed with large, concha and hypoplastic porpoises, antirrhydrates and lobes), long neck, preauricular and/or branchy fistulas and/or kysts, hypoplastic cervical and sloping muscles. The gene is characterised by distincted facial properties: long triangular face, wide front, narrow-nose and lower-nose, highly arched palate. (III) Waardenburg-PAX3 gene: main traits often include significant facial abnormality; unusually decreased hair, skin and/or iris (irides) coloration; and/or congenital deafness. Waardenburg-PAX3 gene:

In particular, some persons affected may have an unusually widening nasal bridge because the inner angles (canthi) of the eye-dystopic canthor are displaced sideway (lateral). (IV) Char-TFAP2B gene syndrome: the disease affects the development of the face, heart and limbs. All the features of this condition are the odd facial look, the heart condition called the patent ductus arteriosus and the deformation of the hand. (V) Current development before and after the birth, leading to short stature; cognitive handicap ranging from mild to severe; and abnormalities of the arms, hands, and fingers in bones, distinctive facial features such as arched eyebrows that often occur in the centre (Synophyrs syndrome), lengthy eye clots, low ears, a small and widely spaced teeth combined with a smooth size (VI) Gene

GNAI3-Auriculocondylar: disorder that impacts face development, especially ear and lower jaw development (mandible), A little ear canal, minor skin tags, background ears flipped with large cheeks, a small mouth, microstomatic asymmetry, the facial structure of both sides of the face asymmetry (facial asymmetry), and a split palate with minimal curves, ears and groovs, and small ear canals (a gap in the roof of the mouth). (VII) Frontal dysplasia-ALX3 gene: Unique, unique facial feature condition marked of poor mid-face structures with or without a split palate with the addition of a bilateral and unilateral split lip [4].

Facial feature biology focuses on genes and processes which lead to embryonic development of the entire face-skeleton and connective tissues produced from NCc-neural crest cells (the transient group of cells that is unique to vertebrates that arise from the embryonic layer of the germination of the ectoderm and in turn lead to different cell lines, including melanocytes, facial feature cartilages and smooth muscle bones. They give rise to a variety of organs, including salivary glands, lachrymal glands, the thymus, and the thyroid. The arteries of the dorsal aortic arch are likewise formed by neural crest cells. Tissue gives rise to arteries and muscles. These cells move, differentiate, and proliferate into cartilage or bone (facial prominence). A multitude of internal and extrinsic signals from adjacent growing cells tightly govern this entire procedure [2].

The complex examinations at the molecular level have confirmed that PRIMATES' cognitive talents have a much larger potential than any other species (representation of conserved mammalian genome). When it comes to evolution, there may be more than 90% resemblance across all primates, which is why they've been labelled as "Ortholog," meaning they gain genes from a common ancestor while maintaining the same function while maintaining speciation and reliability. The increased corticogenesis in primates (the process in which the brain's cortex is generated throughout the development of the nervous system) leads to the spatial and temporal expression of genes in a clearly defined box in which humans take the lead because of their more cognitive capabilities than others. The reasons why humans are more varied than the rest of the orthologs are transcription factors, their binding locations, and regulatory areas [5]. Approximately 90% of variably

expressed genes in the human genome lineage are found in the brain, emphasizing the importance of the brain area and the genomic changes linked with developmental mechanisms.

To fully comprehend this unique predicament, researchers must study the biological mechanisms of cranio-facial development and how they illuminate the genetic architecture in other closely related genomes. Although the genetic regulatory systems are complex, it is comforting to know that the human genome shares a high degree of commonality with the genomes other ethnic taxonomical groupings (orthologs). There has been a lot of focus on facial morphology and the anomalies that come with it, but the underlying biological variety must also be considered. Understanding the genesis of Human Facial Morphology in comparison to other primates will allow us to go deeper into the subject and proceed with more intelligible scenarios. Information about a person's facial morphology has a variety of clinical and forensic applications, including informing person-specific models, reducing the need for broad surgical interventions (for facial anomalies), facial form renovation, skeletal remains prediction, and detection of suspect DNA [1].

In primate speciation, functional characterisation of genetic differences is thought to be critical. Comparing primate genomes reveals fresh information on biological mechanisms that govern primates' interactions with one another; Primate genomics. Comparative genomics, human and not human primates, allow us to uncover shared or particular, species-specific genomic features to show phenotypic differences or genetic similarities. When the neural crest cells were compared and examined by both species; some areas of these regulation components differed in one species from other, which suggests that they are more active in one species than in the other, namely "species-oriented regions." Because of the varied manifestations of the same regulatory regions, this explicitly endorses the question of "why we look the way we do."Today's large-scale technology to sequence DNA opens up new opportunities for non-human primate genomes. Though the main focus of genomics research is on human genetics and their relationship to disease, researchers also want to see comparative genomic primates. There are two major motivations for a systematic study of non-human primate genomes.

The application of this research knowledge in the use of primates as model for analysing human condition; and comparative developmental analysis which reconstruct the history and processes of genomic change with an emphasis on human genome origin [6].

This poses the question of why transcription factors are not to be compared with their binding locations and the corresponding phenotypical changes are not to be seen alongside the effect on gene expression between two species if the changes in enhancer sequences occurred in recent developments.

The ancestry and physical appearance are very connected, and the observable characteristics of individuals may result from conceivable evolutionary manifestation in recent genes. The way in which sites binding on transcription factor change and keep the link between evolutionary lines illustrates the varied differential affinity for species.

1.1 Aim of study

The main goal of this research is to uncover all of the transcriptional processes that are strongly linked to facial feature development in humans and non-human primates, as well as their aesthetically current phenotypic characteristics.

This research will focus on human facial enhancers and their respective orthologs. The goal of this research is to identify areas in human facial enhancers with SNPs inside the transcriptional factor binding sites that have functional implications (non-coding variants).

1.2 Objectives of study

This research will highlight about facial morphology, medical conditions and associated genes which infer etiological link of individual's genetic data mostly enhancers and their link with genetic variants upon ancestral differences.

1. Identification of key regulatory regions in genetic control involves in facial feature features among Human and other primates.

2. Comparison of cis-regulatory elements among Human and Non-Human primates (Lineage-specific variants).

3. Elucidation of genetic control of facial morphological features among human and other primates upon comparison on transcriptional processes.

4. Functional characterization of key regulatory regions in Human and Non-Human primates; upon comparison with positional and structural placements.

Chapter 2

Literature Review

Understanding the genetic architecture of face morphology has major applications in both treatment and diagnosis in medicine; correct corresponding information of certain candidate genes which play their function in facial morphology must be established. The normal face development depends entirely on crest cells of the NCC and the fusion of facial processes depends on the sequence of events that combine cell migration, proliferation, connectance, differentiation and death. Fusion disturbance can lead to total or partial anomalies in all facial processes. In the cellular attachments with the linked genes the epithel pioneer (dermal layer) is involved. The genetics have their relevance in the face posture of an individual; they were examined for their facial cover, facial precision, face pattern, reliable facial grooves (including protrusion of eyes, nose, lips, cheeks and chin). The features of the component such as face distances, ratio and angle, face shells and signatures can be further explained in these points [2].

This study has a major influence on earlier facial details (face height, relative prominence, facial symmetry, facial width, front and psychic folding, form of eye and depth, width of eye and body curvature, nasal width, height and shape between the eye and nosal bridges, chin protrusions and lip demarcations). These tissues were further investigated for their related genes and their functional significance in the formation of facial feature. Each gene or combination of genes for a specific phänotype has recognised functions with a certain scale. There are around

30 key genes that were involved in the anterior development of the face and their functions and phenotype have been displayed in Figure 2.1.

In the presence of transcription factor (TF), gene regulatory networks begins cell-definite expression. These factors in transcription require support (I) Direct interaction; adjoining binding of factors (coding regions). (II) Indirect interaction; Cis-regulatory regions dependence (non coding regions). In certain anatomical locations the presence of cis-regulatory regions controls the manifestation of developmental genes in different timeframes. The genome consists of 3.2 billion nucleotides packaged in histone units (chromatin) which have the function of activating and controlling the associated expression of periodic genes in these cis-regulatory areas in normal craniofactant development.

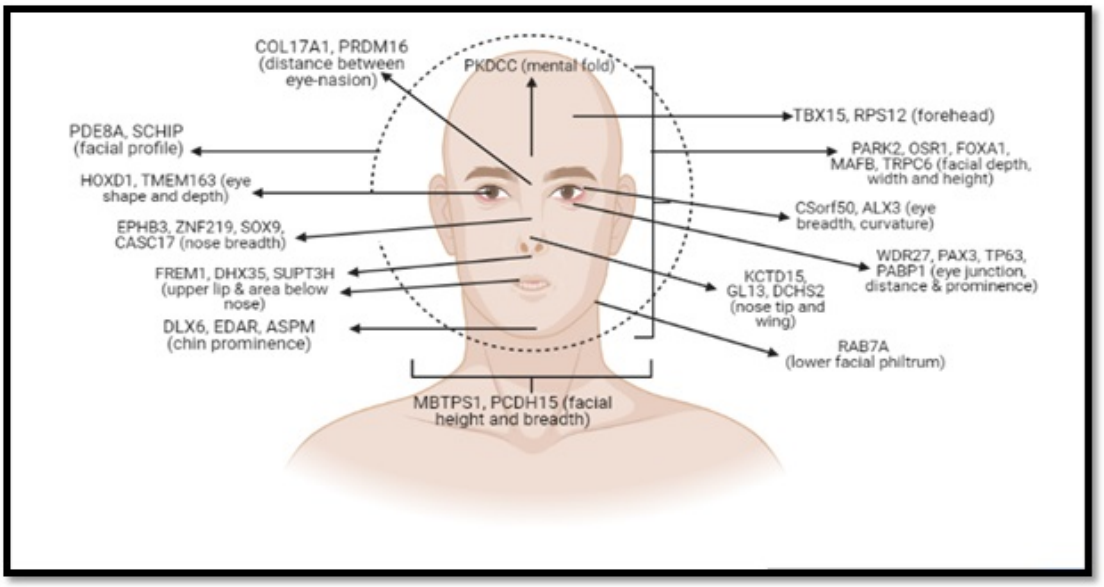

FIGURE 2.1: Reported facial genes that are involved in facial morphology.

The transcriptional beginnings take place in the chromatin function and when the transcription mechanism is initiated by many regulatory areas which have been dispersed throughout 98 percent of the human genome. and thus named as "ENHANCERS" [5]. They are quite variable in their position and persistence, which may be either close to the gene or far from the gene. They can be altered by physiological, environmental and pathological factors, but are strongly preserved. DNA loops can be used to express functionality and to work on a complicated chromatin framework.

The general hypothesis is that the enhancer-promotor looping is used to supply the relevant tissue at the appropriate time with factors (polymerase & transactivators). Since chromatin is flexible, the structured loop is controlled by the energy input that causes it to bend. Thus the total radius is enough to compact the 2 sequences within the range of 1Mb and still stabilize the mechanism.

2.1 Chromatin Looping Mechanism

The GTF's and RNA polymerase II are used to improve the pre-initiation complex (PIC). Long loop creation increases the ability of the component to offer PIC in close proximity to the promoter, as demonstrated in Figure 3. There is another complex termed the Mediator (MED) during the development of the PIC complex which mediates the upstream coactivators and polymerase II. The MED-complex interacts with the GATA (signalling machinery important erythroid transcription factor). It binds with the promoter, and begins transcription of specific genes to the associated expression for specifying the LCR (locus control region).

Enhancers are able to use their transcript to influence their host genome's transcript. The fact that the enhancer is regularly translated into non-coding RNAs is widely known; the special eRNA identifies the active enhancer that eRNAs are mainly part of mRNAs and that the entire transcription regulatory mechanism is based on the confirmation of the enhancer-promoter loop [7]. In extraordinary cell types, eRNA play an important part in the gene expression of the scaffold of different genes.

The enhancer loop then has its main function of providing a transcription mechanism with the activated gene; in case the migration and implications of activated gene interruptions could have not been achieved and abnormal reactions were brought about in the form of mutations/varyings or genetic syndrome in some cases. The importance of enhancers rests just in the fact that the regulatory tract of a certain gene consists of multiple enhancers acting in concert, with a higher sensitivity to functionality.

According to the World Enhancer GWAS Genome Association (GWAS), the overall reliability of the genes is 80% and every sort of single SNP polymorphism will cause the biggest part of variation/mutation as per the loci distribution input [7].

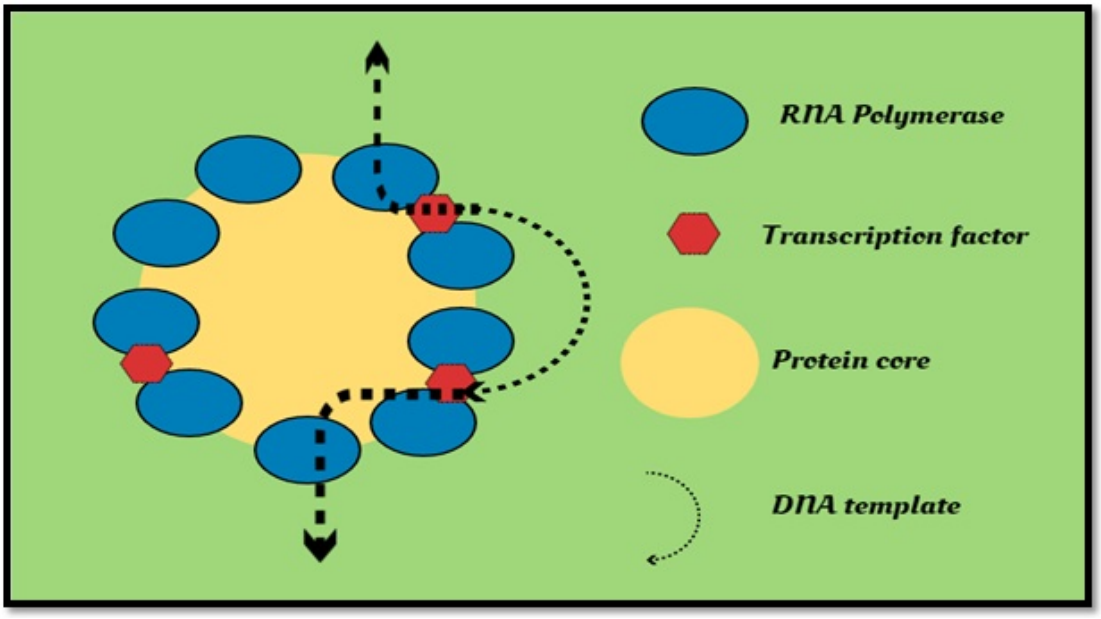

FIGURE 2.2: Chromatin looping well explained during transcription mechanism.

In order to regulate expression/activation of the particular gene, the role of tissue-specific enhancers is to operate as "nucleation" and construction of a montage of the crucial PIC complex. During the development, differentiation and cell cycle, the timing and position of gene activity is all critical for GTF in linking to promoters and regulating gene transcription in each and every specific tissue. In summary, the entire sum would be more than 30 genes according to their expression in the muscle tissues and their genotyping involves all of the reported genes which have been enlisted in human morphology.

The genetic and phenotypical facial data below assist us understand more about the tissue-specific gene expression. It is apparent that the GTFs have their own direct role and are considered as the main participatory component in genetic expression; moreover, the wide angle clear picture of the differentiation is easily formed for the very practical cause by digging into these functional modules more precisely connection and relatedness in between genes and particular mode of phenotypic gesture. A crucial detailed view of all the reported genes along with

their reported expression and function that are involved in the Human Facial Morphology at a particular time zone has been clearly shown in Table 2.1.

2.2 Significance of Facial Expressions in Primates

Primates are highly social organisms, as compared to most other species. For the most, if not all, of your life, primary life is spent in a social context. The size of the social groupings of Great ape's ranges from 2 to over 100,000 people, with considerable variations in diversity and stability of species. In this group environment, the person's maturity, development and premature knowledge all occur. At least portion of the time both primates spend with others to rise, develop, and learn. Also all primates, with a few exceptions, are sleeping with at least one other species, as opposed to most other species. This degree of social organization is not commonly found in other mammalian orders. Although some other mammalian lines, such as tamed hyänas, tigers, wolves, elephants, and many aquatic species, can very well be adapted to at least some part of their life, the primary order is the only one in which every species in everyday life is exceedingly social. The world in which they dwell communicates with both creatures.

The bulk do so in a cognitive way and have complicated morphological adjustments. Primate mammals are essentially general quadruped foragers and are, instead, the least phenotypically sophisticated mammals. One element that distinguishes primates from other animals is their cognitive and behavioral communication with their environment, which is a primordial cause in their interface [8]. Primates have incredibly high brains, which are known to be a trademark of how they handle different societal levels in comparison with their bodily levels.

Despite widespread adoption, the 'social brain' explanation of ape cognition is still contentious. One approach for primates to deal with these complicated social networks is to produce and process facial expressions as a method of close contact with conspecifics. Facial characteristics can produce a physical effect on the audience, elicit similar sensations and enable empathy to develop. The objective

of this analysis is to provide a proportional and ancestral knowledge framework; (I) the responsibility of facial-expressions in primate interaction and relational intelligence. (II) focus on facial production in humans and non-human primates.

2.3 Continuity of Phylogenetics of Facial Expressions in Primates

Darwin explored the pioneering role of "facial expressions," how human face changes and how this behaviour might have evolved, as well as his observations on the main factors in biological evolution that affected these attributes, in his pioneering study, "The Expression of Emotions in Main and Animal. Human face characteristics were developed from monkey antecedents, although Darwin was effectively anti-Darwinian in the argument that the human and non-human primate body language had separate meanings.

He argued that in all mammals, the ability to make face emotions is inherent and often involuntary. So every creature has its own repertoire; in many human civilizations and in people who are born blind or deaf, for instance, "universal facial expressions" are seen. Hatred, fear, joy, astonishment, grief and rage are among them.

A number of tests by the chimpanzee (*Pan troglodytes*) which have been conducted by our neighbourhood family member have been made known to help them to recognise their families through gazing at their faces. Certain human face motions, such laughing and smiling, are similar to those of non-human primates.

These results clearly show the role of facial expression in primates to develop as a significant tool in the operation of social groups and in face and non-human primates' way of acquiring and processing emotional information about the conspecific. Moreover, such talents can play a crucial part in the growth of sensitivity and morality in bigger primates (including human, Chimpanzee and Gorilla and less to *Macaca mulatta*).

Table 2.1: List of reported genes enrolled in facial morphology along with their associated functions. (refer to Appendix A Table 5.1).

2.4 Facial Processing in Human and Non-Human Primates

For humans and our nearest living relatives, Face is extremely essential, because it is subject to stimulations, as is the topic of this study Chimpanzee (Pan Troglodyte), Gorilla and Macaca, with newborns preferring facial signals and face signals in both species. Human people and nonhuman Great apes are unbelievably qualified to distinguish between familiar and unknown faces. Facial ontogeny or proper identification and marking begins considerably before human birth. Since human children can't voice their wants, they largely communicate by emotional cues. Human beings have to be able to differentiate between faces and their underlying feelings, and studies suggest that healthy infant development is linked to subsequent social skills. These growing abilities help the infant to properly respond to a situation, particularly in the light of the emotional conditions of the mother. A lot of time is used by newborns to scan people's faces and fast gain facial recognition skills, in order to identify facial expressions accurately by the age of 12 months. The facial dispensation of infants has evolved rather well, and only little different from that of adults, according to current studies [8].

Chimpanzees have highly evolved facial characteristics which, as widely reported, appear to be part of a broad and dynamic social platform. Little is known about chimpanzee, gorilla and macaca as compared with humans; the ontogeneity of their face processing. These non-human primates share 98% of the human genome and demonstrate a certain amount of emotions in their early lives, starting with smiles and furious emotions. Chimp babies (Pan troglodytes) must be subject to a conspecific in order to make best progress in their facial recognition. Multiple areas of the monkey brain help to properly detect a facial expression emotion. This is not the subject of identification, however, but the function of the indicator and

the unique identification of the individual face can be transmitted via a combination. According to various research, humans and chimpanzees carefully design the facial properties in the identification of faces and people. The facial configuration has two components. Relational attributes of the first order concern the relative arrangements of face features and are utilised to identify faces from distant felt objects [8].

The presence of two eyes, particularly that of the nose and lips, defines the faces in proportion. Some research indicate that people can be born with the capacity to use these skills. Human facial characteristics used to identify individuals are called relational features of the second order. The importance both of attributes for humans and non-Human primates has been illustrated with inverted faces in a number of tests, both reducing accuracy as well as minimising identification times. More evidence for facial processing is derived from human and chimpanzee research, which modifies facial composition by integrating the top and bottom parts of celebrities' faces into one composite face. In these trials, it was shown that the precision and speed at which interviewees indicated that each individual was identified in the group. Human and chimpanzee seem during interpretation to pay attention to the face layout, meaning that humans can inspect and interpret the face above and below differently. The fascination of this research is the fact that people have different cellular brains that govern the mobility of the face top and bottom. It is difficult to assign flow directions and "higher" or "subcortical" channels to each of these regions because of the recurrent and dynamic processing of the facial processing technology. The tiny framework for knowledge transmission across processes needed to create, on the other hand, accords these elements crucial responsibilities in facial recognition.

2.4.1 Neurological Influence on Facial Movements

The neurological pathways generate facial motion are the faces of the pons and the motor cortex facial area. Involuntary (emotional) facial motion is controlled by the facial nucleus, while freely moving is controlled by the face of the motor cortex.

According to current research, humans and large apes have the most primate face nucleus. These studies also reveal the reorganisation of the primary motor cortex faces from the orbital, insular, or other cortical regions, possibly strengthening cortical integration [9].

2.4.2 Dominancy of Emotional Influence on Brain

The processing of the human face is asymmetrical, with most emotional treatment on the right side of the brain. The hemisphere model shows that the right side of the brain possesses social communication capabilities which complement the left side's linguistic specialisation. Some scientists have proposed a "valence model" in which negative emotions are processed on the right part of the brain and pleasant emotions on the left. Inhuman primates, including chimpanzee and Macaca rhesus, also demonstrated patterns of emotional activation [9].

2.5 Primate Facial Muscle Identity Recognition and its Evolution

Muscles that emerge from the second (hyoid) arc of the pharynx and are innervated by the face nerve, the seventh cranial nerve, are called facial speaking musculature. This musculature is widely retained in the spinal kingdom. It can only be employed in particular mammalian divisions to communicate with the facial morphology. With a few exceptions, mammalian animals are the only organisms who live and participate in regular social relations in the great social classes. In a move away from supporting only the pharynx, the ring and hyoid (getting fed and ventilation), Mammalian imitated musculature spreads throughout the face and into structural evolutionary components, such as (a new auditory function). The mimic mammalian muscles link the dense dermis and mobilise the facial mask. Some visual expressions, as well as the sensory information, are included in the vibrissaes and outside ears of select species, such as friendly canids and felids [10].

Although lectures on mammalian facial musculature have been greatly needed, there is an increasing amount of information of Great ape's muscles. A working environment that explains the face musculature of monkeys (and certain mammals) was certainly referred to as a particularly significant base for improving knowledge of primates and general mammals about phylogeny. These muscular data can be combined with paleoecological information to help us to understand that human cognitive functions have arisen and the impact of the facial traits. Human progress has shown much about primate interactivity training and specialisation.

Although we can't presume that all primates follow the same direction, we can use non-human processes as a guide for primates in general. The only creatures with significantly complex facial expression patterns were humans and chimpanzees. However, in some Macaca and baboon animals complicated facial expression patterns were identified. Many experts have recognised their attached upper lip as an impediment to facial expression when examining the relative absence of sophisticate face motions in strepsirhine. People are still at the peak, trailed by chimpanzees, gorillas, orangutans and macacas, and ultimately galagos, one of the most prehistoric primates in existence. This common interpretation is mostly based on a species' evolutionary standing, with more advanced species having lesser behavioral and muscle complexity and higher species having higher complexity; as summarized in Figure 2.3. The hypothesis that biological and social variables also demonstrate that microanatomic studies of ape face muscles play a greater part than previously assumed in the genesis of primate facial muscle. The packaging of muscle fibres into fascicles shows details of the potential contraction of the muscle.

2.5.1 Gestating the Primate Facial Musculature

In human and non-human primates (chimpanzees, gorilla, and rhesus monkey (Macaca mulatta), almost same number of face muscles are observed, which is 9-10 common muscles, which is why the structure of the muscle is almost identical. In these three catarrhan species, the prominence of eyebrow fascia and nasolabia

reflects their ecological nicknames and social classes. All three (M. Mulatta, Gorilla and Pan troglodytes) live with complex, tough and convoluted social networks in enormous, spatially and temporally powerful factions in the biosphere and seem to be fairly advanced, possibly the greatest facial display configurations, This can incorporate eyebrow shapes with less emphasis on external ear movement, muscular upper deck and nasolabia area. Tight packed fascicles with few binding tissues can provide more strength, even more connective tissue than loosely packed fascicles.

Given that Chimpanzees, Gorilla and Macaca mulatta's facial musculature resembles that of human beings and all of these species have extremely complex social systems and repertoire of face expressions, what makes human facial expression different from others? One response could be found in the development and evolution of human speech and language characteristics. If we focus on all shared facial muscles of human and non-human primates, chimpanzee, Gorillas and Rhesus Monkeys. Comparative embryological investigations are necessary to understand how ape mimetic musculature develops and how this development is preserved. In the strepsirrhines, the basic living primates, all of these paths should be pursued. They could allow us to correlate generic mammalian face expression to its related musculature despite the limited amount of data in the specimens [9].

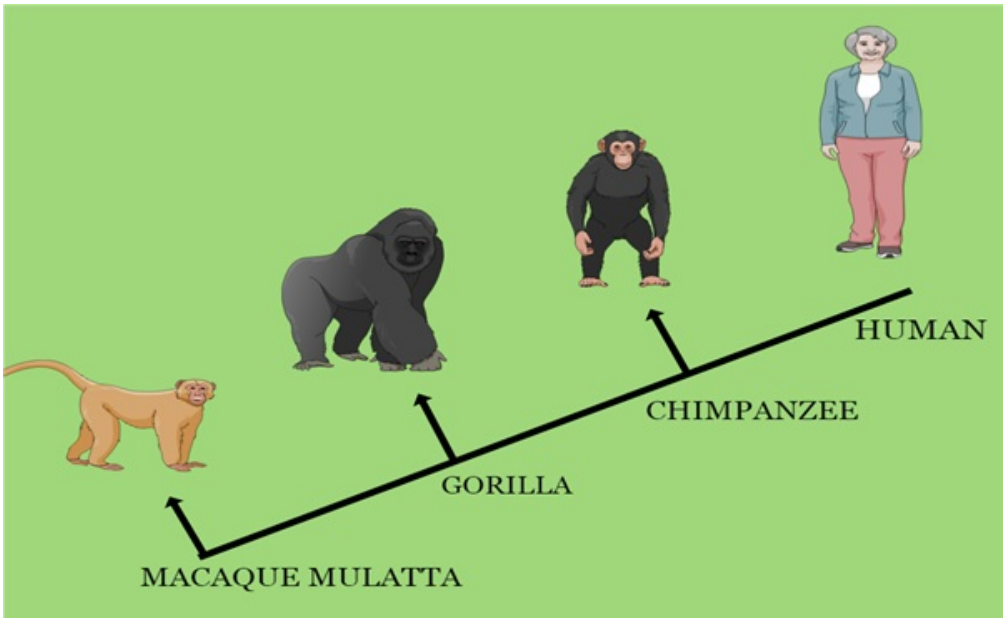

FIGURE 2.3: A simplified view of primate phylogeny.

2.5.2 Profiling of Primate Facial Musculature

In the HUMAN FACE shown on it, Gorilla and Rhesus monkey (Chimpanzee) exposes the common mark changes of the densely packaging fascilla in its associated specific contraction along with a few important prospective conclusions, so that the genetic monopoly behind the system is shown by the fact that they have only commonly shared facial muscles between human and non-human primates. In each and every muscle the classification is controlled by its genes or genes which expresses itself inside the transcription cycle in a particular time and in a very precise way (under the supervision of ENHANCERS-regulatory regions).

The lateral view of your head muscles is a single platysma muscle. There are two platysma muscles on both sides of the spine. Each is a big muscle plate, which spans most of the front neck of the side of the body. To expose the lower muscles, the platysma muscles are taken off, while the other anterior muscles are underneath them. Plastic muscles help the lower jaw to be pulled down (mandible.) Under the platysm there are two sternocleidomastoid muscles. There is one on per side of the neck. These muscles have two roots: first on the sternum and secondly on the collarbone. You join the mastoid process of the temporal bone.

They are able to flex or extend your head and move you to your shoulders. The epicranius muscle is likewise very wide, covering much of the head. Aponeurosis is entirely in the middle section of the epicranio muscle. The epicranius frontalis and the back of your head are the only sites on the organ where muscle tissue is present the front belly of the epicranius (the portion of the muscle called the epicranius occipitalis; sometimes called the occipital belly of the epicranius). If the buccinator muscles are tightened on either side of the face, the cheeks are squeezed. There are two masseter muscles on each side of the face. It closes as you tighten your mouth. Its name originates from the same Greek root as the chewing term. The main and minor zygomatic muscles on both sides of the face arise from the zygomatic bone. Both can raise the mouth to vary its form. As described above, between the humans and the non-human primates (Chimpanzees, Gorillas and Rhesus monkey (Macaca mulatta), there are almost comparable numbers of face

muscles), therefore the functions are also somewhat similar. For example, the chimpanzee orbicularis muscle has more packaged fascicles for fewer bridges and a bigger fibre thickness than humans. This is due to the regular application of the upper lip as a weapon for personal care and coercion and its significance in the facial toolbox of the chimpanzee. Although earlier microanatomical data in primary facial studies have not been studied, it is clear that this can really help us to understand the application of certain muscle types in distinct creatures. More than one histochemical study has shown fascinating data about fibre distribution in human facial muscles.

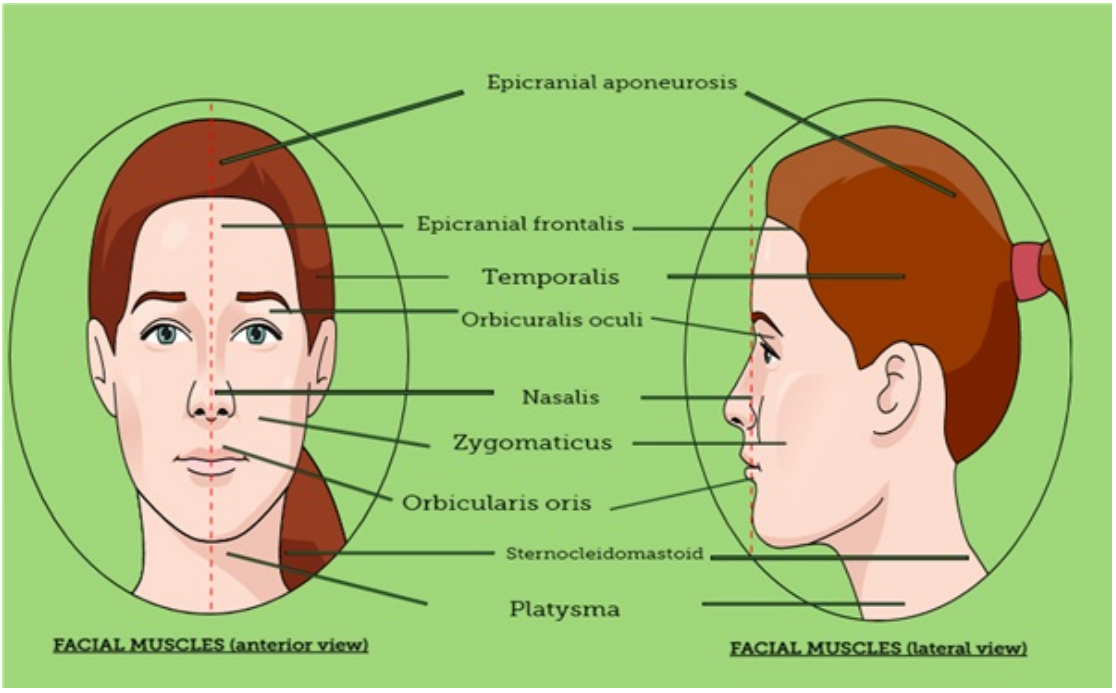

FIGURE 2.4: Human Facial Musculature System.

These muscles are shown to contain a higher proportion of type II (fast-twitch, tired) fibres compared to the limb and trunk muscles, as study shows. The higher rate of type II fibres implies that these muscles should contract quickly and transfer a particular area of the face mask without maintaining contractions, particularly in humans, for prolonged periods. Muscular spindles, which are stretch receptor organs dispersed throughout muscular fascicles and which govern the fibre length, were certainly detected in the facial structure of human beings. They are present in the muscle of the mammalian limb/trunk and in articular muscles, but in human mechanistic muscles [9].

These findings are especially notable because of their rapid neurological processing time described for human face, particularly in combination with the dominance of type 2 fibres. This occurs when there is a rapid muscle contraction for a very short time and the contraction is unnecessary. The focus of this study is on EYES and NASOLABIAL areas, with the expressions of the faces obviously more than two specified regions within the categories (Human and Non-human primates).

2.6 Emerging Patterns of Facial Musculature and its Dynamics

The human face is a spatio-temporal characteristic comprising of distinct features of dimensions, form, and structure (e.g., eyes, nose, chin, and mouth). Researchers still do not know, however, which genetic variances generate human face variation. This work focuses on functional chromatin signatures in the human neural crest cells, a type of foetal cell which eventually leads to a large number of facial feature characteristics (which are mostly governed by transcription factors). Due to the importance of facial feature formation and the need for contact between different types of cells, the embryonic colony occurs during 3–6 weeks of gestation plays a critical role in the creation of the face plan and determining species-specific and individual variation. The CNCCs migrate away from the neural tube during development, forming a mass of cranial mesenchyma that then turns into facial bone, cartilage, and connective tissue. In non-coding sections of the genome, the vast amount of biological change related to humans segmenting research and a set is found. Variants of these are believed to be made available in cis-regulatory elements. Conversely, the cis-regulatory elements, in particular remote enhancer, show cell-type specific activity patterns (in cell-type-specific regions of chromatin labelling) that emphasise the requirement for cell-type activities in the evaluation of genotype.

Revealing the connection between human facial alteration, disease and regulatory element improvement that differs from that of Chimpanzees (Pan Troglodyte), the

Gorillas and the Rhesus monkey (Macaca mulatta) raises the exciting prospect of gene mutation affecting the species-specific and individual facially produced mixed results within an intertwining set of loci/regulatory elements. The importance of GENETICS is underscored by the fact that factors of transcription and enhancers have a significant influence in determining the future of certain phenotype fields. More biological study of the human face characteristics could lead to new discoveries of facial feature morphology, an understanding that in facial feature disturbance and birth problems there is a tight link between genotype and phenotype, and consequently, a justification to attempt in several ways to forecast facial characteristics, ranging from early provision to medicine.

2.6.1 Genetic Variation Among Primate Species (Facial Musculature)

Complex signaling pathways and accurately time-taken expression of embryonic genes as well as molecular and cellular activities affect the facial feature structure. As people get old, hormones, dietary and biomechanical variables all influence the face. As a result of this, the facial structure is modular, with facial characteristics at various levels which indicate the efficacy of the organization but stay relative separated from other aspects.. This is predicted to affect the human side of numerous genes, some of which only affect the localized portion of the face. The conceptual face segmentation is therefore identical to the concepts of modularity and convergence in morphological research, with the main variation is the hierarchy of partitioning with globally integrated and locally focused modules. This allows the study of the consequences of facial form to spread to different scales.

The global and local face patterns of the discovered loci can help clarify its significance in facial feature development. Many of the genes in these loci are manifested during embryo development in associated tissues, which is clear. Early expression and patterns of function with possible morphogenesis are a big difficulty. The type of finely tuned facial effects demonstrated here help clarify molecular-morphological linkages. Consider two genes related to various characteristics of

the nasal form, such as the first gene (KCTD15), which have a very focused nasal tip influence. These very technical phenotypical effects could show many things about how these two genes function during the formation of human facial morphogenesis. The KCTD15 control of TFAP2A, a gene involved in the creation of neural crest, was shown to lead to a lower snout length, among other problems, if mutation into mice. SUPT3H affects the nasal form of the maxillary and narrow nose portions and KCTD15 controls the shape of the nasal tip in humans as cartilage is influenced by the septum proliferation [10].

2.6.2 Dysmorphic Morphological Changes (Facial Musculature)

Many hereditary dysmorphic side disorders are distinguished by modest morphs, frequently with quantitative characteristics with continuous distribution. The range of differences between the affected people and healthy ones for each certain face feature often greatly overlaps. Understanding the genetic elements contributing to normal face feature variation could influence the development of the causes of CFD, for example orofacial keys, common facial feature birth abnormalities.

Both research identified a relationship between PAX3 and interorbital variants, an interesting discovery given that PAX3 mutations produce Anatomical changes in interorbital areas; Waardenburg type 1 syndrome characterised by high altitude and various morphological anomalies. Both research identified strong relationships with nasal projection measurements in their cohorts, but various genetic areas were involved. In addition, several other loci gene studies involving facial feature or facial feature developmental pathways have related one or more facial feature measurements or formal characteristics to a small number of genes. At least three study candidates suggest a minor link between standard FGFR1 and normal variance in facial feature morphology, however each investigation has considered a distinct set of characteristics. In the two earlier GWAS studies on face shape, none of those genes discovered, including the FGFR1, were included. Although earlier research has found a small number of biologically feasible genes

connected to facial feature variations, these attempts simply scrape the surface and much more can be learned [11].

2.7 Importance of Transcriptomics in Primate Studies

This study focuses on genetic links between mechanisms of transcription control and its phenotypes and human anatomical differences. The extraction and subsequent identification of associated polymorphisms in humans compared to non-human primates of cranium facial-encrypted genes provided us with a chance to address the issue of where the distinction rests. A genetic relationship between facial development genes and quantitative features defines the essential components of the mean facial feature complex is demonstrated.

These associations can help us understand the wide range of phenotypic expression and seriousness experienced in particular extraordinary genetic diseases, as well as enhancing our overall understanding of those elements which support the many face features that we observe in people. Dysmorphic phenotypic expression in such settings may play a function in variants in a range of genes or regulatory elements.

Similar connections in non-human primates can also help to find indications of the aetiology of facial feature abnormalities which are even more typical [9]. For example, three of the features mentioned here (crane base width, NO width and intercantal broadness) have previously been involved in orofacial clefting, the most common facial feature birth abnormality in humans, as potential phenotypical risk factors.

The detection of genes that influence these characteristics could result in the discovery of important clefting hazards. The candidate gene was therefore extracted from the genetic material of non-human primates that associated with specific phenotypic features, with the explication as to where the disparity lies. The introduction of genetics and genomics in evolution, and phylogenetics has broadened

the scope of research into the evolution of organisms' genomes. In fact, since all organisms have similar genetic codes, they may compare DNA sequences in and among species.

These distinctions are difficult to interpret because of the diversity of methodologies available, and the way in which genotypes, i.e. genotyping or sequencing, are determined, as well as their characteristics or locusts. As many papers examining the effects of natural selection on the reported patterns of variety indicate, the degree to which observable variations on DNA sequences adapt was already significant for understanding development.

2.8 Importance of Enhancers in Primate Genomics

Enhancers are often characterized as DNA areas which encourage gene transcription and work in either direction and at different distances from their aim promoter. For a variety of reasons, the identification of enhancer was challenging in the past. Firstly, 98% of the human genome that does not encode proteins are spread to create an immense search space Enhancer (billions of base pairs of DNA). In the second place, while enhancers are well known for modulating genes in the cis, their position varies considerably relative to the gene product (or genes) of enhancers.

In addition, the closest promoter is not always regulated, but genes can potentially further regulate the chromosome by travelling through nearby genes. In certain cases, individual enhancers were discovered to regulate many genes and add to the complexity of their functional annotation. Thirdly, unlike protein code genes with a well-defined sequence code, enhancers are not well-understood if there is something like a generic sequence code. As a consequence, enhancers cannot be detected reliably using DNA sequence-only computing approaches. In the end, the activation of the enhancer can be confined to a single tissue, cell type or biological period, pathogenic, or environmental conditions. Since enhancers are adaptive, they can thrive when, where and how much our genes express themselves. It also

limits the discovery of enhancer and proper research in the genome due to its genomic role [12].

2.8.1 Enhancers in Eukaryotic Genome

The current findings from tissue and cell line examinations for the large number of enhancer in our genome came as a surprise. Hundreds of thousands of enhancers and well over the 20,000 protein encoding genes are estimated to be present in the human genome. The need to direct gene expression as an initial step to influence the genome and ultimately the function of the organism is emphasised.

This finding Epigenomic techniques show how poorly retained enhancers are in a certain tissue. This includes improvements in a variety of vertebrates and also certain enhancers in mouse and cardiac tissues discovered in human beings. Most of these investigations are designed to refocus their attention away from animal models and human cell lines and towards real human tissues, with a view to uncovering enhancers such as ENCODE. Taking into account the importance of genomic enhancers, the current group of research approaches include a number of inconvenient features. There is currently no unique "enhancer brand" to be used to identify all enhancer genomic locations and to properly forecast whether or not enhancers are present or missing in a certain cell or tissue. All relatively high and super-low enhancer recognition technologies are required to bridge this disparity, including the capacity to segregate or interact with decreased amount of tissue data (even single cells).

2.8.2 Chromatin Template-Closed Compaction

The formation of a loop between different sequences, chromatin modification and chromatin remodeling factors will boost the potential for mistakes of these factors resulting to extra protein-protein and protein-DNA compounds. Similarly, an increase in the concentration of localised proteins has recently been a fundamental method for reducing the inhibition of lake repressor. What happens if DNA loops

that cannot be inserted in the nucleoplasm are presented in enhancer and genetic destinations? In a number of these circumstances, separating chromatin appeared packed and was therefore quite close to enhancer and promoter with one another (200–400 nm). The enormous quantities of transcription factors and membrane proteins nucleated by activating the enhancer can nonetheless get through all this constrained nuclear environment to discover and initiate transcription from the ultimate promoter. Enhancers are genomic structures which should lead to an increasing number of gene transcriptions or genes. These components work in a way that brings the enhancer and receiver gene closer together by forming chromatin loops. DNA-compatible transcription factors interacting with promoters and enhancers were thought to communicate or acquire "looping" molecules that influence long-term links in the chromosomal configuration capture (3C); CTCF and cohesin, insulator-binding proteins, may also increase enhancer–promoter interactions, according to new research.

2.8.3 Enhancer RNA - eRNA

Indeed, enhancers can influence the transcription of their target genes by modifying their transcripts. For many years all was found to generate sensitive and contradictory transcripts by specific enhancer, but the importance of the transcripts was not known. Is RNA or transcription of itself relevant, or is it merely the result of the transcription of a repeating gene? Thus enhancers are often translated into non-coding RNAs of varied lengths, levels of polyadenation and mast characteristics according to genome-wide research. Really used to find functional enhancer RNAs (eRNAs), which implies that enhancer transfer has a role in enhancer functionality [7].

The transcription of eRNAs corresponded to mRNA production in the adjacent genes which showed that they play a transcript regulatory function. When a subset of eRNAs are removed, it is probable that transcription is only a side effect of the activation of the target gene. This shows that the enhancer activity is entirely dependent on the RNA element, rather than on the gene transcription of

the non-coding RNA (ncRNA). There is a tantalising potential that eRNAs play a key role in the establishment or stabilisation of enhancer-promoter networks. In view of evolutionary constraints, portions which are nonprotein code represent approximately 85% of human DNA, with cis-regulatory regions providing for a major fraction of this. It is also no surprise that genetic differences can develop complicated features in some regulatory regions and are involved in causing the sickness of humans. Roughly three decades later, the observation that the translocations of the - globin genome create thalassemias was one of the examples of altered gene regulation as little more than a different disease process. The disease comes to the fore in the absence of globin, which disrupts the direct connection between certain globin genes and their cis-regulatory downstream sequence.

2.8.4 Impact of Chromosomal Translocation – Genetic Variations

Genetic diversity has been associated to some human mendelian illnesses in remote enhancers. Changes in a Sonic Hedgehog Activity (SHH) Enrichor from a megabase radius were made in the initial illustration of pre-axial polydactyly generated by people. Patients that are chromosomally altered and delete the enhancer from neighbouring SHH are present in this phenotype. Enhancer anomalies could only have phenotypic impacts, irrespective of their link with the same gene, which are independent of protein-coding modifications.

Enhancer variations primarily affect transcription in cis, while polymorphisms may change a larger array of gene regulatory variables, such as RNA processing and integrity, protein folding and the like in protein coding sequences. The correct determination of variations on disease-causing regulators at GWAS sites remains a key difficulty, in particular with regard to the proposed technique for the apparent functional implications of these changes. Most regulation variations in GWAS enhancers are already scientifically identified and valuable information is provided [6]. At first the risk of a range of ailments can be influenced by a solitary modification. Second, new disease processes have been discovered, identified or established

as a contributing factor for coronary heart conditions with an affected sensitivity to inflammatory signalling. Third, it usually requires the use of relevant cell lines and/or animal models to determine the major impact of regulatory modifications.

In addition, genetic disease accessory or signalling structures would lead to changeover from scientific proof. The evolution of genomes is driven by mutagenesis. It may arise if an adaptive genetic variable enhances fitness from a single minimum context in which locus is employed, but has low impact on performance in other contexts. In other cells and tissues the bulk of human genes are already transcribed. When using the transcript a comparative genetic mutation could cause complications for the organism in any setting that raises the likelihood of a catastrophic result. The regulation zone of a gene is often the average multiple enhancer stimuli, however (and some other cis-regulatory elements) [14].

Chapter 3

Materials And Methods

This whole workflow involves the prediction of how Human specific Facial-Enhancers causes Genetic variation in Facial Morphology (on comparison with other Non-Human Primates → Orthologs) because of its Transcriptional activities which can be clearly analyzed by structural analysis in between Facial Enhancers (DNA) and Transcription factors (Protein) on respective sites.

3.1 Extraction of Genes Involved in Human Facial Morphology

There are plenty of genes already reported that are solely responsible for Human Facial Morphology. Maintaining the repository for facial genes is crucial, for this prime purpose NCBI (National Centre for Biotechnology Information) available at https://www.ncbi.nlm.nih.gov/ has been consulted. Approximately around 27 genes has been listed down for this study. NCBI is a component of the National Institutes of Health's National Library of Medicine (NLM) (NIH). In order to characterise and organise gene data, the National Center for Biotechnology Information (NCBI) employs genes. As a main node, the map, appearance, sequence, function of protein, structure and the data on homology are included. Each gene record is issued a specific GeneID, which may be watched over the whole review

period. Genes are here determined for known or forecast genes and are delineated by map coordinates or genetic markers.

Gene's successor, LocusLink, has been enhanced multiple times: improved integration with other NCBI databases, a broader environment, and the increased query and retrieval possibilities for the Enterz database system.

3.2 Extraction of Human Facial Enhancers

Enhancers are cis-regulatory regions that participate in the transcriptional pattern; VISTA-Enhancer Browser available at https://enhancer.lbl.gov/ was used to extract related enhancers.

This database classifies putative enhancer elements in the human genome using comparative genome analysis, followed by in-vivo enhancer activity testing in transgenic mice. It has over 250 DNA fragments that have been examined in the lab, with over 100 of them being confirmed as tissue-specific enhancers. It is not necessary to register to use this web-database since it is free to use.

3.3 Filtration of Human Enhancers (NOSE and EYES)

Because the expression patterns of all positive enhancers are annotated, it is possible to obtain sets of enhancers that drive expression in specific anatomical locations or tissues (this database contains over 145000 conserved non-coding sequences). On the basis of their appearance in the nose and eyes, approximately 103 enhancers have been identified for this research. Each of which contains sequence-related information, experimentally validated positive elements, image data, and protein sequence information. On the basis of their appearance in the nose and eyes, approximately 103 enhancers have been identified for this research. Each of which contains sequence-related information.

3.4 Formation of Ortholog Sequences – Non Human Primates

The aim of this phase is to test RBH (reciprocal best hit), which is widely used in comparative genomics for orthology, to determine which genome has the best hit scoring match. For this prime step BLAT has been done through Ensemble Genome Browser available at http://asia.ensembl.org/index.html.

This browser aids research in comparative genomics, evolution, sequence variation, and transcriptional control, and its BLAT feature is designed to find very similar nucleotide sequences between genomes that are similar (same genes-different species). As a query, BLAT accepts DNA or protein sequences and compares only genome sequences. A DNA query can be compared to a DNA database using BLAT. BLAT, on the other hand, will operate in a translated mode.

3.5 Identification of Transcription Factors and Binding sites (NOSE and EYES)

Every type of cell or tissues has a unique pattern of functional Transcription Factors (TF) that self - assemble Transcription Factor Binding sites (TFB's) in regulatory sections of genes, demonstrating their expression at a specific developmental stage or in response to an external stimulus.

For the scrutiny of putative TF's and its TFB's TRANSFAC interconnected database geneXplain available at http://gene-regulation.com/ has been consulted. Its weight matrix based technique enabling the visual representation of DNA factor on strand, TF's and TFB's, as well as the associated protein sequences from your DNA-input, lets us easy to discover the potential location by using the library of PWMs. The TRANSFAC Professional Incentives are required for this database to be registered (MATCH tool). There are various registered match tools available online that can be used for this purpose.

3.6 Manual Inspection of Human Specific Transcription Factors (Muscle Specific)

Out of above extracted 103 Enhancers whose expression can be seen, here is the time where the difference has been laid down in a way that only those Transcription Factors narrowed down aside which are having expression (Nose and Eyes) in only Human but not in other Non-Human Primate group. This requires manual inspection over a huge data set of Transcription Factors by keeping in mind that each Enhancer has more than 3 Transcription Factors along with its differential binding sites. 45 Enhancers has been listed down aside entitled as Human Specific DataSet of Transcription Factors.

3.7 Manual Inspection of Human Specific Transcription Factor Binding Sites

Keeping in mind the above 45 Enhancers just selected upon which Human Specific Transcription Factors has been putted aside too; now it's time to focus on Transcription Factor Binding sites because of which the genetic variations can be analyzed on later stages. 5 Enhancers and it's corresponding 5 Transcription factors and its binding sites has been selected after manual inspection due to the very fact that the major focus is to find where the difference has been actually laid down in between Human and Non-Human Primate group. Gained and Loss idea has been selected in order to narrow down the best selection between the groups.

3.8 Elucidation of Genetic Variations

Color coded technique has been used to find and filter out the Transcription Factor Binding sites inside the nucleotide sequence of Enhancers and named accordingly as it's Transcription Factor. This step is majorly required so as to jump straight

into one narrow area rather than to explore the whole genetic sequence of Human Enhancer, this allows a major ease so as to pick the binding site and compare accordingly the genetic variations in other Non-Human Primate group for analyzing the Loss and Gain variable components.

3.9 DNA-Protein Docking for Human Specific Structural Analysis

With the possible data of Human Specific Transcription Factors and their binding sites (major expression in the nose and eyes), it's now time to recognize the contact surface on DNA that corresponds to its protein, so that a visual representation can be used to understand more about the identification and its genetic expression. The conformational complexes provide information about the precise position in the nucleotide sequence where the protein binds and has its specific expression in humans, as opposed to non-human primates, where the expression has been lost. HADDOCK (High Ambiguity induced DOCKing) is available at https://wenmr.science.uu.nl/haddock2.4/ for this important phase. has been made use of this is by far the best forum for DNA-protein docking, because it guides the docking process' improve settings using non-structural experimental data (Van der Waals energy, Electrostatic energy, Desolvation energy, Violation energy, buried surface area in terms of spatial disposition). This platform includes registration and manages the workspace on a professional level. It takes.pdb structures (protein and nucleic acid) as input and outputs the best clusters. Protein structures has been made by RCSB PDB available at https://www.rcsb.org/ on contrary DNA structures has been made by SCFBio which is freely available and doesn't require registration at http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp.

3.10 Overview of Methodology

Overview of methodology opted for this study is shown in figure 3.1

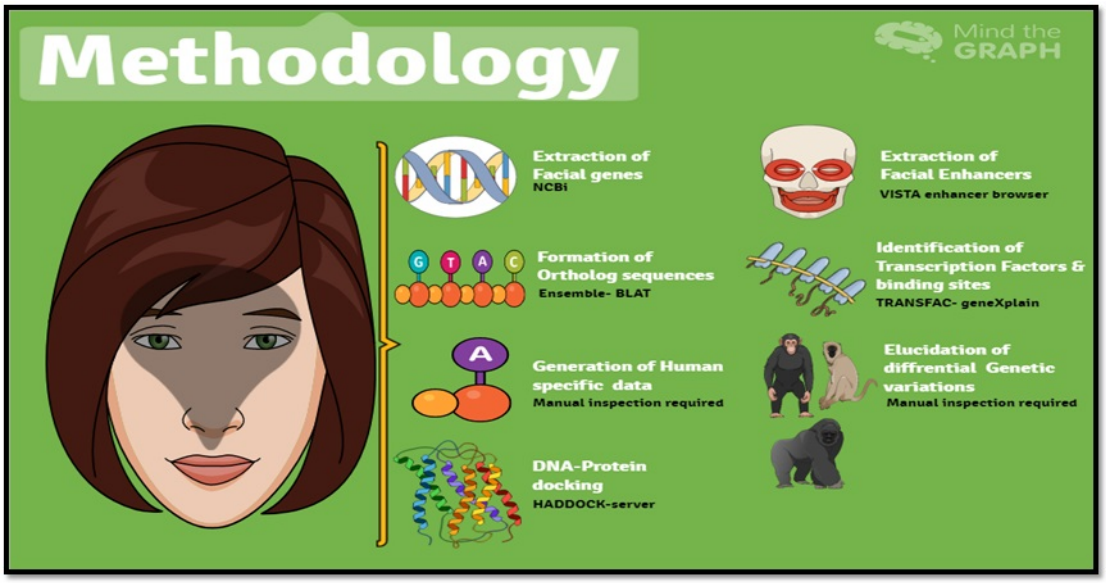

FIGURE 3.1: Methodology opted for this study.

Chapter 4

Results and Discussion

Initially the facial feature structure was regulated by complex embryonic gene expression and even by some molecular pathways. The way a person's face emerges has an impact on hormones, diet, and biochemical factors. The facial frame has a modular structure, with suites of facial features on various sizes that show internal integration while defining features as a natural result of these forces and restrictions during facial morphogenesis and development. As a consequence, several genes are expected to influence the human face while others will only have limited impacts on specific parts of the face.

4.1 Extraction of Genes Involved in Human Facial Morphology

Many data shows that genes play an influence in the face's look This is particularly obvious when we think of our own family, because we share face traits with close relatives less than with other relatives. Although there is limited documentation of how GD affects the types of face distinguishing features that give us our various identities, such as nasal morphology or the distance between our eyes. There were a total of 37 main genes which provide a specific cranio-facial phenotype for the genesis of face morphology. These 37 genes are similar in between Human and

Non-Human primate Chimpanzee (Pan troglodytes), Gorilla and Rhesus monkey (Macaca mulatta).

It has been investigated and contrasted the muscle organisation and appearance of prosimians, cercopithes and humans. In the investigation, a slim pupil collis muscle, one two-layered zygomatic major, and a separate risorius muscle, previously identified exclusively in adult prosimians, were found. The presence in such a defined form of these muscles supports prior research showing that the species has a complicated A face system with a large degree of difference from other non-human sub-species. The structural differences between the two groups are furthermore small (Humans and Non-Human primate) [8].

The exploratory analysis further enhances the functions of all extracted genes as shown in Table 4.1 shown below.

4.1.1 DCHS2 Gene

This gene encodes a big protein that is considered to play a function in cell adhesion with a number of cadherin domains. Van Maldergem's syndrome (Telecanthus, epicanthus, wide flattening of the nose, large inverted W-shaped mouth and malformed ears, malformed limbs (camptodactyly, clinodactyly, interdigital and joint laxity, mental retardation), adult and Pineoblastomas (PBs) are both diseases rarely of the kind of cancer) Two diseases are associated with the Van Maldergem syndrome.

Table 4.1: Functional description of extracted facial genes (refer to Appendix A Table 5.2)

In children aged 1 to 12 years, they are most frequently diagnosed, while adults are quite regularly identified. As a consequence, there is scant evidence of adults in literature and often comes from paediatric practise). Hippo pathway – one of its linked tracks is many species. A protein called Protocadherin-23 (PCDH23) or Cadherin-27 has been encoded in a DCHS2 gene (CDH27). The angle of the nose is connected with DCHS2 (how much a nose is upturned).

4.1.2 PDE8A Gene

This codes a member, especially of the PDE8 subfamily, of the cyclic nucléotide phosphodiesterase (PDE) family. The hydrolysis of cAMP, a second messenger to regulate and mediate different cellular responses to extracellular signals, is catalysed by this PDE. This protein is engaged in a range of key physiological processes because of its capacity to control the cellular cAMP content. This gene is expressed largely on the face and helps conserve the basic morphology of the face. In the PDE8A area linked to the allometrical component of PSD, certain additional face traits were also minorly associated.

4.1.3 SCHIP Gene

The SCHIP is a metric of the whole face size, which is unrelated to characteristics such as facial height and nasal width. It is found in a range of tissues, including the developing face, where it has decreased expression in mesenchymal nasal processes, maxillary processes and mandibular processes. In the SCHIP knockout gene, neural crest derived skeleton, palatal processes and snout improperly grow, suggesting that they are important for the facial development as well. It operates as an early reaction gene in the PDGF signalling cascade and stimulates cell motility in response to cytoskeleton reorganizations [3].

4.1.4 ASPM Gene

This gene is the human equivalent of the Drosophila melanogaster 'abnormal spindle' gene (asp), which is needed in embryonic neuroblasts for a normal mitotic spindle function. According to animal research, this gene can play a function in control of mitotic spindles and influence neurogenesis. The predominant kind of microcephaly has been associated with these gene alterations.

There have been multiple transcript variations in this gene which coded different isoforms. Protein codes are used to identify the gene of ASPM (Assembly Factor

for Spindle Microtubules). Some of the disorders connected to ASPM include microcephalus, primary, autosomal recessionary, and primary autosomal recessive microcephaly.

4.1.5 DLX6 Gene

This gene is connected to the distal gene of Drosophila and is part of the family of the homeobox transcription factor genes. This family comprises six members, each with a protein that contributes to the growth of the forebrain and facial feature. A protein codes for the gene (Distal-Less Homeobox 6). Two related disorders are the isolated Split Hand-Split Foot and Rapp Hodgkin Syndrome. MECP2 and Associated Rett Syndrome (Ectodermal Dysplasia is an abnormal development category of two or more ectodermal structures. Ecodermal Dysplasia All of the symptoms are anhydrotic ectodermal dysplasia, splinter lip and palate. Two related pathways are the distinctive appearance of the face (narrow nose and mouth), wire, slow-growing and incombable hair, thin eyelashes, obstructed lacrimal puncta/epiphora, bilateral stenosis of external auditory canals, microsomalia, hypodonia, cone shapes, hypoplasia of the size of the enamel, dystrophic nails and cleft lip / cleft palate. Two gene ontology (GO) annotations are DNA-binding transcription factor activity and DNA-specific sequence-binding. This gene's DLX1 is a major paralog.

4.1.6 EDAR Gene

EDA (including EDA, EDAR and EDARADD) defines the position, scale and shape of the ectodermal womb annexe (including the EDAR banding death-domain adapter protein) (such as hair follicles, teeth and glands). The effect of epithelial mesenchymal interactions throughout the development of the facial feature mice is hypothesised to have on the mandibular shape. The revelation that EDAR influences chin protrusion adds to the increasing number of evidence of phenotypic effects in this gene. The most frequent investigation in this development was the

ectoderm, which produces skin, hair, nails, teeth, mother's and sweat glands.

4.1.7 HOXD1 Gene

Hoxd1 can be present in the crest cells, face and forebrain primordia. The loci related to the nose were linked to the various features of the nose, including nose protruding, nasal height and nasal shape, suggesting that nose underpins precise genetic regulations and the associated facial traits are shown in fore, side and eye figures and on the right facial angle [15].

4.1.8 WDR27 Gene

The WDR27 gene (also known as IFT121) encodes a protein that helps cilia generate and maintain minuscule finger-like projected from the surface of a cell. Cilia, including in the kidney and liver and luminescent tissue in the back of the eye, is important for the production and function of a range of tissue and cells (the retina). Cilia is also involved in the production of bones, but it's not recognised as the process.

4.1.9 PAX3 Gene

PAX3 is a member of the PAX gene family involved during embryonic development in the formation of tissues and organs. The PAX gene family is also vital to support appropriate post-birth cell function. During embryonic development, the PAX3 gene is active in neural crest cells. These cells move from the spinal cord of the embryo to certain locations. The PAX3 gene generates a protein which governs the actions of other genes, such as nerve tissues and melanocyte generating pigmentary cells, which are called neural crest cells. Most PAX members have a paired box domain and a paired homeodomain. These genes have a crucial role to perform in foetal development. The disease of Waardenburg, facial feature deafness and alveolar rhabdomyosarcoma are all associated with mutations in the Box 3 pair.

The translocation t(2;13)(q35;q14), which mirrors a PAX3 fusion and the Forkhead gene, is often present in alveolar Rhabdomyosarcoma [3].

4.1.10 TP63 Gene

For early development, the tp63 protein is necessary. It is particularly important in the appropriate development of ectodermal tissues such as skin, hair, teeth and nails. It could also contribute to the development, according to the research, of the limbs, facial features, urinary and other organs and tissues. In addition to its embryonic roles, the tp63 proteins appear to be necessary for the maintenance of many cells and tissues later in life. These isoforms are all impacted by skin growth and retention, adult stem/pronatal cell regulation, heart expansion, and precocious ageing. These isoforms are all impacted by skin growth and retention, adult stem/pronatal cell regulation, heart expansion, and precocious ageing. Some isoforms were shown to protect the germinate by eliminating oocytes or testes that had been damaged by DNA.

Ectodermal and cleft lip-palate syndrome 3 (EEC3); SHFM4 splits-hand-foot malformation; ADULT syndrome (acro-dermato-ungual-lacrimal-tooth); limb-mammary syndrome; Rap-hodgkin syndrome; Orofacial cleft 8 all related to mutations with it. This can cause a disruption to the skin, which causes a disorder, which can be caused by the disease

4.1.11 PABP1 Gene

This nucleolary protein possesses a significant nascent poly(A) affinity. For the gradual and successful polymerisation of poly(A) tails at the three ends of eukaritic transcripts, the protein is necessary and the poly(A) tail size is limited to approx. 250 nt. This protein is located in a stable nucleus while the cytoplasm contains another poly(A) binding protein. The 5” end of the coding area for this gene has a GCG repeat and the extension from the conventional 6 copies to 8-13 copies causes autosomal dominant muscle degeneration of the oculopharyngeal (OPMD).

(Hereditary condition, usually above age 40 that causes muscle weakness in people; "oculopharyngeal" are the eyes and a part of the throat known as pharynx (-pharyngeal). The first thing that weakens the muscles in both eyes is that the affected people have droopy lids (ptosis). Ptosis can deteriorate with time, hinder the view and impede eye motion under specific circumstances. In addition, ptosis makes it much harder to swallow the throat of affected people (dysphagia).

4.1.12 TMEM163 Gene

A highly-conserved gene in mammals is TMEM163 which codes for a transmembrane protein produced in brain and retina as well as for several other tissues. TMEM163 is a binding partner of MCOLN1, the lysosomal storage disorder of Mucolipidosis type IV (MLIV) caused by loss of function mutations (including deletions and point mutations) in a new investigation. Although MLIV does not often exhibit distinct facial characteristics, few examples of unique facial dysmorphia, especially around the eyelids, have been described. The pathophysiology of TMEM 163 may play a role through affecting levels of cellular zinc that increase in MLIV and when TMEM 163 is eliminated. MCOLN1 orthologists are expressed throughout the development of zebra-fish eyes, showing that the relationship TMEM 163–MCOLN1 can have a role in determining the morphology of the eyes. These results demonstrate that TMEM163 may be a potential functional candidate for the PC1 eye phenotype.

4.1.13 COL17A1 Gene

Collagen XVII is a kind of collagen in the skin. The gene COL17A1 encodes the collagen type XVII, a transmembrane type II protein, having an end of the C-terminal cytoplasm N ending and a terminal end of C which traverses the lucid lamina on the outside cell side. This protein is a trimmer with three-chain 1 connected via contact with the protein plectin and BP230 to the basement membrane. A variety of symptoms result from pathogenic mutations in COL17A1.

Although different researchers have examined likely correlations between genotype and phenotype, the only evident link is that levels of expression of collagen XVII have a direct influence on clinical expression. This issue is due to mutations in the COL17A1 Gene, coded to a structural component of hemidesmosomes (NO-Herlits-type epidermolysis bullosa), a skin blistering disorder accompanied by dental enamel defects. It is a transmembrane collagen that runs from a hemidesmosomic component of the intracellular basement membrane zone of the lower lamina lucida of the basement keratinocytes. It also thought it was a vital protein for preserving the integrity of the corneal epithel. Current corneal erosions are associated with mutations in the gene and their expression is aberrant, as are generalised atrophic benign and junctional epidermolytic bullossa in many malignancies.

4.1.14 PARK2 Gene

PARK2 was first found to be a cause of auto-recessive Parkinson-like juvenile disease and is since associated with autism spectral disorders and hyperactivity disorders with attention-deficit. A cytoskeleton-relevant protéin, which is expressed in midbrain, basal, cerebral and cerebellar neuronal processes and cellular bodies, Parkin is encrypted with PARK2, which can be located on 6q26. Parkin is a transcriptional protein ligase p53 that causes developmental delays, abnormal eye motion and peripheral hypertonia in the axial ligase.

4.1.15 OSR Gene

All predicts DNA-bounding factor activity in transcription, RNA polymersis II-specific DNA binding activity and sequence-Specific DNA-bounding activities of RNA polymerase II transcription regulatory region. Cell difference in angioblast, glomerular visceral differentiation in epithelial cells and pronephric development of nephron are all processes involving this protein. The focus in the nucleus is predicted. Among the structures in which it has been expressed are endoderm, germ ring, head, Mesoderm and vascular. OSR1 is human-standard for OSR1

(odd-skipped related transcription factor 1). It contributes to the development and suppression of the tooth palate after adult teeth have erupted. During development of the mouse, Osr2mRN A expression designates the medial maxilla areas in which the palate grows early and persists strongly in the palatal regiments that grow.

4.1.16 FOXA Gene

Establishing the expression of tissue-based genes and regulating gene expressions in differentiated tissues; designed to operate as a "pioneer" factor by engaging with nucleosomal core histones and replacing the link histones on locations of destination enhancers and/or promoters, opening chromatin to other proteins. Foxa1,2,3 is a member of the Foxa winged helix/forkhead box (Fox) subfamily that plays critically important function in the early development and organogenesis of mammals. In post-born life, the Foxa factors are vital since they regulate many genes in the liver, pancreas and the adiposal tissue to control metabolism and homeostasis.

4.1.17 MAFB Gene

This gene creates a transcription factor of a basic leucine zip (bZIP) that regulates hematopoiesis specific to the lineage. The encoded nuclear protein suppresses the transcription of erythroid-specific genes in myeloid cells through ETS1. In this gene there are no introns. It acts as a linespecific hamatopoesis-regulating transcription/repressor in myeloid cells by eliminating ETS1-mediated erythroid-specific gene transcription. This protein is necessary for the formation of monocyte, macrophage, osteoclast, podocyte and islet beta cells. The promoters of insulin and glucagon are activated [16]. DRS is an affinity of the movement of the eye, characterised by a limited view and retraction of the eye. The heterozygous MAPB loss of function and the dominant negative MAPB mutation are responsible for RSD and deafness. In postmortem DRS tests, Abducens nerve hypoplasia and an aberrant lateral rectal nerve internvation are found. Aberry oculomotor

nerve branches that appear at developmental decision places next to the target extraocular muscles, when the abdicant nerve is selectively damaged, are the cause of secondary innervation of the lateral rectum muscle [9].

4.1.18 TRPC6 Gene

The gene produced protein generates a calcium channel in the cell membrane triggered by receptors. A second messenger system of phosphatidylinositol is intended to modulate the channel caused by diacylglycerine. The channel is not triggered by low amounts of cellular energy or by the protein kinase C. Focal segmental glomerulosclerosis 2 induces mutations in this gene (FSGS2). TRPC6 is a member of the Subfamily C Protein Family of Protein Transient Receptor Potential Channel. Genetic Steroid-Resistant Nephrotic Syndrome and Focal Segmental Glomerulosclerosis 2, are two TRPC6 related disorders. TRPCs, TRPVs, TRPMs, canonical (melastatin), The 6 key subfamilies of the mammalian TRP (transient receptor potential) are TRPP (polycystin), TRPML (mucolipin) and TRPA (ancyrine). These subfamilies consist of 28 ion channels, serving as different mobile sensors. All channels can go through monovalent cations, and most can go through Ca2+ as well. Throughout odontogenesis, TRPC canals are expressed at varied levels according to immunostaining. The genes found are unique opportunities that are thought to be vital for the production of rat tooth germs. The results contribute to an increased awareness of the gene regulating mechanisms.

4.1.19 TBX15 Gene

T-box transcription factors are evolutionary and play a major role in biochemical modifications, including the formation and patterning of mesoderms, and in morphologies, in both vertebrates and invertebrates. The relationship between the mutations leads to congenital morphogenesis, such as cardiac, cranio-facial and limb malformations, shows how important T-box genes are to human development. Individuals with pelvic and scapular hypoplasia (Cousin syndrome),

cranial, cervical, atrial and skeletal syndrome that seeks to simulated the abnormal face phenotype are associated with cardiac, facial feature and T-box mutations. and limb malformations. Tbx15 A group of 18 genes that are evolutionarily consistent (T-domain) and engaged in a variety of developmental events, including mesoderm specification with a similar DNA-binding domain. Tbx15 has an agouti protein distribution-covering dorsoventral pattern. Inactivation of the gene Tbx15 in rodents and human variants of Tbx15 induce serious skeletal deformities. When interacting with Groucho family co-repressors, Tbx 15 inhibits transcript of a lot of genes, which means that Tbx15 is mostly a repressor [10].

4.1.20 RPS12 Gene

The ribosome, the organells that catalyse protein synthesis, consist of a minute 40S component and a large 60S subunit. These sub-units consist of four types of RNAs and around 80 proteins which are structurally distinct. This gene, which covers ribosomal protein, is in a 40S subunit of the ribosome. The S12E ribosomal protein family contains the protein. The cytoplasm can be discovered. This gene was found in colorectal tumours to be higher than that in normal colonic mucosa. The genome contains many processed pseudogenes, as is common for genes that produce ribosomal proteins.. Dental construction in the embryonic skull, a complicated and sensitive process that is guided by a unique genetic programming process. The development of the face is a complicated number and must be coordinated of embryonic processes.

In the fourth week of pregnancy the human face begins to take shape and has a close link to cranial crest neural cells. Genetic mutations and environmental factors can interfere with the delivery of these programmes, leading to facial or oral abnormalities. Two prevalent dental problems include orofacial clefts, and hypodontia and oligodontia. Orofaciales and dental structural malformations arise during embryogenesis from interference in normal development of the head.. For the initial supplement of 46 chromosomes to be rebuilt in the new somatic cell, germ cells have to be half as many chromosomes (the haploid number) as somatic

cells. Meiosis is the method by which germ cells produce the somatic cell by half a chromosome. The process of mitosis divides somatic cells. About 10 percent of all human disorders are responsible for a single gene mutation. These modifications are transmitted in many ways, two of which are very essential [3].

4.1.21 FREM1 Gene

The gene encodes the main gene that supports the trigeminal and renal systems. Mutations in the gene were associated with anorectal and renal abnormalities with or without a bifid nose. Alternatively, spliced variants of transcript encoding different isoforms have been found.

The shortest isoform (Toll-like/interleukin-1 regulator, TILRR) is thought to be the co-receiver of the family of interleukin 1 recipients, which regulates the receptor function and cannabinoid transcription factors like Toll-like receptor/interleukin 1, and thus is capable of helping to control inflammation.

Although genes play a major role during cranial bone fusion, it has been difficult to pinpoint specific genes, even when the illnesses of craniosynostosis are predominantly inherited.

In all cases of craniosynostosis, early fusion of the forehead bones, termed as metopic craniolynosis (MC), accounts for between 5 and 15%. Premature frontal fusion creates trigonocephaly, a distinctly altered morphology of the skull that usually requires surgical correction.

Surprisingly, there is no recognised cause for most MC cases (idiopathic). A big number of unisutural metopic craniosynostosis individuals were present with the FREM 1 gene. Micro CT image and quantitative analysis of the skull shape indicate Frem1's role in metopic suture control revealed premature fusion of the PF suture, and abnormalities in the frontal bone form. In combination with the results of Frem1 gene expression and protein expression.

These findings show that FREM1 mutations cause metopic craniosynostosis [11].

4.1.22 HOXD Gene

The more later Hox genes in the hindbrain and branch arches have a significant influence primarily on mesodermic-derived cranial bones, but those that were more prevalent can influence more anterior branch and crest-derived structures in the hindbrain and branch arches. The Hox branch code, represented by combination structures of gene expression Hox, has been designed as a strategy of development in which an axial location in the neural pipe is transferred by means of the migrant neural crest, into the periphery (the arches).

The embryonic head identifies 4 paralog Hox sets (HOX A,B,C and D) as "caudal rhombomerics." When neural crest cells migrate from the hind brain into the branch ark on identical axial levels, face mesenchyma is generated. Genetic expression in diverse tissue-tissue interactions can be seen throughout the development of embryonic vertebrate. For optimal facial feature, limb and ectodermal organ development gene-mediated inductive interactions are necessary. The most frequent congenital birth disorders in humans are craneiofacial anomalies such facial clefting and craniosynostosis and result from mutations in genes which disrupt any of these processes [15].

4.1.23 PKDCC Gene

A potential proteins kinase identified in the mesenchyma of the arches as well as in the limb buds is encoded by Pkdcc. The cartilages of the previous neurocranium were derived from the viscerocranium and neural crest; both were phased out as a result of genes and dose-dependent.. Pkdcc encodes proteins that are respectively identical to 56% and 40% amino acid.

The morphants displayed a ventral head collapse, eye abnormalities, increasing yolk and a developmental delay, as well as loss of ceratobranchi and ceratohyal reversals. Long bones are shortened due to delayed endochondral expression of Pkdcc protein ossification, and facial feature defects include a tiny and shortened nasal capsule and maxilla [13].

4.1.24 C5orf50 Gene

C5orf50 is expected to encode an uncharacterized transmembrane protein, which can be discovered inside a 1.24Mb duplicated area in a patient with preaxial polydacty and holoprosencephaly (HPE). The frontonasal suture, which connects the frontal bone nasal area with the nasal bones, is its most anterior point. The nasal bones are interspersed with the frontonasal suture and the inner suture.

It looks as a conspicuous sunken region between the eyes, exactly above the nasal bridge. It is an immediate cerebrospinal marker below the glabella. The most feasible option in this doubled area is FBXW11, a gene connected to the sonic signalling of hedgehogs, the main disturbed route in HPE. C5orf50 mutations probably affect the segmentation of the cranium-facial structure through changing production of FBXW11, but the loading factor for C5orf50 is also likely to effect the face more directly and clearly [4].

4.1.25 EPHB3 Gene

Eph and ephrins are cell surface receptors, which bind to each other, respectively initiating forward and reverse signalling pathways. The removal alone and in combination with the loss of EphB2 and EphB3 of the protein ephrinB1 induces palate splitting. Due to the bidirectional signals of such molecules, the cleft palate in the ephrinB1 protein null, EphB2 and EphB3 molecule null, was confused whether forward or backward signalling was induced. The Neural Crest (NC) is an early developmental group of mammalian cells that pass through many things and specialise in many different things. EPHRIN-B1 is possible to bind and signal tyrosine kinases to the receptor EPHB1, EPHB2 and EPHB3, although the signalling partners of CFNS continue to be unknown. According to the geometric moral analysis of EFB1 allel series, Ephb2; Ephb3 mutant embryos, but total loss from EPHB1-3 does not fully recapitulate the severity of heterozygositis, CFNB1-like EPHB1; influes cell segregation in facial feature development (CFNS). EPHB2 and the CFNB3 are key receptors mediated by HFE 1 hemisygous-like phenotypes [14].

4.1.26 ZNF219 Gene

The Zinc finger protein 219 (ZNF219) partner SOX9 transcription factor is essential for regulating the development of mammalian chondrocytes. The collagen type 2 gene alpha 1a (col2a1a) gene is necessary in the zebrafish notochord to control the expression. It consists of nine C2H2 type zinc finger fields, which in the luciferase test discover reasons and regulate promoter activity. In addition, the znf219L morphological knockdown lowered the expression of the endogenous promoter in the notochord. ZNF219L can alter notochord-related genes expression throughout the early embryonic phases.

The 723 amino-acid protein is produced with nine different zinc finger domains of Cys2His2 (C2H2), eight of which are separated-paired. In general, there is three varieties of zinc finger-type C2H2 proteins, three-fold C2H2, one-fold C2H2, one-paired C2H2, and one-paired C2H2 protein. They offer a wide range of functions because to their zinc finger design [16].

4.1.27 SOX9 Gene

A number of developmental processes were associated with the transcriptional control systems of the Sox family. SOX9, a member of the family, was initially discovered as a possible gene for the condition of the human skeletal and testicular development of campomelic dysplasia (CD). The majority of the endochondral bones of the face do not develop micrognathy, palate dizziness and face dysfunction in these patients.

SOX9 is discovered as a possible CD gene, a skeletal disturbance, cranial dysmorphology, and sexual revolution. In cartilage development, Sox9 appears to be a critical role. In all skeletal components developed,, it plays a critical role in the specification and differentiation of mesenchymal cells toward the chondrogenic lineage. Sox9 works together and governs the activation of chondrocyte-specific enhancers of several genes crucial for the development of chondrocytes, including collagens type-II, Col2a1, the principal protein mature cartilage.

Sox9b is expressed previously as a neural crest progenitor than sox9a and seems to have a part in the development of neural crests, as indicated by less expression in Sox9b mutant embryos of the early neural crest markers snail1b, foxd3 and sox10. In addition, overexpression tests demonstrate that the effects of sox9a and sox9b on each other's expression, resulting in ectopic expression of several early neural crest markers, indicating that sox9b has an early role in neural crest specification [17].

4.1.28 CASC17 Gene

The Homo sapiens identifies this 2,077-nucleotide- long lncRNA gene sequence. Acts during the formation of chondrocytes and supports anti Muellerian gene transcription (AMH) in combination with steroidogenic factor (SOX9 Interacts with CASC17 via transcription). The tip of the nose is recessive and wider along the side of the nose. A RNA gene which belongs to the family of long non-coding RNA (lncRNA) is a CASC17 (Cancer Susceptibility 17) [1].

Interstitial nephritis is the disease associated with this mutant gene variant (a kidney illness characterised by swelling between the renal tubules.) Is a CASC17-related illness. The basic tasks of your kidneys are to filter your blood and remove waste. The kidney tubules reabsorb water from the filtrated blood and extract organic molecules that you don't need into the urine for disposal. Swelling can lead to various moderate to severe renal conditions. The tubules are swelling.

4.1.29 KCTD15 Gene

Kctd15 is a powerful transcription factor inhibitor for the enhancer binding protein 2 (AP-2), which is essential in the formation of neural crest at different stages. Although Kctd15 is connected to AP-2, its interaction with similar sites and times is not hindered or in the nucleus is prevented. Kctd15 binds to the activation area of AP-2 and suppresses transcriptional fusion protein activation, which includes the Gal4 DNA binding frame and mechanisms, and the AP-2 activation motif. The

domain of the NC is defined by the inhibitory role of the Kctd15 in the generation of NC which impedes the placodal domain laterally. Whilst Kctd15 suppresses the canonical Wnt pathway, it is a crucial signal in NC induction, the Wnt inhibitory level was insufficient to explain its efficacy in NC formation suppressing, which suggests that KCTD1 can interact in human cells with AP-2. Kctd15 interacts with AP-2 and deletes 3 members of the AP-2 family from transcription activity. Kctd15 binds to the AP-2 coder in the reconstituted model system, which prevents it from functioning. AP-2 inhibition in the NC formation regulation is a significant mechanism in the function of Kctd15 [18].

4.1.30 PRDM16 Gene

TGF-β(in mice and humans it has an influence on orofacial development. TGFs, activins and BMPs all form part of a broad family of ligands that interact with the heterodimeric serine/threonine kinase transmembrane sensors.) is a transcriptional cofactor that is found in a range of tissues, including a cranial neural crest and a brain, and that affects the facial feature development.

Because of the lack of function Prdm16, Prdm16 mutant animals have secondary palate and micrognathia non-syndrome splitting. The deletion of Prdm16 led to a significant decrease in histone 3 lysine 9 in the palatal racks, but had no impact on histone 3 lysine 4 methylation. Prdm16 maintains the temporal and spatial control of gene regulatory networks necessary to produce cNCC, with these roles preserved as well as differing amongst vertebrates [19].

4.1.31 DHX35 Gene

The protein gene codes for DHX35 (DEAH-Box Helicase 35). Putative RNA helicases are proteins that have a preserved motive Asp-Glu-Ala-Asp that contribute to cell development and division/protein import into the mitochondrium. A range of cell activities, including beginning of translation, nuclear and mitochondrial splicing, and assembled ribosome and sliceosomes, to name just a few, alter RNA

secondary structures. Based on its distribution patterns, embryogenesis, spermatogenesis and cellular development, the quantification of several features of nose morphology such as wing nose width, nose tip shape or nose profile is involved.

4.1.32 SUPT3H Gene

A protein encoding gene is SUPT3H (SPT3 Homolog, SAGA And STAGA Complex Component). The variation of SUPT3H is hypothesised to influence the angle, shape and the thickness of the head of the cornea, whereas SUPT3H influences the centroid size of the variation (squared root of the squared distances of all landmarks of the face from the centroid). Defines the depression of the nasal nasal side, which is superiorly positioned and wider. On the subnasal level there is also a slight depression. Cleidoco-cranial dysostose (CCD) is a congenital disorder known as cleidoco-cranial dysplasia that mostly affects the bones and teeths and causes a mutation The collarbone can be pushed together securely because of the generally undeveloped or lacking colours. Front of the skull often does not cover until later in life, and the affected ones are usually short of average. All probable signs are a broad front, wide eyes, peculiar teeth, and a flat neck [22]. The symptoms vary amongst people, although IQ is often unchanged. Either the disease is legacy of its parents or arises as a consequence of a new transformation. It is transmitted into an autosomal dominant pattern over generations. It is due to an RUNX2 gene mutation that governs bone growth which controls bone development. Symptoms and X-rays are used to make a diagnosis, which is then confirmed through genetic testing [2].

4.1.33 ALX3 Gene

The crest neurons are rich in Alx transcription factors for the frontonaal population. According to genetic mutation research the Alx3 influences differentiation timing and cell morphologies across frontonasal neural crest subpopulations. This gene generates a nuclear protein which governs cell type distinction and develop-

ment with a homeobox DNA-binding domain. ALX3 (ALX Homeobox 3) is a protein gene. The ALX3-associated disorders are frontonasal dysplasia 1 and basal encéphalocytes. ALX4 is an important gene paralog. For this null allle the lacZ coding sequence is inseparable from the wild type within the Alx3 mutant. On the other hand, Alx3 & Alx4 have substantial facial feature deformations which are missing with single mutants from Alx4. Double mutant infants in Alx3/Alx4 have divided nasal regions. The bulk of the facial bones are malformed, truncated, or totally absent, as with many other neural crest elements of the skull. In Alx3/ALX4, double mutants become physically visible on the 10.5 embryonic day, when the nasal processes seem to be inappropriately positioned. The failure of the medial nasal processes to fuse in the face midline is most likely the resulted [20].

4.1.34 GL13 Gene

The GLI3 gene belongs to a set of genes which, during early development before birth, affect the form (shape) of various tissues and organisations. GLI genes produce proteins that attach to some portions of DNA and help modulate if genetic factors are activated or disabled (gene expression). GLI proteins are therefore called factors of transcription. Sonic hedgehog is a protein which operates like GLI proteins in the same molecular path.

This approach requires early development. It affects cell development, cell specialisation, and structural patterns such as brain and limbs. The GLI3 protein in response to Sonic Hedgehog signals the GLI3 protein can either activate (activate) or repress (repress) other genes. During the development process, scientists strive to determine which genes are targeting the GLI3 protein. Acrocallosal syndrome is a condition characterised in the genetic disorder by cognitive deficiencies, additional fingers and toes, and odd facial features, such as hypertelorism and a high front. Patients with acrocallosal syndrome have at least two GLI3 mutations. Greig cephalopolysyndactyly syndrome has extremely similar signs and symptoms as well as symptoms. The GLI3 gene mutations that affect the blocks of single protein (amino acidifyers) in a certain section of the GLI3 protein are

causing Acrocallosal syndrome which interrupts its function. The faulty protein most likely alters the expression of specific genes during early development. The relevance of GLI3 in brain and foot structures could explain why variants cause cognitive impairment, polydactyly, and other acrocallosal symptoms.

GLI3 gene is carried by persons with polydactic, hyperteloric, broad forehead and exceptionally large heads Greig cephalopolysindactyl syndroma, a unique condition characterised (macrocephaly). Changes in individual DNA (base pairs) building blocks for crucial gene sections and insertions and/or removals of a tiny amount of DNA, are all linked to Greig cephalopolysyndactyly syndrome. This condition can also be caused by chromosomal abnormalities inside the area of chromosome 7 which codes the GLI3 gene. Grey cephalopolysyndactyly syndrome is caused by genetic malformations which inhibit one copy of the gene for each cell that precludes the production of any functional GLI3 protein. With just about half of the typical amount of the protein, the expression of target genes during early development is affected. How a GLI3 shortcoming affects your health is still unclear. Determines the development of the limb, head, and face leading to specific Greig cerebral syndrome symptoms [21].

4.1.35 RAB7A Gene

Small GTPase alternating between active and inactive GTP countries. It interacts to a number of effector proteins in its active state, which have a major function in regulating endo-lysosomal trade. The Early-to-Late endosomal maturation, minus-end and Plus-end directed endosomal migration and placement microtubule regulates transport of endosome lysosomes by multiple cascade protein-protein interactions. Not just in endosomal traffic but also in a range of other cell and physiological processes including cell signalling through growth factor, nutrient absorption as a nutrient transporter, axon transport by neurotrophins and lipid metabolism. Charcot-Marie- Distal muscle weakening and atrophy as well as high prevalence of foot ulcers, infection and toe amputations from recurring infections are a distinctive feature of the clinically known type 2B dental (CMT2B).

The 3q13-q22 chromosome is associated with CMT2B. Two misense mutations (Leu129Phe and Val162Met) in the tiny GTP-ase late endosomal protein RAB7 cause the CMT2B phenotype in three larger families and three patients with a positive family history. These two mutations are aiming, according to the RAB7 alignment, at well preserved amino acid positions. We found RAB7 in sensory and motor neurons, and we were extensively expressed [22]. Helps to melt the phagosomes and lysosomes. It plays a crucial function in the infection and survival of microbial pathogens and in the virus life cycle. Microbial disease agents have RAB7A-driven survival strategies include use of RAB7A (e.g. Salmonella) function and ignore RAB7A (e.g. coli function) (e.g. Mycobacterium). Works in conjunction with RAC1 on the creation of osteoclasts (ruffled borders). NTRK1/TRKA modulates the trafficking in endosomal products and signals of neuritis. Regulate endocytic transactions inside the EGF-EGFR complex by managing the degradation of its lysosome.

4.1.36 MBTPS1 Gene

That gene code includes the proteasses that process protein and peptide precursors while passing through the secretary route, either regulated or constituent, for a member of the subtillin-like proprotein convertase family. The encoded protein passes a main automatalytic mechanism in the ER which leads to a heterodimer escaping the ER to Golgi, where everything performs additional autocatalytic sorting and catalytic results. The code is used to manage cholesterol or the lipid homeostasis through cleaning up non-basic residues in substrates for membrane-bonded type 1. Dysfunction of the lysosome may be linked to mutations in this gene.

A breakdown in somitogenesis or, more frequently, a lack of production and pattern of bone could produce a spinal deformity in caudal regression syndrome. According to our findings the MBTPS1/SKI-1/S1P protein is implicated in somitogenesis and the lumbar/sacral vertebral anomalies aetiology, which starts trans-membrane transcription factor. A living mouse with a kinky short tail, deformed, joined, and

condensed lumbar and sacral vertebrae resulted in conditionally removing Mbtps1. The lumbar and the holy vertebrae are joined together, reducing the number of the lumbar and the sacred vertebrae. (i) Mbtps1 is required in the presomitititic mesoderm to support Fgf8's 'wavefront,' which is either axial elongation basis, (ii) establish an oscillation' clock' which determines the rate of change in somite formation, (iii) set the components and persona of the somite organic matrix incorporating fibronectin, fibrillin2 and laminin and (iv) restore the contents and idents. We hypothesise that MBTPS1 impaired mutations are the cause of caudal regression syndrome based on this spinal phénotype and recognised roles in the MBTPS 1 [22].

4.1.37 PCDH15 Gene

This same gene is part of the cadherin superfamily. Members of the family encode membrane proteins which facilitate calcium-dependent adherence to cells. It is vital to maintain an adequate retinal and cochlear function. Hearing loss and IF Usher syndrome are due to gene alterations (USH1F) The orthological mouse had a wide range of alternatives, which led to several isoforms. Human beings are believed to have similar spliced transcripts and additional variants are likely. A bilateral condition of Usher type I(USH1) occurring throughout adolescence is important, congenitally conductive, vestibular and retinitis-pigmentsa characteristics.

RP develops slowly, bilaterally and symmetrically degenerating the rod and cone functions of the retina in teens, resulting in gradually decreased vision and visual field. Individuals usually only need a cochlear implant to develop speech. RP develops slowly, bilaterally and symmetrically degenerating the rod and cone functions of the retina in teens, resulting in gradually decreased vision and visual field. PCDH15, a superfamily member of the cadherina, has been discovered that influences neuronal development and function in a risk-gene for Neuropsychiatric diseases. Researchers have recently studied rare variations with considerable effects in order to better understand the aetiology of these complicated illnesses. In

addition, neuropsychiatric diseases with overlapping genetic etiologies include both schizophrenia (SCZ) or autism spectrum disorder (ASD). And it demonstrates the link to that gene [24].

4.2 Extraction of Human Facial Enhancers

Human genetic information and vast cDNA data bases have provided an almost complete overview of human genes and proteins they make, as have efficient genetic forecasting models. In fact, we also have extensively investigated genome, our understanding of the positioning and activity of cis-regulatory components, particularly remote enhancers. The more efficient analytical enhancer accessible now is backed by the conclusion that cis-regulatory elements are usually strongly reduced and hence may be established in biologically distant species by their stability.

Enhancer were located, for example, in a gene-sparse section of the human genome, to systematically search for preserved non-coding sections. With the recent rise of whole-genome sequences of vertebrate and the introduction of strong comparative genomic approaches, a Genome-wide accuracy assessment to facilitate in vivo research has become important.

4.2.1 VISTA Enhancer Browser

The VISTA Enhancer Browser has been developed as a public supply for the enhancer activity to be allowed to exploit sealed sequence sections. The main data set for the VISTA Enhancer Browser consists of experimental in vivo data for the tissue-specific enhancer. These elements differ from their unexpectedly large conservation in mammals, including "ultra-"conservation [82 percent for at least 200 bp between human beings, mouse and rats], by their conservation over long (chicken and frol) or exceptionally long evolutionary distances (pufferfish and zebrafish). The VISTA Enhancer Browser main objective is to provide a standard

method for empirical findings, but also a programmed data store of around 145 000 conserved non-coding sequences. Based on their high conservation level (P 0.001) across humans, mouse and rats these key principles are selected. The VISTA Enhancer Browser records sequence-related information, picture data and tissue specificity information for experimentally tested positive elements. The sequence data comprises locations, nucleotide sequence, often used PCR primers, the chunk, the names of neighbouring genes. and comparable information concerning the conservation depth of each element.

4.2.1.1 Ultra-conservation Properties

Candidate elements are amplified by human genomic DNA and the heat shock protein 68 (hsp68) and the LacZ-reporter gene in the transgenic mice experiment used to investigate the tissue specificity of putative enhancers, generally of between the dimensions between 200 bp to 2 kb. Candidate elements are Importantly, the Hsp68 promoter alone has no activity in embryonic mouse tissues, but it increases the expression of reporter genes effectively when coupled with tissue-specific enhancers. Before injecting into fertilised mice eggs and subsequently reimplanting it into fake pregnant women, the reporter construct is linearised. Embryo bags for genotyping will be removed on E11.5, embryos will be labelled for the activities of a LacZ reporter for the gene. As a result, the same construct in different (random) sites is likely to have transgenic embryos stably integrated inside the genome. The outcome is that the comparison of numerous distinct transgenic embryos enables cases of expression to be identified due to positions (i.e. their integration close to endogenous regulatory components) which usually result in the reporter's activity patterns seen in a single embryo from the many more transgenic animals that are produced throughout this experiment.

4.2.1.2 Types of Enhancers

Positive enhancers are characterised as locations with consistent gene expression in three or more embryos, while negative enhancers are defined in at least five

transgénic embryos as elements which do not display a reproducible pattern. Positive enhancer tissue specialties for developmental mouse anatomy are defined in terms mainly consistent with the accepted mouse nomenclature.

4.2.1.3 Experimental Dataset

You can use gene-like symbols, accession numbers, or co-ordinates to do simple queries for elements of genomic interest from the main page. These searches are by default restricted to items that contain experimental data. In the 'computational data set' component of the database, the substantially larger data set of predicted enhancers without experimental data is also searchable.

More search options for the experimental dataset may be found in the "advanced query" section. Since all positive patterns of expression of enhancers have been established, groupings of enhancers can be found in certain anatomical zones or tissue that drive expression. It is also possible simply to isolate components which have no measurable activity for enhancer.

Remember that the test records just one developmental time point and does not exclude a negative result which may boost the growth sooner or later of the related element.

4.2.1.4 Technical Background - Directions

An integrated database has been developed for functional in vivo data on human tissue-specific enhancers. This database aims to give the public with access to an extensive and consistent collection of these enhancements for experimental and computational biologists. Conservation of candidate non-coding regions throughout human and other genomes will be utilised to identify applicant non-coding sections for testing. The future ambitions of VISTA Enhancer Browser include improving interoperability with other resources and developing capabilities to store and explore data sets contributed by other users to optimise its utility as a centralised resource for in vivo enhancer data.

4.2.2 Dataset Retrieval

The results of a search are displayed as a list, with each row representing data for a single construct. Three separate entries are supplied in circumstances;

- The ID of each dataset is found in column "ID," with the prefixes "hs" referring to human that was evaluated for enhancer activity.
- Hyperlinked coordinates designate the genomic site in the species from which the enhancer was cloned.
- Bracketing genes which is basically the encapsulation of respective targeted genes – flanking from both side in order to ease the follow up for genome mapping.

As per according to the expression pattern (embryonic stage) given in experimental data for each enhancer; the data set has been retrieved for only NOSE and EYES with the +ve selection of enhancers along with their accurate chromosomal position (validated by USCS genome browser) Fasta sequence and Primer sequence (forward and reverse). The primary focus has been laid down on NOSE and EYES because these two regions represents the Primate Sensory Capabilities which are highly reactive in response to external stimuli.

For sight and scent reaction at very high consistent levels, human and non-human primates react considerably the same. For their behavioural views, scent- and visual signals for Non-Human primates have become important: olfactory participation in social behaviours has been observed in a number of Non-Human primates, including rank-forming and rank-conservation activities, territorial defence, the identification of sexual partners, group recognition and communication of reproductive status. Non-human primates are particularly sensitive to visual stimuli and are widely used to collect visual information on their environment.

So there is an urge to find out why the Human and Non-Human Primates responds approximately the same and what are the genetic reasons behind this logic. Narrowing down our analysis; if we dig in more and associate this all point with facial

musculature system it can be seen that all this is due to genetic changes. Following Table 4.2 is the extraction of 5 out of 103 Enhancers (VISTA Enhancer Browser) that shows positive expression pattern in both NOSE and EYES.

Table 4.2: Dataset retrieval of VISTA ENHANCER (5 OUT OF 103) (refer to Appendix A Table 5.3)

4.3 Formation of Ortholog Sequences

The correlation of genotype with phenotype is one of the main aims of studies that investigate genomic changes in human lines. Since direct human testing is not permitted, scientists in the past rely on variations and chromosomal mutations that have occurred naturally that are different for Homo sapiens from others, with an emphasis on recent progress.

There are various genomic sites which have experienced a major sequence or rearrangement that requires changes in the Homo sapiens genome unique from those seen in other remaining primates. Changes may all be affected by coding areas, non-coding regulatory areas and repeated sequence contents. When comparing human and nonhuman primates, screening the genome for evidence of positive selection is a common strategy for detecting small scale alterations.

4.3.1 Ensembl Genome Browser

This is the correct collection of robust human-genome sequence inspections based on external sources and is available as a dynamic website or as csv files. It is an accurate collection of genetic predictions. It looks to be an open source code infrastructure project to construct a portable system which handles very big genomes and all the associated requirements, from sequencing data to storage and representation of information. The Ensembl site has been the world's major source of annotation of the human genome sequence, and the publishing by the human genome project throughout the world has been analysed extensively.

4.3.2 BLAT (Blast Like Alignment Tool)

Reciprocal Best Hits (RBH) a frequent orthological agent in comparative genomics. The essential discovery of RBH is when proteins encrypted by two genes, which diverge after a speciation event and which are known as "identical genes" in different species, are homologous genes. The BLAST-Like Alignment Tool (BLAT) is employed in genomic sequences corresponding to a protein or DNA sequence supplied to users. BLAT is used most often to detect homologous sequences within the same or closely related species. It was intended to expedite the alignment to the human genome of millions of expressed sequence tags. BLAT accepts a sequence of DNA or protein and only compares sequences for the genome. You can compare a DNA query with a BLAT DNA database. High similarity implies 95% identity or greater for searches of the nucleotide sequence, and 80% or more for the amino acid sequence [25].

4.3.3 Dataset Retrieval

The input sequence data modes requires "Fasta sequence" either DNA or protein, in our case Human Enhancer nucleotide sequence has been pasted one by one (all 103). The next requirement is to search your sequence against given Genomic specie database, upon which comparison has been made out against (Chimpanzee, Gorilla and Rhesus monkey). Ortholog nucleotide sequences has been retrieved from database in .txt format accordingly in a pattern already maintained. Table 4.3 shows only the 5 out of 103 ortholog sequences against human enhancer. **Table 4.3: Dataset of 5 out of 103 Ortholog sequences made by Ensembl** (refer to Table 5.4)

4.4 Transcription Factors and Binding Sites

In order to understand the basic principles of gene regulation, computer techniques are required to anticipate DNA TF binding areas. Genetically speaking,

tools have been developed for identifying the exact binding sites targeted by these transcriptors, techniques such footprinting with DN Ase I, delaying of gel, interference with methylation and protection, to name a few. A fast growing range of transcription factor binding sites (TFBSs), with information regarding the gene in response generally being the first, was also found and mapped to eucaryotist promoters or enhancers. The purpose was also to decrypt the DNA recognition code of DNA-binding area, taking into account their required places and their DNA-binding specificity, for methodically gathering data on transcription factors. If we could figure out, at least within a different class of TFs, how the DNA-binding proteins achieve sequence specificity, we could foresee future binding sites for thousands of TFs for whom we don't know where they bind.

4.4.1 TRANSFAC-gene Xplain Database

In Match-Professional, there are some tissue-specific profiles not in the public edition. Moreover, only the 'best selection' profiles are available to Match-Professional users. The cuts are finished with the best possible series of genuine binding sites from TRANSFAC and these profiles contain the most trustworthy matrices (unlike traditional profiles, which refine cuts to the use of nuclear synthesis methods) (see minFN above).

In Match-Professional a matrix generating tool is supplied that is not truly open to the public for the Match. Users can use this tool to produce a series of aligned sequences to generate their own vectors. It is vital to select the suitable values for the core and matrix similarity cuts when using Match to discover the supposed TF sites. The choice of a cut-off is dictated largely by the objectives of the user. We have calculated three alternative cuts for each matrix in the TRANSFAC database:

- to cut down the amount of false positives (over-prediction error).
- eliminate false negatives (under-prediction mistake); and
- reduce errors greatly.

4.4.2 Match 1.0 Public

Match receives DNA sequences as an input, searches for a PWM library to find potential TF-binding sites and lists potential sites that are detected, along with a graphical representation of their sequence positions. The search results use the matrix-like value (MSS), as well as the core sequence value (CSS) (CSS). These two values vary from 0.0 and 1.0, with an exact match between the 1.0 values, and they measure the match quality between the sequence and the matrix. The first five consecutive positions of the matrix are known as the core of the matrix. In both MSS and CSS scoring, the same formula is employed. You can select the taxa that you want to see from the matrix section (a vertebrate, an insect, a plant, a fungus, or any of the above).

Only by using the option tag 'high quality' may high quality matrices be used. Around 70 percent of all TRANSFAC matrices are represented by TRANSFAC matrices with the lowest false positive rate. The following criterion was used to choose these matrices. When using a matrix with a cut-off that allows 50 per cent false negative, the number of matches found in exon2 sequences (false positive rate) must fall below 1 match per 1 kb. You have three alternatives to choose from with the matrix independent press (minFN, minFP, and minSUM). MatchTM contains a number of important pre-loading functions including several matrices for optimum selection and a choice of tissue / cell-specific (liver, muscle, immune cell) or cell-type practises (cell cycles). For each profile, the TRANSFAC data bank was built for sets of transcription factors known to be active in a specific fabric or process. These transcription factors were then associated with the TRANSFAC matrices. We chose the one with the lowest false-positive rate when there were numerous matrices connected with a transcription factor [26].

4.4.3 Dataset Retrieval

Following the submission of the form to the server, the Match-Professional application searches for TF binding sites using the parameters provided. In the results

table, each match detected by the programme is displayed on its own line. It contains the following information: (I) matrix ID and match position (II) strand +ve or -ve, which indicates the orientation of the matrix in the match, two match scores, related subsequence (III) identification of transcription factors that are connected to the matrix. It's worth remembering that the + strand of the sequence always determines the place of the match. A simulation method of the selected matches is provided after hitting the 'graphic' button. The +ve or -ve placement of the sites, which are displayed above the sequence, correlates to the orientation of the '>' symbol.

The name of the matrix is also given. Given the complexity of regulatory functions maintained by promoters of genes that must be encoded in their structure by a system of TF site combinations, we can expect many more TF sites to somehow be uncovered experimentally in the near term. All Match-Professional search predictions can be used to generate well-supported hypotheses. If we are actually talking about human particular figures, that we have to compare the transcript data set of humans one by one, and the number of transcript factors together with their binding websites have been acquireed and lost. A total of 287 tfactors are involved, but we are more concerned about those tfactors that were gained/specifically gained and lost in humans; the collective gained tfactors are 190 and the lost tfactors collectively are 97. as a whole data set transcription factor. Table 4.4 shows Tfactor dataset among those 5 out of 103 previously selected enhancers.

Table 4.4: Dataset of 5 out 103 Transcription factors along with binding sites made by TRANSFAC-geneXplan database (refer to Appendix A Table 5.5)

4.5 Generation of Human Specific Data

Now we have all the data related to Transcription factors along with their binding sites upon with comparison with Non-Human Primate group that are involved in facial morphology; it's high time to filter out Human Specific data so as to

narrow down the analysis which can ease our path in making concluding remarks regarding where the actual difference lies in between the genome of Human and Non-Human Primate group. This needs manual curation over the huge dataset of Transcription factors one by one; so there is an alternative way where we can first inspect and maintain the separate dataset of all the Transcription factors that are involved solely in Human.

Identifying human specific transcription factors can be prevented by a number of circumstances that make it difficult to compare accurately the genes and the genome sequence. Gene annotation is often sloppy and may change across the builds of genomes, preventing it from being clear whether the changes are true or caused by a computing or assembly error. Modifying the genetics is a key approach for introducing adaptive changes to evolution. Transcriptional splicing changes are typical for changing the gene structure.

Several studies indicated that genes are split differently in human lineage, including changes in brain expression and a wide range of genes involved in metabolism and morphology. Changes in gene expression are a typical kind of evolutionary change and can be due to a range of genetic modifications including changes in regulative DNA affecting promoters, enhancers, and suppressors, and CNV-related dosage variations. These modifications can all impact the gene amounts, timing and even tissue gene expression [27].

The aim of obtaining human specific data is to confirm that the phenotype keeps its genetic presence and absence (NOSE and EYES). There are many evidence suggesting genes play a part in the appearance of the face. It may therefore be most clear that we are more likely to share personality features with close relatives than outsiders when considering own particular groups. With this, few understand how genetic diversity has an impact on forms of differentiated face functions, such the shape and size of nostrils or the distance between our eyes. According to an increasing body, genes are regarded to influence the face.

Table 4.4 has been shown below containing the dataset of major Transcription Factors that are involved in NOSE and EYES (Human Specific). Total of 28

Transcription factors are involved in the phenotypic expression; and then later on this dataset has been compared with the dataset of Transcription factors mentioned in Table 5 (overlap comparison) so that we may have only few Transcription factors (Human Specific) whose presence or absence creates the definite phenotypic change in Human NOSE and EYES shown in Table 4.5.

Table 4.5: 27 Major Transcription factors involved in Nose and Eyes (Human specific) (refer to Appendix A Table 5.6)

Table 4.6: Human specific transcription factors (definite phenotypic expression) (refer to Appendix A Table 5.7)

4.6 Elucidation of Differential Genetic Variations

At this very stage we have Human Enhancers (along with their respective Transcription factors and binding sites); and we have also scrutinized those Transcription factors that are solely responsible for the phenotypic expression in Human NOSE and EYES. The prime objective of this very step is vital for the very fact that gaining and losing of some genetic content makes a specific expression (either in Human or Non-Human Primate group). We are more concerned around Human Specific dataset, so we have to see where exactly the Transcription Content (TFactors and binding sites) lies in the genomic texture of Human-Enhancers.

With the help of the data shown in Table 4.7; it is easier for us to manually find out where the binding site actually reside in the genomic sequence of its respective Human Enhancer. The dataset shown in Table 4.6 is the wholesome of all the Transcription factors involved in facial morphology (Human-specific) and here is the point where we maintained a separate repository for only those transcription factors that are involved in Human Specific (phenotypic expression – NOSE and EYES). The total number of those factors are 6; and here we have used the COLOR CODED TECHNIQUE to manage each of the factor separately and hence made them easier to find out. Each factor has its own identity and name alongside with

their separate allocated color. The details are as follow which shows only those Transcription factors that do overlaps with the already founded factor list mentioned in Table 4.6. (I) SOX (mainly involved in nasolabial angle) red color (II) PAX (mainly involved in retina formation) yellow color (III) OCT (mainly involved in retinal and choroidal vasculature) green color (IV) MYOD (mainly involved in ecoptic eye structue) blue color (V) GATA (mainly involved in lens fibre cells) pink color (VI) SRF (mainly involved in visual cortex: ocular dominance) orange color. This strategy of color coding to find out manually the binding sites of above stated factors inside the genomic sequence of Human Enhancers made a lot much easier for us to find out where to focus on one specific area rather than to idiotically scrolling and exploring the whole genome. Now we have to see which of the above stated factor is appearing more upon which the phenotypic expression has a huge impact. By manually inspecting inside the 5 above stated Enhancers (shown in Table 4.7) we came across this very fact that only 2 transcription factors (OCT and PAX) has impactful existence and do also overlaps in the genome of Non-Human Primate group but shows different genetic expression in order shown as (I) Enhancer 141 (OCT) (II) Enhancer 156 (PAX) (III) Enhancer 1122 (OCT) (IV) Enhancer 1833 (OCT) (V) Enhancer 2328 (OCT)

4.7 DNA – Protein Docking

In a variety of biological activities, DNA and proteins between proteins and proteins are vital. The complex architecture of these interactions is vital to understand, and molecular docking might help. The two most critical organic constituents in the cell are proteins and nucleic acids. All of these linkages are instances of biological and other biological functions such as signal transmission, cellular regulation, protein synthesis, DNA replication and repair and RNA transcription. Therefore, it is necessary to identify their complex architectures in order to comprehend biological processes at the atomic level and to design therapies or medicines. that target these interconnections. Docking is a process of sampling and scoring. In view of two unique structures, the purpose of docking is to analyse

all feasible binding modes of one linked structure. The binding methods sampled were measured during and/or after sampling compared to other optimisation.

This would be tremendously valuable in their research, for strong approaches to protein-DNA docking. The paucity of information necessary to identify the DNA binding interface and the inherent flexibility of DNA prevented efficient docking approaches. Whenever the protein-protein docking is mentioned, the interaction surfaces of the docking partners are typically identified with suitable information (e.g. sequence, sustainability, or biological information). This information is used to support docking and to limit the space to be checked. It seems more difficult to locate the surface of interaction on the DNA than to find the surface of interaction on proteins. There is no common rule of recognition, and global conformance with the industry is indeed an ongoing effort. The future interaction surface of DNA can be influenced. If DNA attaches to a protein, it undergoes major conformation changes, which can radically affect the interface geometry. This has enlarged the overall physiological space to be filled. In order to find acceptable conformations, the region which needs to be studied is growing more and more. DNA flexibility can be divided from the global and local components. The only two basic motions allowing for global flexibility are bending and twisting. It results from a mixture of intricate development of the extensible base and combinations and the backbone of sugar-phosphate. Allowing for internal and external flexibility in DNA while preserving the required shape is a fundamental challenge in protein-DNA docking [27].

4.7.1 HADDOCK 2.4

It is a quantitative doctrinal approach, which directs the empirical and biological data docking process. The success of the NMR-protein–DNA and protein–RNA complexes structure calculation. A local copy of the PDB data base is kept and regularly updated on the HADDOCK web server in order to enhance computational efficiency and minimal service interruption. It should be noted that although users can provide sequence or structure as protein inputs.

There are some differences in the pipeline between the two. The main advantage from this server is that the "workspace save forum" saves all actions and parameters on time.

HADDOCK is an information encoded with AIR energies that reduces the amount of contact space searched and increases the number of single residue solutions is considered active or passive. This information encoded with AIR energy. While AIRs are defined. Active residues, or any other experimentally relevant data, are characterised as those which are vital to interaction based on data from conservation, mutation, or interference with ethylation The solvent-accessible companions of active residues are called passive residues.

4.7.1.1 Scoring Function

For ranking of buildings, the HADDOCK score is employed. It consists of electrostatic covalent bonds (Elec), van der Waals (vdW), solvability (Dsolv) and AIR energy, as well as a BSA (superficial area) term: AIR rigid-body score 14 1.0 * Elec + 1.0 * vdW + 1.0 * AIR rigid-body score 14 1.0 * Dsolv + 1.0 * Dsolv + 1.0 * AIR, final score 141.0 * Elec + 1.0 * Dsolv + 1.0 * Air * Dsolv + 1. A cluster analysis was performed with a minimal cluster sized four for the resulting docking solutions. The clustering cut-off was determined individually for each docking experiment.

Once the protein interface residues are correctly positioned, the root mean square devioation (r.m.s.d.) matrix was built over the interface atoms in the interface of the DNA. The final structures within a cluster were established by the helix shape and the summary base pair and base pair step deformation energy (classified as B-DNA).

4.7.1.2 Input Parameters

For the sake of DNA-Protein docking; keeping this in mind that the PDB structure of Enhancer file has been made manually by SCFBio (we haven't given the whole

nucleotide genomic sequence as an input because it is of no use; rather than we are focused on TFBinding site so we took few bp from left and right respectively and made a separate file for this short Fasta sequence of selected Enhancer). INPUT PARAMETERS has to be set first as per following; From the ensembles of initial structures, for each protein-DNA interaction, a total of 100 structures were constructed.

Each docking attempt was repeated ten times, with the lowest HADDOCK score being maintained. (I) Molecule 1 (PDB structure of Enhancer nucleotide sequence) with the selection of only Chain A. (II) Molecule 2 (PDB structure of respective protein) with the selection of ALL chains. (III) Active site residue selection for Molecule 1 (by encapsulating the TFBinding site sequence in between; do select bp accordingly). (IV) Active site residue selection for Molecule 2 (whole sequence)

4.7.1.3 Output Parameters

The output interface contains best of 10 clusters with different values and scores as per their Cluster size, Coverage area, Vander Waals energy, Conservation energies, RMSD score and Desolvation energy. Out of all best possible structures (automatically the best cluster came up to the top with minimum Z-score). Upon comparison with all the scores; RMSD plot and FS plot can also be seen there for individual clusters and for the combined visualization as well. The best docked complex has been selected on the basis of RMSD scoring (<2.0 is the best selected).

4.7.1.4 PDBe PISA

Now we have the Docked File Complexes of every Human Enhancer; it's prime time to analyse these complexes and visualize the interactions in between DNA and Protein. The interaction can be made authenticated further alongside with the visualization by the help of some parameters served by the server PDBe PISA, available at https://www.ebi.ac.uk/pdbe/pisa/. PDBePISA is a web-based interactive tool for the study of macromolecular interfaces, which may be used to

calculate pre-Calculated findings for the whole PDB database. The results are computed in particular by the structural and chemical properties of the macromolecular surfaces and interfaces and the potentially quaternary structures (assemblies). Use a number of parametres, including multimeric status, symmedium number, space group, accessible/buried surface area, free energy of dissociation, to search PISA for pre computed findings, presence/absence of salt bridges and disulphide bonds, homomeric type, and ligands.

The relevance of macromolecular interactions are determined by keywords (biological function). To download and examine structures, interfaces and assemblies use Rasmol, Unix/linux platforms, Rastop, Windows PCs and Jmol (platform-independent server-side java viewer). "PISA is an interactive tool for macromolecular interface (protein, DNA/RNA, and ligand), quaternary structuring (assembly), structurally comparable database search for interfaces and assemblies, and for searching on different assembly and PDB entry criteria. "PISA is included as a key component of CCP4 (Collaborative Computational Project No. 4). Crystallography software of X- Ray Macromolecular is a collection of macromolecular structural applications used by crystallographers. PDBePISA is an advanced service that analyses macromolecular interfaces in their crystal environment to predict the quaternary structure of your protein. PDBePISA calculates and reduces these assembles to the more physiologically important the predicted strength of each macromolecule-to-macromolecule interacting. PDBePISA can help in the construction of biological assemblies if crystal-symmetric techniques are necessary. PDBe PISA analyse the given input file PDB file by following parameters shown in Table 4.7 below;

Table 4.7: Summary chart of PDBe PISA analysis parameters (refer to Appendix A Table 5.8)

PDBePISA is a sophisticated service that analyses macromolecule interfaces in their crystal environment, allowing you to forecast your protein's quaternary structure.PDBePISA calculates the expected strength of all macromolecule - to - macromolecule interactions and reduces these assemblies to the most physiologically important ones. If crystal symmetry procedures are required to construct biological

assemblies, PDBePISA can help. PDBe PISA analyse the given input file PDB file by following parameters shown in Table 4.7;

4.7.2 Dataset Retrieval

The Docked Complex has been analyzed named as "Interface". It has its own summary chart including (number of residues, number of atoms, solvent accessible area and salvation energy) for our input files with comparison. Interface has been further viewed by the help of PyMol available for download at https://pymol.org/2/ upon which interactions in between DNA and Protein has been labeled properly. The next section describes the Final Results along with their analysis chart at the end of each image where magenta color dictates Protein and red color dictates DNA.

4.7.2.1 Docked File 141-OCT

The interaction can be clearly seen in between DNA nucleotides and Protein residues with the dotted line, which clearly makes sense of OCT transcription factor.

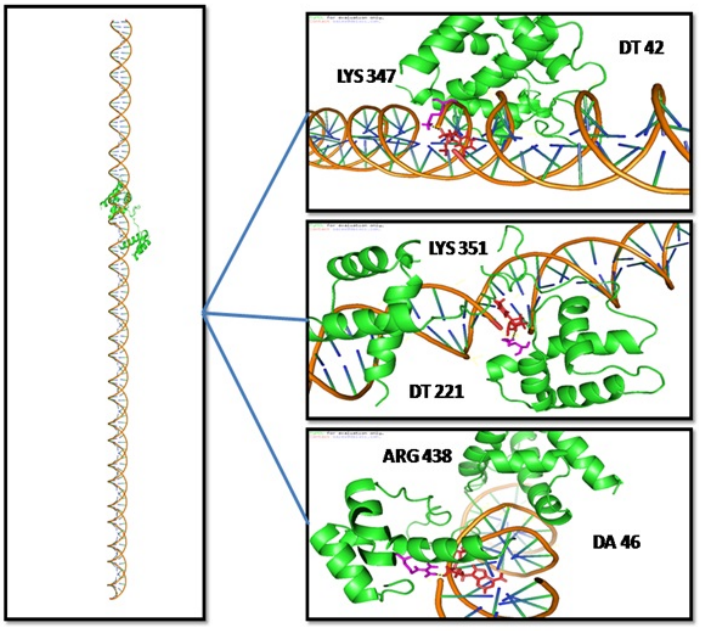

FIGURE 4.1: Docked Complex (DNA-Protein) for Enhancer 141-OCT.

1. LYS 347 forms hydrogen bond with DT 42 maintaining distance of 1.94 Angstrom.

2. LYS 351 forms hydrogen bond with DT 221 maintaining distance of 2.47 Angstrom.

3. ARG 438 forms hydrogen bond with DA 46 maintaining distance of 1.90 Angstrom

4.7.2.2 Docked File 141-OCT

The interaction can be clearly seen in between DNA nucleotides and Protein residues with the dotted line, which clearly makes sense of OCT transcription factor.

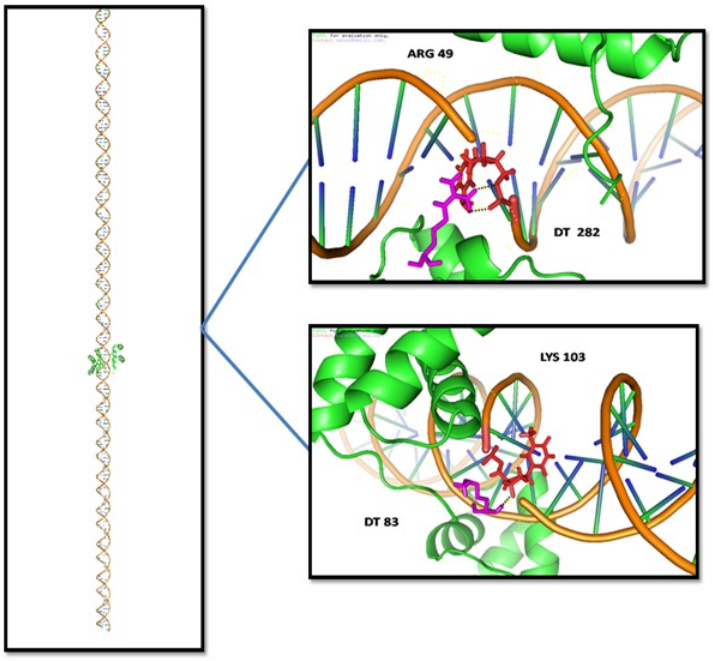

FIGURE 4.2: Docked Complex (DNA-Protein) for Enhancer 141-OCT.

1. ARG 49 forms hydrogen bond with DT 282 maintaining distance of 2.25 Angstrom.

2. LYS 103 forms hydrogen bond with DT 83 maintaining distance of 2.71 Angstrom.

4.7.2.3 Docked File 156-PAX

The interaction can be clearly seen in between DNA nucleotides and Protein residues with the dotted line, which clearly makes sense of OCT transcription factor.

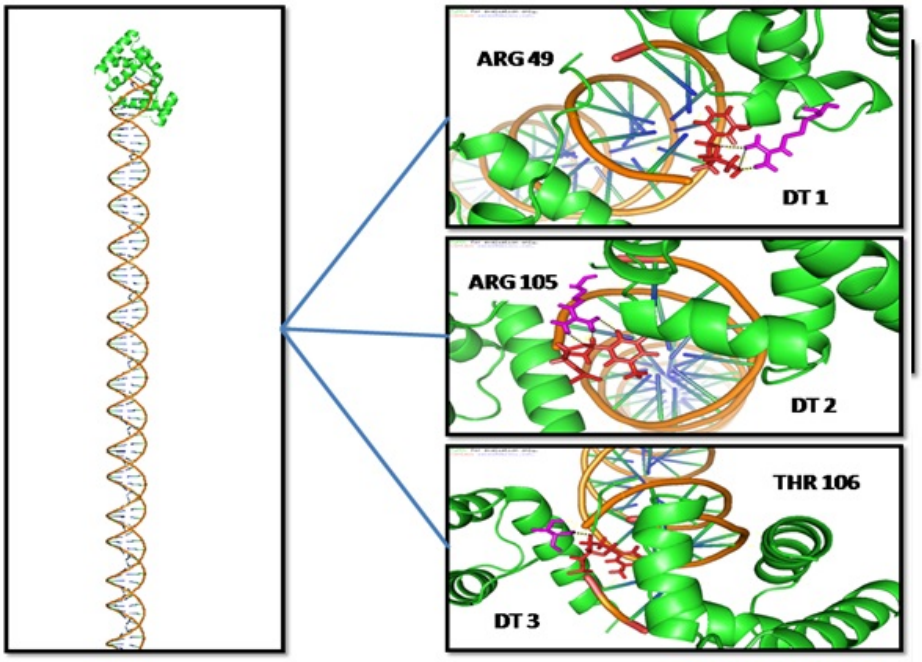

FIGURE 4.3: Docked Complex (DNA-Protein) for Enhancer 156-PAX.

1. ARG 7 forms hydrogen bond with DT 30 maintaining distance of 1.71 Angstrom.
2. LYS 80 forms hydrogen bond with DG 12 maintaining distance of 1.82 Angstrom.
3. ASP 118 forms hydrogen bond with DA 161 maintaining distance of 1.67 Angstrom.

4.7.2.4 Docked File 1122-Oct

The interaction can be clearly seen in between DNA nucleotides and Protein residues with the dotted line, which clearly makes sense of OCT transcription factor.

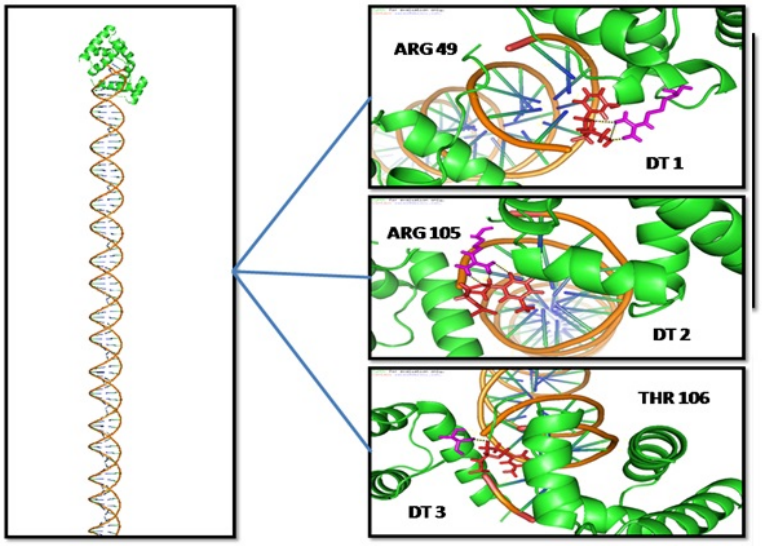

FIGURE 4.4: Docked Complex (DNA-Protein) for Enhancer 1122-Oct.

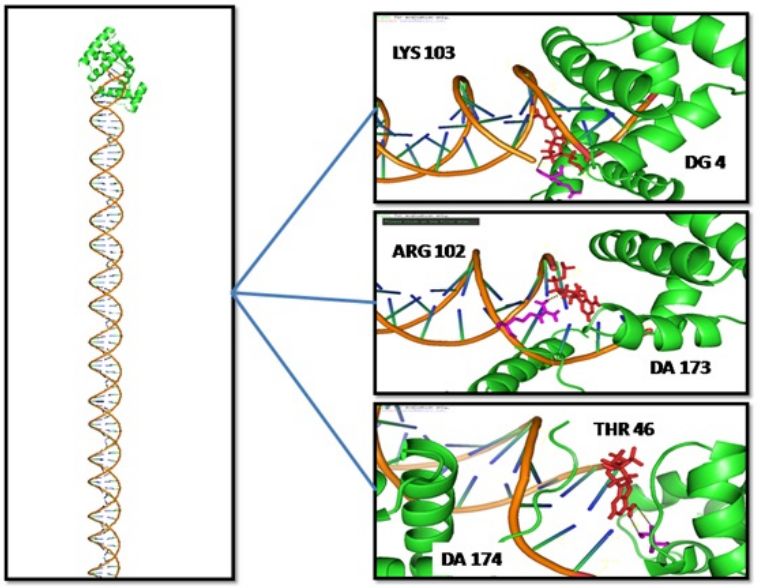

FIGURE 4.5: Docked Complex (DNA-Protein) for Enhancer 156-PAX.

1. ARG 49 forms two hydrogen bonds with DT 1 maintaining distance of 1.64 Angstrom and 2.49 Angstrom.

2. ARG 105 forms two hydrogen bonds with DT 2 maintaining distance of 2.31 Angstrom and 2.33 Angstrom.

3. THR 106 forms hydrogen bond with DT 3 maintaining distance of 3.28 Angstrom.

4. LYS 103 forms hydrogen bond with DG 4 maintaining distance of 3.36 Angstrom.

5. ARG 102 forms hydrogen bond with DA 173 maintaining distance of 2.11 Angstrom.

6. THR 46 forms hydrogen bond with DA 174 maintaining distance of 3.71 Angstrom.

4.7.2.5 Docked File 1833-Oct

The interaction can be clearly seen in between DNA nucleotides and Protein residues with the dotted line, which clearly makes sense of OCT transcription factor.

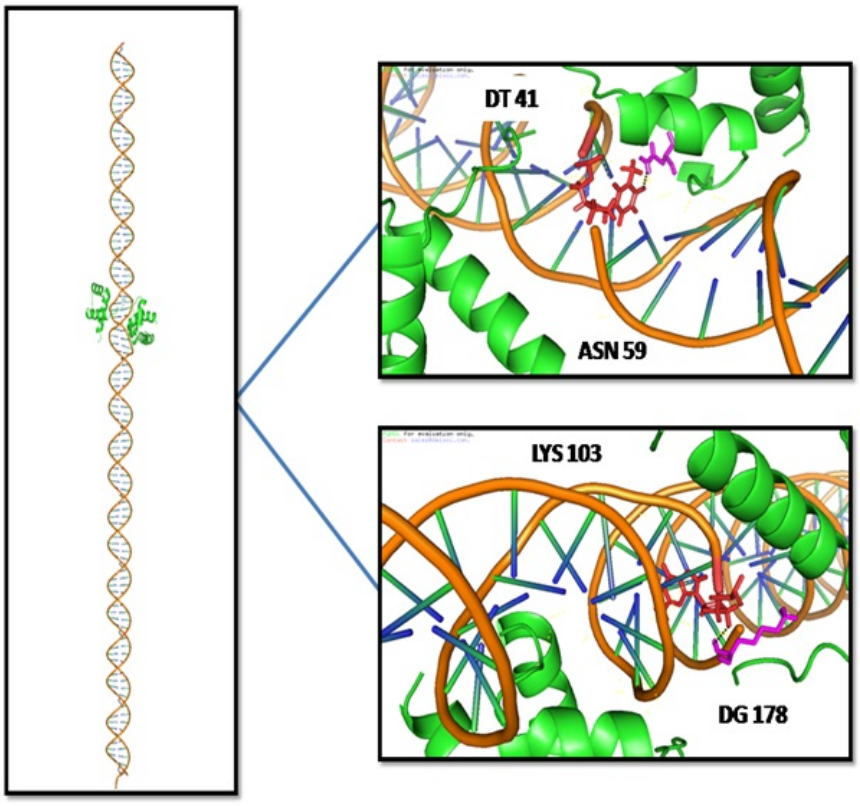

FIGURE 4.6: Docked Complex (DNA-Protein) for Enhancer 1833-Oct.

1. ASN 59 forms hydrogen bond with DT 41 maintaining distance of 2.04 Angstrom.

2. LYS 103 form hydrogen bond with DG 178 maintaining distance of 3.31 Angstrom

3. ASN 54 forms hydrogen bond with DA 173 maintaining distance of 1.87 Angstrom.

4. SER 56 forms hydrogen bond with DA 174 maintaining distance of 1.89 Angstrom.

4.7.2.6 Docked File 2328-Oct

The interaction can be clearly seen in between DNA nucleotides and Protein residues with the dotted line, which clearly makes sense of OCT transcription factor.

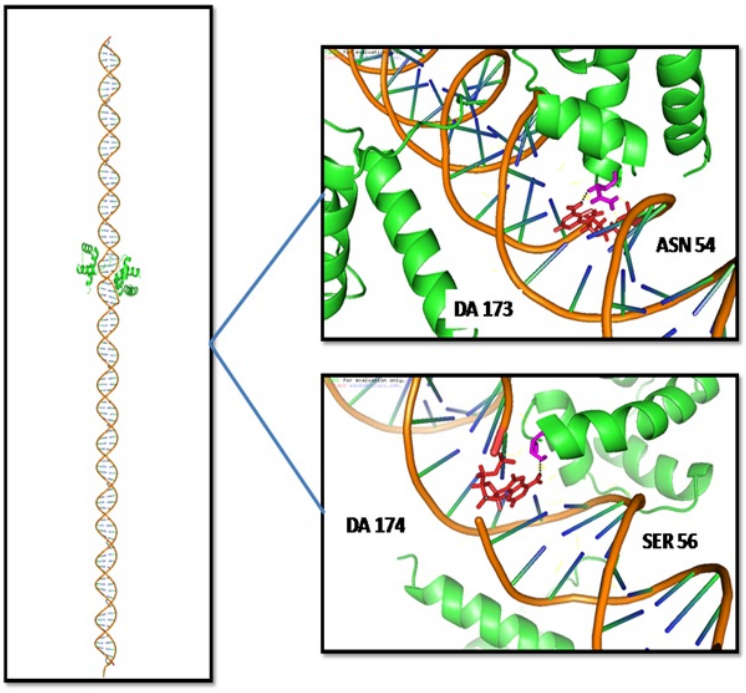

FIGURE 4.7: Docked Complex (DNA-Protein) for Enhancer 1833-Oct.

1. ARG 49 forms hydrogen bond with DC 41 maintaining distance of 2.36 Angstrom.

2. LYS 103 forms hydrogen bond with DG 150 maintaining distance of 3.73 Angstrom.

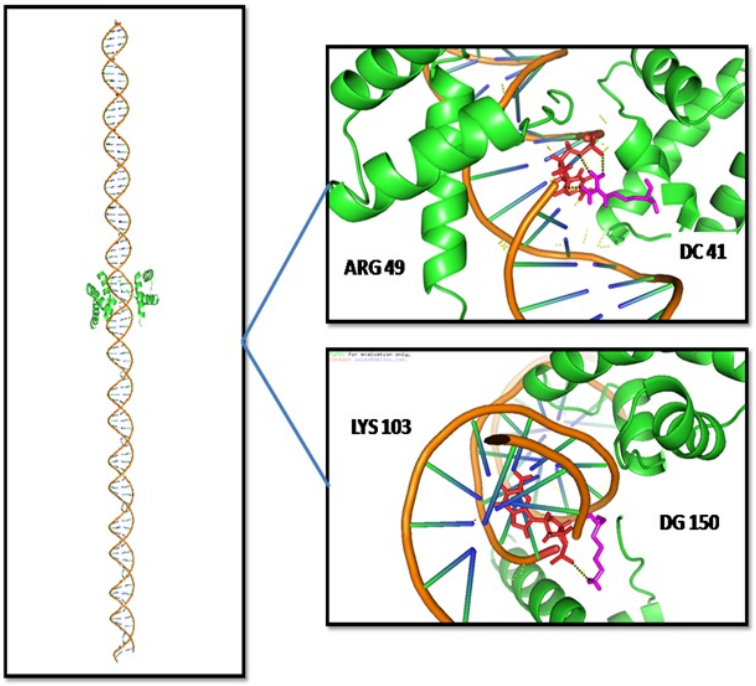

FIGURE 4.8: Docked Complex (DNA-Protein) for Enhancer 2328-Oct.

Further these Docked Complexes can be analyzed by the values/parameters setted by PDBe PISA. The results of the completely automated technique for evaluating the server were identical to those obtained by using separate docking software and methods for scoring protein–DNA interactions manually. To provide a comprehensive database of high-quality macromolecular structures and related information. This is accomplished through creating and maintaining advanced structural bioinformatics databases and services that are at the cutting edge of the field, if not the leading edge. They should be kept up-to-date in order to keep up with the growth of the PDB archive, and they should ideally be available 24 hours a day, 365 days a year. Maintaining in-house knowledge in all key structure determination techniques; in order to stay aware of technical and methodological changes in these domains, as well as collaborating with the community on mutually beneficial issues (For example, data processing, extraction, protocols and specifications, or fundamental report generation).

Table 4.8: PDBe PISA summary chart for individual docked interfaces(refer to Appendix A Table 5.9)

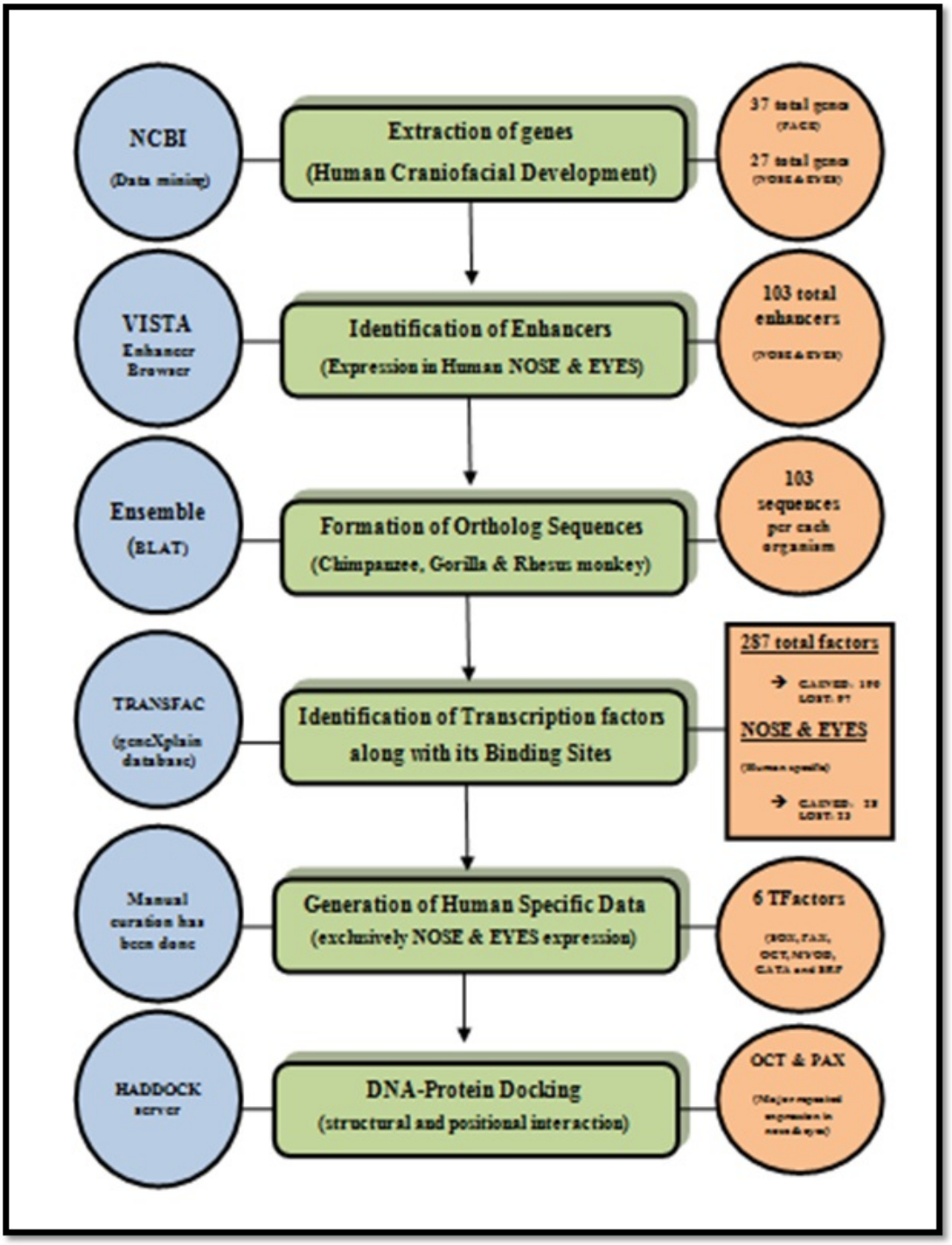

FIGURE 4.9: Summary of Research Project.

Chapter 5

Conclusion and Future Prospects

A greater grasp of the genetic nature can provide insights on orofacial morphological methods, A clearer sense and basis for predicting face features in a range of applications, in light of early identification of health, are the complicated interaction between genetic variations in facial feature problems and birth defects. The cranio-facial complex is initially influenced by complicated signaling networks and carefully timed nascent genetic material's expression, as well as other processes. Hormones, nutritional state, and biomechanical considerations all have an impact on the face as people age. The face demonstrates a modular organization as a natural consequence of these forces and limits underlying face morphology and maturation, there are suites of facial features of various sizes that demonstrate organization's effectiveness but are mostly independent of other aspects. As a result, together with their regulatory regions, we estimate that multiple genes will have a considerable impact on the human face. However, the genetic differences responsible for human face variability are still unknown. Whilst the primary emphasis of genomics study is human genetics and its relevance to illness, comparative primate genomics are being researched. The genomes of non-human primates are vital for the heritage and differences between the same type used as a disease model.

The first objective of this study was to explore and co-relate all the key regulatory genetic information available for facial feature processes in humans; Moreover,

during embryonic development, a number of genes are expressed in important organs. Total of 37 genes are involved in full facial musculature system. The genetic mechanisms can affect natural facial variation and facial feature abnormalities both. Understanding the biological processes that are critical and impactful in nose and eyes; total of 27 genes has been extracted aside that shows expression exclusively in nose and eyes.

The second objective is to extract solely the Enhancers (cis-regulatory elements) that are expressed in nose and eyes of human (as most of the genes shows an impactful expression in this region). Particularly distal enhancers, exhibits tissue specific patterns). Total of 103 enhancers has been finalized that shows distinctive expression. Non-protein coding sequences represent approximately 85% of the human DNA under evolutionary restrictions, and the primary part of this is the cisregulatory regions. Enhancer mutations primarily influence cis transcription, while mutations can influence a wider array of gene expression components, such as RNA processing and stability, protein folding, etc. The discovery of a link between human facial variation, disease, and regulatory element enrichment that differs from that of Chimpanzee (*Pan troglodytes*)s, Gorillas, and Rhesus monkey (*Macaca mulatta*) raises possibility of genetic variety affecting both species-specific and individual facial forms of an overlapping array of regulatory components.

The third objective is to create Lineage Specific Connection in between Human and Non-Human primates (i.e orthologs). This purpose has been accomplished by the formation of ortholog sequences for each of the respective primate upon comparison with human.

The fourth objective is to understand and accomplish the importance of transcriptomics in primate genomics. The genetic connection between facial development genes and quantitative characteristics characterizes major components of the average healthy facial feature complex is demonstrated. These relationships can help us to comprehend the variety of expressive and intensive expression presented in some rare genetic conditions and increase our overall understanding of the variables supporting the great spectrum of human face characteristics. For this purpose to

be fulfilled total of 190 out of 287 transcription factors has been seen as commonality in between Human and Non-Human primates. If we further analyze this data 5 of the common transcription factors (SOX, PAX, OCT, MYOD and SRF) exclusively gained by 28 times and lost 23 times in the whole human genomic data with respect to non-Human primate group.

The fifth objective is to emphasize only on those exclusive transcription factors that can be functionally characterize in human specific phenotypic expression. Another purpose for meticulously collecting information on transcription factors, the DNA-binding specificity was designed to try to decode the DNA-binding domain recognition code. We might be able to forecast prospective hundreds of TFs for which binding locations we don't know where they bind or what they govern if we can figure out how DNA-binding proteins achieve sequence specificity, at least within various Transcriptional Factors classes. Total of 2 transcription factors (OCT and PAX) has been in crown out of all the above data that is majorly involved in the phenotypic expression of human eye and nose which cannot be seen in commonality on the other side of non-Human primate group.There is a need to identify the intricate architecture of these interactions and molecular docking. Proteins and nucleic acids are the two most significant macronutrients in the cell.

The sixth goal is to perform positional and structural docks to observe interactions, either by placing the contact surface of the protein exclusively to identify the DNA, in order to enable a visual representation to gain information about identification and genetic expression. The confirmative complexes show how exactly the protein is attached and expressed in the nucleotide sequence where the protein attaches and expresses itself in humans, as opposed to non-human primates, where the expression is lost.

To conclude this all study in a single frame; the first part of this paper summarizes what is known about the content of primate genomes and the differences between them. Following that, there is some novel insights into genetic differentiation and speciation, with a focus on human evolution (exclusive focus on nose and eyes). Finally, we demonstrate how genetic data is expanding and improving the use

of nonhuman primates in human health and disease research. Previous research has also demonstrated that most or all primates share a substantial proportion of human protein-coding genes. However, until recently, thorough comparisons across all components of the genome were unattainable. Researchers may now investigate the content and function of genomic characteristics across many species, bringing new insight into the genetic basis of similarity and differences in humans and other primates that can gives an insight to treat with respective genetic disease in particular mentioned regions as well.

Bibliography

A. Achilleos, N. T. Huffman, E. Marcinkiewicyz, N. G. Seidah, Q. Chen, S. L. Dallas, P. A. Trainor, and J. P. Gorski, "MBTPS1/SKI-1/S1P proprotein convertase is required for ECM signaling and axial elongation during somitogenesis and vertebral development," Human Molecular Genetics, vol. 24, no. 10, pp. 2884–2898, 2015.

A. Beverdam, A. Brouwer, M. Reijnen, J. Korving, and F. Meijlink, "Severe nasal clefting and abnormal embryonic apoptosis in Alx3/Alx4 double mutant mice," Development, 15-Oct-2001. [Online]. Available: https://journals.biologists.com/dev/article/128/20/3975/41482/ [Accessed: 15-Aug-2021].

A. Burrows, "The facial expression musculature in primates and its evolutionary significance", BioEssays, vol. 30, no. 3, pp. 212-225, 2008. Available: 10.1002/bies.20719 [Accessed 15 August 2021].

A. E. Kel, "MATCHTM: a tool for searching transcription factor binding sites in DNA sequences," Nucleic Acids Research, vol. 31, no. 13, pp. 3576–3579, Jul. 2003, doi: 10.1093/nar/gkg585.

D. Masthan, D. Priya.P, D. Babu and D. Rajesh, "Genetics And Its Association With The Developmental Anomalies Of", European Journal of Molecular & Clinical Medicine, vol. 7, no. 5, 2020. [Accessed 15 August 2021].

E. Lausch et al., "TBX15 Mutations Cause Craniofacial Dysmorphism, Hypoplasia of Scapula and Pelvis, and Short Stature in Cousin Syndrome", The American

Journal of Human Genetics, vol. 83, no. 5, pp. 649-655, 2008. Available: 10.1016/j.ajhg.2008.10.011 [Accessed 15 August 2021].

F. Liu et al., "A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans", PLoS Genetics, vol. 8, no. 9, p. e1002932, 2012. Available: 10.1371/journal.pgen.1002932 [Accessed 15 August 2021].

I. Vieille-Grosjean, P. Hunt, M. Gulisano, E. Boncinelli, and P. Thorogood, "Branchial HOX Gene Expression and Human Craniofacial Development," Developmental Biology, vol. 183, no. 1, pp. 49–60, 1997.

J. Cole and R. Spritz, "The Genetics of Facial Morphology", eLS, pp. 1-9, 2017. Available: 10.1002/9780470015902.a0027240 [Accessed 15 August 2021].

J. Park et al., "Loss of MAFB Function in Humans and Mice Causes Duane Syndrome, Aberrant Extraocular Muscle Innervation, and Inner-Ear Defects", The American Journal of Human Genetics, vol. 98, no. 6, pp. 1220-1227, 2016. Available: 10.1016/j.ajhg.2016.03.023 [Accessed 15 August 2021].

J. Yu, J. M. Madison, S. Mundlos, F. Winston, and B. R. Olsen, "Characterization of a Human Homologue of theSaccharomyces cerevisiaeTranscription Factor Spt3 (SUPT3H)," Genomics, vol. 53, no. 1, pp. 90–96, 1998.

K. Ishizuka et al., "Investigation of Rare Single-Nucleotide PCDH15 Variants in Schizophrenia and Autism Spectrum Disorders," PLOS ONE, vol. 11, no. 4, p. e0153224, Apr. 2016, doi: 10.1371/journal.pone.0153224.

K. L. Jones, M. C. Jones, and M. del Campo, Smiths recognizable patterns of human malformation. Philadelphia, PA: Elsevier Saunders, 2013.

K. Verhoeven, P. D. Jonghe, K. Coen, N. Verpoorten, M. Auer-Grumbach, J. M. Kwon, D. Fitzpatrick, E. Schmedding, E. D. Vriendt, A. Jacobs, V. V. Gerwen, K. Wagner, H.-P. Hartung, and V. Timmerman, "Mutations in the Small GTP-ase Late Endosomal Protein RAB7 Cause Charcot-Marie-Tooth Type 2B Neuropathy," The American Journal of Human Genetics, vol. 72, no. 3, pp. 722–727, 2003.

L. C. Shull, H. M. Kim, E. S. Lencer, S. Goyama, M. Kurokawa, J. C. Costello, K. Jones, and K. B. Artinger, "PRDM proteins control Wnt/β-catenin activity to regulate craniofacial chondrocyte differentiation," bioRxiv, 01-Jan-2021. [Online]. Available: https://www.biorxiv.org/content/ [Accessed: 15-Aug-2021].

L. Pennacchio, W. Bickmore, A. Dean, M. Nobrega and G. Bejerano, "Enhancers: five essential questions", Nature Reviews Genetics, vol. 14, no. 4, pp. 288-295, 2013. Available: 10.1038/nrg3458 [Accessed 15 August 2021].

L. Vissers et al., "Heterozygous Mutations of FREM1 Are Associated with an Increased Risk of Isolated Metopic Craniosynostosis in Humans and Mice", PLoS Genetics, vol. 7, no. 9, p. e1002278, 2011. Available: 10.1371/journal.pgen.1002278 [Accessed 15 August 2021].

M. O'Bleness, V. B. Searles, A. Varki, P. Gagneux, and J. M. Sikela, "Evolution of genetic and genomic features unique to the human lineage," Nature Reviews Genetics, vol. 13, no. 12, pp. 853–866, Nov. 2012, doi: 10.1038/nrg3336.

M. Risley, D. Garrod, M. Henkemeyer and W. McLean, "EphB2 and EphB3 forward signalling are required for palate development", Mechanisms of Development, vol. 126, no. 3-4, pp. 230-239, 2009. Available: 10.1016/j.mod [Accessed 15 August 2021].

M. van Dijk, "Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility," Nucleic Acids Research, vol. 34, no. 11, pp. 3317–3325, Jun. 2006, doi: 10.1093/nar/gkl412.

N. Ward and G. Moreno-Hagelsieb, "Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss?," PLoS ONE, vol. 9, no. 7, p. e101850, Jul. 2014, doi: 10.1371/journal.pone.0101850.

R. Zehra and A. Abbasi, "Homo sapiens-Specific Binding Site Variants within Brain Exclusive Enhancers Are Subject to Accelerated Divergence across Human Population", Genome Biology and Evolution, vol. 10, no. 3, pp. 956-966, 2018. Available: 10.1093/gbe/evy052 [Accessed 15 August 2021].

S. Boehringer et al., "Genetic determination of human facial morphology: links between cleft-lips and normal variation", European Journal of Human Genetics, vol. 19, no. 11, pp. 1192-1197, 2011. Available: 10.1038/ejhg.2011.110 [Accessed 15 August 2021].

S. Richmond, L. Howe, S. Lewis, E. Stergiakouli and A. Zhurov, "Facial Genetics: A Brief Overview", Frontiers in Genetics, vol. 9, 2018. Available: 10.3389/fgene.2018.00462 [Accessed 15 August 2021].

T. Gabaldón and E. V. Koonin, "Functional and evolutionary implications of gene orthology," Nature Reviews Genetics, vol. 14, no. 5, pp. 360–366, 2013.

T. Niethamer, T. Teng, M. Franco, Y. Du, C. Percival and J. Bush, "Aberrant cell segregation in the craniofacial primordium and the emergence of facial dysmorphology in craniofrontonasal syndrome", PLOS Genetics, vol. 16, no. 2, p. e1008300, 2020. Available: 10.1371/journal.pgen.1008300.16. Zinc finger protein 219-like (ZNF219L) and Sox9a regulate synuclein-c2 (sncgb) expression in the developing notochord of zebrafish.

V. E. Zarelli and I. B. Dawid, "Inhibition of neural crest formation by Kctd15 involves regulation of transcription factor AP-2," Proceedings of the National Academy of Sciences, vol. 110, no. 8, pp. 2870–2875, 2013.

V. Melvin, W. Feng, L. Hernandez-Lagunas, K. Artinger and T. Williams, "A morpholino-based screen to identify novel genes involved in craniofacial morphogenesis", Developmental Dynamics, vol. 242, no. 7, pp. 817-831, 2013. Available: 10.1002/dvdy.23969 [Accessed 15 August 2021].

Y.-H. Lee and J.-P. Saint-Jeannet, "Sox9 function in craniofacial development and disease," genesis, vol. 49, no. 4, pp. 200–208, 2011.

Appendix A

Table 5.1: List of reported genes enrolled in facial morphology along with their associated functions.

S.No.	Gene	Facial feature	Facial phenotype	Processes
1.	DCHS2	Ala aperture	Subtle entry to nostrils	Subtle entry to nostrils
2.	PDE8A	Allometry	Overall face shape and width	Overall face shape and width
3.	SCHIP	Centroid size	Face height and width	Face height and width
4.	ASPM	Chin prominence	Mandibular recession in lip	Mandibular recession in lip
5.	DLX6	Chin prominence	Chin button association	Chin button association
6.	EDAR	Chin shape	Chin protrusion	Chin protrusion
7.	HOXD1	Eye shape	Curvature of eyelid	Curvature of eyelid
8.	WDR27	Eye prominence	Eye tail length	Eye tail length
9.	PAX3	Eye width	Distance between eyeballs	Distance between eyeballs

10.	TP63	Eye distance	Eyeballs curvature	Eyeballs curvature
11.	PABP1	Eye junction	Inner canthi distance	Inner canthi distance
12.	TMEM163	Eye depth	Eye height	Eye height
13.	COL17A1	Eye nasion	Distance-eye nasion	Distance-eye nasion
14.	PARK2	Face elevation	Mid face stature	Mid face height
15.	OSR1	Face depth	Right gonion	Right gonion
16.	FOXA1	Face width	Inter tragi distance	Inter tragi distance
17.	MAFB	Face width	Inter tragi distance	Inter tragi distance
18.	TRPC6	Facial depth	Left tragus to nasion	Left tragus to nasion
19.	TBX15	Forehead	Forehead prominence	Forehead prominence
20.	RPS12	Forehead	Central portion of forehead	Central portion of forehead
21.	FREM1	Upper lip	Lip height	Lip height
22.	HOXD	Lip prominence	Narrowing - nasolabial crease	Narrowing - nasolabial
23.	PKDCC	Mental fold	Subtle mandibular crease	Subtle mandibular crease

24.	C5orf50	Eye prominence	Nasion – zygoma position	Nasion – zygoma position
25.	EPHB3	Nose bridge	Nose ridges - prominence	Nose ridges - prominence
26.	ZNF219	Nose prominence	Left ala to nose tip	Left ala to nose tip
27.	SOX9	Nose prominence	Nasal labial angle	Nasal labial angle
28.	CASC17	Nose prominence	Sidewall of nose	Sidewall of nose
29.	KCTD15	Nose tip	Superior tip of nose	Superior tip of nose
30.	PRDM16	Nose width	Nose height	Nose height
31.	DHX35	Alae width	Nose breadth	Nose breadth
32.	SUPT3H	Nose bridge	Nose breadth	Nose breadth
33.	ALX3	Eye width	Intercanthal width	Intercanthal width
34.	GL13	Nose wing	Nose depression	Nose depression
35.	RAB7A	Philtrum	Prominent subnasale	Prominent subnasale
36.	MBTPS1	Facial height	Skeletal pattern	Skeletal pattern
37.	PCDH15	Facial profile	Mandibular profile	Mandibular profile

TABLE 5.2: Functional description of extracted facial genes

S.No.	Gene	Facial Feature	Facial Phenotype	Function
1.	DCHS2	Ala aperture	Subtle entry to nostrils	Cell adhesion
2.	PDE8A	Allometry	face shape and width.	High expression in ectoderm
3.	SCHIP	Centroid size	Face height and width	Transcriptional element.
4.	ASPM	Chin prominence	Mandibular lip recession	mitotic spindle activity
5.	DLX6	Chin prominence	Chin button association	facial feature development
6.	EDAR	Chin shape	Chin protrusion	Encodes receptor family
7.	HOXD1	Eye shape	Curvature of eyelid	Transcription factor
8.	WDR27	Eye prominence	Eye tail length	Protein-protein interactions
9.	PAX3	Eye width	Distance between eyeballs	Multi-potent cell population
10.	TP63	Eye distance	Eyeballs curvature	Epithelial morphogenesis
11.	PABP1	Eye junction	Inner canthi distance	Poly A binding protein
12.	TMEM163	Eye depth	Eye height	Cadherin-zinc transporter
13.	COL17A1	Eye nasion	Distance-eye nasion	facial feature patterning

14.	PARK2	Face height	Mid face height	Proteasomal degradation
15.	OSR1	Face depth	Right gonion	Expression in signal transduction
16.	FOXA1	Face width	Inter tragi distance	Dental development
17.	MAFB	Face width	Inter tragi distance	Patterning of neural crest cells
18.	TRPC6	Facial depth	Left tragus to nasion	Cation channel subunit
19.	TBX15	Forehead	Forehead prominence	Transcription factor - regulations
20.	RPS12	Forehead	portion of forehead	Transcriptional activator
21.	FREM1	Upper lip	Lip height	Expression - medial nasal fuses
22.	HOXD	Lip prominence	nasolabial crease	Morphogenesis
23.	PKDCC	Mental fold	Subtle mandibular crease	Protein coding
24.	C5orf50	Eye prominence	Nasion – zygoma position	facial feature patterning
25.	EPHB3	Nose bridge	Nose ridges - prominence	Ephrin receptors – development

26.	ZNF219	Nose prominence	Left ala to nose tip	Transcriptional partner of SOX9
27.	SOX9	Nose prominence	Nasal labial angle	Chondrocyte differentiation
28.	CASC17	Nose prominence	Sidewall of nose	ncRNA
29.	KCTD15	Nose tip	Superior tip of nose	Protein coding
30.	PRDM16	Nose width	Nose height	Cartilage differentiation
31.	DHX35	Alae width	Nose breadth	Cellular growth – division
32.	SUPT3H	Nose bridge	Nose breadth	Bone development
33.	ALX3	Eye width	Intercanthal width	Pharyngeal arch patterning
34.	GL13	Nose wing	Nose depression	Protein coding
35.	RAB7A	Philtrum	Prominent subnasale	Vesicle traffic endosome
36.	MBTPS1	Facial height	Skeletal pattern	facial feature patterning
37.	PCDH15	Facial profile	Mandibular profile	Signaling pathway

TABLE 5.3: Dataset retrieval of VISTA ENHANCER (5 OUT OF 103)

S. No	**ID**	**Coordinates**	**Bracketing Genes**	**Expression**	**Fasta Sequence**

1.	Hs141	chr13:73,001,083-73,003,053	DACH1-MZT1	eye	Refer to ANNEXURE-I
2.	Hs156	chr16:53,994,058-53,995,943	FTO (intragenic)	nose	Refer to ANNEXURE-I
3.	Hs1122	chr12:103,345,263-103,346,680	PAH-ASCL1	eye	Refer to ANNEXURE-I
4.	Hs1833	chr1:8,130,439-8,131,887	ERRFI1-SLC45A1	eye nose	Refer to ANNEXURE-I
5.	Hs2328	chr7:84,442,634-84,444,146	LOC101927378-SEMA3D	eye nose	Refer to ANNEXURE-I

TABLE 5.4: Dataset of 5 out of 103 Ortholog sequences made by Ensembl

S.No	**ID**	**Fasta sequence**
1.	141	chromosome:Pan_tro_3.0:X:25189269:25191014:1 chromosome:gorGor4:X:25170469:25172214:1 primary_assembly:Mmul_10:X:24835757:24837504:1 Refer to ANNEXURE-II
2.	156	chromosome:Pan_tro_3.0:16:44571721:44574199:1 chromosome:gorGor4:16:44510058:44512534:1 primary_assembly:Mmul_10:20:41085976:41088460:1 Refer to ANNEXURE-II
3.	1122	chromosome:Pan_tro_3.0:12:105809291:105810810:1 chromosome:gorGor4:12:102807166:102808684:1 primary_assembly:Mmul_10:11:102699963:102701485:1 Refer to ANNEXURE-II

4.	1833	chromosome:Pan_tro_3.0:1:7353667:7355142:1 chromosome:gorGor4:1:7738126:7739605:1 primary_assembly:Mmul_10:1:216448211:216449688:-1 Refer to ANNEXURE-II
5.	2328	chromosome:Pan_tro_3.0:7:87642466:87644209:1 chromosome:gorGor4:7:93429215:93430955:-1 primary_assembly:Mmul_10:3:121327064:121328242:-1 Refer to ANNEXURE-II

Table 5.5: Dataset of 5 out 103 Transcription factors along with binding sites made by TRANSFAC-geneXplan database

S.No	Enhancer	TF-HUMAN	TF-CHIMPANZEE (PAN TROGLODYTES)	TF-GORILLA	TF-MACACA mulatta
1.	141	Oct-1 (attaTTTGTatttat)	FOXJ2 (caaaatATTATttt)	FOXJ2 (caaaatATTATttt)	MyoD (aagCAGGTgttg)
		c-Rel (tgggtTTTCC)	Pax-6 (ttagtTTAAGcctgagtttca)	MyoD (aagCAGGTgttg)	Evi-1 (tgctttgaaAAGA Ta)
		Evi-1 (ACAAGataa)	MyoD (aagCAGGTgttg)	Evi-1 (tgctttgaaAAGA Ta)	Pax-6 (agtttgtcatgCGT GAcatgt)
		FOXD3 (gaATGTTttctt)	Evi-1 (tgctttgaaAAGATa)	Pax-6 (agtttgtcatgCGT GAcatgt)	Pax-4 (gtttgTCATGcg tgacatgtc)
		Oct-1 (tctcTTTGCatagtg)	Pax-6 (agtttgtcatgCGTGAcatgt)	Pax-4 (gtttgTCATGcgt gacatgtc)	COMP1 (aattgaTATTGa ccacaaaaaaaa)

			Pax-4 (gtttgTCATGcgtgacatgtc) COMP1 (aattgaTATTGaccacaaaacaca)		
2.	156	Pax-4 (tgggaTCATGtgtgcccagct) Pax-4 (ggttttgacccCTTGActccc) Evi-1 (ACAAGataa) XFD-2 (tgcttgTTTATagt) Nkx2-5 (CACTTga)	HNF-4 (atatgtgCAAAGgcccagg) Pax-4 (ggttttgacccCTTGActccc) Evi-1 (ACAAGataa) XFD-2 (tgcttgTTTATagt) Nkx2-5 (CACTTga)	HNF-4 (atatgtgCAAAGgcccagg) Pax-4 (ggttttgacccCTTGActccc) Evi-1 (ACAAGataa) XFD-2 (tgcttgTTTATagt) Nkx2-5 (CACTTga)	HNF-4 (atatgtgCAAAGgcccaag) Pax-4 (ggttttgacccCTTGActccc) Evi-1 (ACAAGataa) Nkx2-5 (CACTTga) Nkx2-5 (CACTTga)

		Nkx2-5 (CACTTga)	Nkx2-5 (CACTTga)	Nkx2-5 (CACTTga)	Oct-1 (gtgaTTTGC actagt)
			Oct-1 (gtgaTTTGCactagt)	Oct-1 (gtgaTTTGCa ctagt)	
			Pax-4 (agaatTCAGGcctggtttgtg)	Pax-4 (agaatTCAGG cctggtttgtg)	
3.	1122	Oct-1 (tttaTTTTCattctt)	Pax-4 (tgccgTCACGgggcgggt tgc)	USF (gTCACGgggc)	COMP-1 (tttctcCATTGa caactccaaaga)
		Oct-1 (cataatGTAAAtggg)	USF (gTCACGgggc)	Oct-1 (tttaTTTTCattctt)	HNF-4 (cccggtgCAA AGggcctgg)
		SRF (ctccttaTTTGGag)	Oct-1 (cataatGTAAAtggg)	Oct-1 (cataatGTAAAtggg)	Oct-1 (tttaTTTTCattctt)

		Oct-1 (gtggctGCAAAtta)	SRF (ctccttaTTTGGag)	SRF (ctccttaTTTGGag)	Oct-1 (cataatGTAAAtggg)
			Oct-1 (gtggctGCAAAttac)	Oct-1 (gtggctGCAAAttac)	SRF (ctccttaTTTGGag)
					Oct-1 (gtggctGCAAAttac)
4.	1833	Pax-4 (gaaggtagactCCT GAggact)	GATA-1 (tccctTATCTctgc)	GATA-1 (tccctTATCTctgc)	GATA-1 (tccctTATCTctgc)
		Nkx2-5 (tcAAGTG)	Pax-6 (tctttTTATGtataatttttt)	Pax-6 (tctttTTATGtataatttttt)	FOXD3 (aaTTGTTttttt)
		NF-Y (ctaATTGGtta)	Freac-7 (gtattgTTTATatata)	XFD-2 (tatatgTTTATata)	NF-kappaB (ggGGGACttcccat)
		Nkx2-5 (ctTAATTg)	HFH-1 (tattGTTTAtat)	Freac-7 (tatatgTTTATatata)	NF-Y (ctaATTGGtta)
		HFH-3 (tagTGTTTattta)	Pax-4 (gaaagtagactCCTGAgg act)	Pax-4 (gaaggtagactC CTGAggact)	HFH-3 (tagTGTTTattta)

		Pax-4 (ggtcttgaactCCT GAcctca)	Nkx2-5 (tcAAGTG)	Nkx2-5 (tcAAGTG)	
		AP-4 (ctgacctcAGCT Gatcca)	NF-Y (ctaATTGGtta)	NF-Y (ctaATTGGtta)	
		CDP CR1 (cATTGAttgg)	Nkx2-5 (ctTAATTg)	Nkx2-5 (ctTAATTg)	
		NF-Y (ttgATTGGttc)	HFH-3 (tagTGTTTattta)	HFH-3 (tagTGTTTattta)	
		HLF (attctGTAAC)	Pax-4 (ggtcttgaactCCTGAcc tca)	Pax-4 (ggtcttgaactC CTGAcctca)	
		Oct-1 (cataTTTGCatgtgt)		AP-4 (ctgacctcAG CTGatcca)	
5.	2328	Oct- (tacaatGCAAAgcca)	Oct-1 (aataatGCAAAggta)	Oct-1 (aataatGCAA Aggta)	Oct-1 (aataatGCAAA ggta)

FOXJ2
(aaaATAATaattag)

Pax-4
(cgcttTCAAGtg
tttttatcc)

Nkx2-5
(tcAAGTG)

Oct-1
(aaatatGTAAAtgat)

HNF-1
(ttttttaaataTTAACa)

HNF-1
(ttttaaatATTAAca)

Oct-1
(tacaatGCAAAgcca)

FOXJ2
(aaaATAATaattag)

Pax-4
(cgcttTCAAGtgtttttatcc)

Nkx2-5
(tcAAGTG)

HNF-1
(ttttttaaataTTAACa)

HNF-1
(ttttaaatATTAAca)

Oct-1
(tacaatGCAA
Agcca)

FOXJ2
(aaaATAATaattag)

Pax-4
(cgcttTCAA
Gtgtttttatcc)

Nkx2-5
(tcAAGTG)

HNF-1
(ttttttaaataTT
AACa)

HNF-1
(ttttaaatATTAAca)

RFX-1
(acGTTTCcatt
gtaacat)

HNF-3beta
(tgtttTGTTT
gttta)

FOXD3
(ttTTGTTtgttt)

HFH-3
(tttTGTTTgttta)

FOXJ2
(ttgttTGTTTa
tttatga)

FOXD3
(gtTTGTTtattt)

HFH-3
(gttTGTTTattta)

	HNF-4 (ttctgtgCAAA Gaccatgg)

Table 5.6: 27 major Transcription factors involved in Nose and Eyes (Human specific)

S.No	Transcription factor	Role in EYES	Role in NOSE	Reference
1	RPGRIP1L (intragenic)	retina: photoreceptor connecting cilium	-	Delous et al. (2007)
2	DACH1	potential master gene in eye formation	-	Hammond et al. (1998)
3	ELP4 (intragenic)	lens epithelium, retina, and iris	-	Bhatia et al. (2013)
4	ZNF536 (repressor)	lower expression: eye vesicles	-	Qin et al. (2009)
5	OR7E156P	-	olfactory receptors: genes	Man et al. (2004)
6	ZNF32	expression: retinal pericytes	-	Jiang et al. (2020)
7	SOX5	-	lower expression: olfactory bulb	Shim et al. (2012)
8	ASCL1	rod photo-receptor: opsin gene diffrentiation	-	Ahmad (1995)

9	SIM1	formation of supraoptic nuclei	-	Holder et al. (2000)
10	PDE8A	-	nasal prominence	Cole et al. (2016)
11	PAX3	distance between eyeballs	-	Liu et al. (2012)
12	TP63	distance between eyeballs	-	Liu et al. (2012)
13	EYA4	ocular development	-	Borsani et al. (1999)
14	ZNF219 (repressor)	-	nasal prominence	Shaffer et al. (2016)
15	CHD8	-	nasal prominence	Shaffer et al. (2016)
16	SOX9	-	nasal labial angle	Cha et al. (2018)
17	HADC8	eye width	-	Shaffer et al. (2016)
18	PRDM16	-	nose width and height	Shaffer et al. (2016)
19	PAX	retina formation	nose width	C. Grindley et. al (1995)
20	SUPT3H	-	nose breadth	Adhikari et al. (2016)
21	RUNX2	-	nose bridge	Adhikari et al. (2016)
22	ALX3	eye width	-	Shaffer et al. (2016)

23	GL13	-	nose wing	Adhikari et al. (2016)
24	Oct	retinal and choroidal vasculature.	-	Rodríguez et. al (2018)
25	MyoD	ecoptic eye structure	-	Andrew B. Lassar (2017)
26	GATA	lens fiber cells	-	Maeda et. al (2009)
27	SRF	ocular dominance: visual cortex	-	S. Pulimood et. al (2017)

TABLE 5.7: Human Specific Transcription factors (definite phenotypic expression)

S.No	**Enhancer**	**GAINED**	**LOST**
1.	141	Oct-1 (attaTTTGTatttat)	MyoD (aagCAGGTgttg)
		c-Rel (tgggtTTTCC)	Evi-1 (tgctttgaaAAGATa)
		Evi-1 (ACAAGataa)	Pax-6 (agtttgtcatgCGTGAcatgt)
		FOXD3 (gaATGTTttctt)	Pax-4 (gtttgTCATGcgtgacatgtc)
		Oct-1 (tctcTTTGCatagtg)	-
2.	156	Pax-4 (tgggaTCATGtgtgcccagct)	HNF-4 (atatgtgCAAAGgcccagg)
3.	1122	Oct-1 (gtggctGCAAAtta)	Oct-1 (gtggctGCAAAttac)

4.	1833	CDP CR1 (cATTGAttgg)	GATA-1 (tccctTATCTctgc)
		NF-Y (ttgATTGGttc)	-
		HLF (attctGTAAC)	-
		Oct-1 (cataTTTGCatgtgt)	-
5.	2328	Oct-1 (aaatatGTAAAtgat)	Oct-1 (aataatGCAAAggta)

*The rest data for all transcription factors against 103 enhancers has been shown in ANNEXURE-IV

TABLE 5.8: Summary chart of PDBe PISA analysis parameters

S.No	**Parameter**	**Detail**
1.	Range:	000000 Range displays the selection range for the relevant interface structure. The range is just a chain ID as shown in the relevant PDB or user-uploaded file for complete monomeric networks. - is used without a chain ID to denote entire chains. [R]C:N is the ligand, where R is the residue name (ligand), C is the chain ID and N is the residue number. [HEM]-:605 specifies Fe (HEM) having protoporhyrin IX in a chain without a chain identifier and a residue number 605.
2.	iNat	The number of interacting atoms in the given structure.
3.	iNres	With the relevant structure, the number of interacting residues.

4.	Surface Å2	It is the overall surface area of the solvent that can be accessed in square Angstroms.
5.	Interface area in Å2	This can be determined by dividing the difference between the total accessible surface areas of isolated and interface structures by two.
6.	ΔiG	The solvent-free energy gain is displayed in kcal/M at the interface. To determine the value, the difference in total solvent energy is used in isolated and interacting structures. Hydrophobic interfaces or the positive affinity of protein are equivalent to negative values. This figure does not include the effect of completed hydrogen connections and salt bridges over the interface.
7.	ΔiG P-value	It reflects the measured energy gain P-value without solving. The P value analyses the chance to get a lower observed value when atoms of the interface are chosen at randomly from a protein surface, such as the observed interface area. The P-value is an interface specificity metric, reflecting how energy-surprising the interface is.
8.	NHB	The amount of hydrogen connections available throughout the interface is shown. The free energy of protein binding for each hydrogen connection is roughly 0.5 kcal/mol (exact value depends on the calibration procedure and may change with version).

9.	NSB	There are likely number of salt bridges over the interaction. Every salt bridge increases the free energy of protein binding by about. 3 kcal/ mol (exact amount relies on calibration technique and may varies with version)..
10.	NDS	It shows the amount of disulfide bonds that may develop across the contact. Each salt bridge increases its free protein binder energy of around 4.0 kcal/mol (exact value depends on the calibration procedure and may change with version number).
11.	CSS	It is the Significance Score for Complexation which represents how significant the interface is for assembly training. The scoring is defined as the greatest proportion of the total free binding energy corresponding to the interface for stable assemblies.

Table 5.9: PDBe PISA summary chart for individual docked interfaces

S. No			Structure 1 (Protein)	Structure 2 (DNA)	Interface
1.	141-Oct	Number of Atoms iNat	119	135	254
		Number of Residues iNres	33	23	56
		Solvent Accessible Area Surface Å2	1093.4	1103.1	2196
		Solvation Energy (isolated structure)	-109.8	478.4	-
		P-value	0.855	0.396	1.25
		Interface area in Å2	-	-	1098.3
		ΔiG	-	-	-18.1
		ΔiG P-value	-	-	0.58
		NHB	-	-	3
		NSB	-	-	0
		NDS	-	-	0
		CSS	-	-	0
2.	141-Oct	Number of Atoms iNat	99	129	228
		Number of Residues iNres	32	16	48
		Solvent Accessible Area Surface Å2	9379	612228	70607
		Solvation Energy (isolated structure)	-105.8	668.1	-

		P-value	0.93	0.59	1.25
		Interface area in Å2	-	-	1008.9
		ΔiG	-	-	-9.7
		ΔiG P-value	-	-	0.74
		NHB	-	-	2
		NSB	-	-	0
		NDS	-	-	0
		CSS	-	-	0
3.	156-Pax	Number of Atoms iNat	177	210	387
		Number of Residues iNres	52	35	87
		Solvent Accessible Area Surface Å2	14005	28613	42618
		Solvation Energy (isolated structure)	-87.7	291.1	-
		P-value	0.84	0.52	1.25
		Interface area in Å2	-	-	1678.8
		ΔiG	-	-	-28.5
		ΔiG P-value	-	-	0.66
		NHB	-	-	3
		NSB	-	-	0
		NDS	-	-	0

		CSS	-	-	0
4.	1122-Oct	Number of Atoms iNat	105	117	222
		Number of Residues iNres	30	13	43
		Solvent Accessible Area Surface Å2	9245	29677	38922
		Solvation Energy (isolated structure)	-100.5	302.1	-
		P-value	0.94	0.84	1.78
		Interface area in Å2	-	-	1013.8
		ΔiG	-	-	-2.0
		ΔiG P-value	-	-	0.89
		NHB	-	-	8
		NSB	-	-	0
		NDS	-	-	0
		CSS	-	-	0
5.	1833-Oct	Number of Atoms iNat	109	138	247
		Number of Residues iNres	33	16	49
		Solvent Accessible Area Surface Å2	9217	36559	45776
		Solvation Energy (isolated structure)	-102.9	390.4	-
		P-value	0.89	0.57	1.46
		Interface area in Å2	-	-	1119.6

		ΔiG	-	-	-12.5
		ΔiG P-value	-	-	0.71
		NHB	-	-	4
		NSB	-	-	0
		NDS	-	-	0
		CSS	-	-	0
6.	2328-Oct	Number of Atoms iNat	119	139	258
		Number of Residues iNres	33	17	50
		Solvent Accessible Area Surface Å2	9101	31947	41048
		Solvation Energy (isolated structure)	-104.8	343.7	-
		P-value	0.90	0.62	1.52
		Interface area in Å2	-	-	1161.2
		ΔiG	-	-	-10.8
		ΔiG P-value	-	-	0.75
		NHB	-	-	2
		NSB	-	-	0
		NDS	-	-	0
		CSS	-	-	0