UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

COMPARATIVE ANALYSIS OF HUMAN CHROMOSOME 22 CES-DGCR SYNTENIC REGIONS IN CHIMPANZEE, BABOON, BOVINE, MOUSE AND ZEBRAFISH AND EXPRESSION PROFILING IN ZEBRAFISH EARLY DEVELOPMENTAL STAGES USING WHOLE MOUNT *IN SITU* HYBRIDIZATION

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

SHELLY K.OOMMEN

Norman, Oklahoma

2006

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/993

UMI Number: 3207532



UMI Microform 3207532

Copyright 2006 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COMPARATIVE ANALYSIS OF HUMAN CHROMOSOME 22 CES-DGCR SYNTENIC REGIONS IN CHIMPANZEE, BABOON, BOVINE, MOUSE AND ZEBRAFISH AND EXPRESSION PROFILING IN ZEBRAFISH EARLY DEVELOPMENTAL STAGES USING WHOLE MOUNT *IN SITU* HYBRIDIZATION

A Dissertation APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY

_____
Bruce A. Roe

_____
Paul Cook

_____
Ann West

_____
George Richter-Addo

_____
Tyrrell Conway

©Copyright by Shelly K. Oommen 2006
All Rights Reserved.

Acknowledgements

I thank God for all His mercies, guidance and motivation all through my life and helping me to achieve what I am thus far and always.

I would like to express my sincere gratitude and deepest appreciation to my major professor Dr Bruce Roe, whose intellectual and scientific guidance, patience and support has always inspired me to do the best and never limit myself in scientific thinking and thus enabling me to complete my research successfully. I would like to also thank my advisory committee, Drs Paul F Cook, Ann H West, G.B Richter-Addo and Tyrell Conway for their continued support, participation and encouragement during the research. I would like to express special thanks to Dr Han Wang for his guidance, training and collaboration with the zebrafish experiments.

I also would like to extend my thanks to all members of Dr. Roe's laboratory, as each member has been helpful and supportive during my research and making the lab a great environment to be a part of. I would like to sincerely thank my colleagues, Christopher Lau, Jiangfeng Li, Hung-Chun (James) Yu for their team work and challenging constructive discussions while working as a team investigating comparative genomics and whole mount *in situ* experiments. I also thank our informatic team, Steve Kenton, Dr Axin Hua, Jim White and Hongshing Lai for their invaluable help and guidance and especially Dr.Fares Najar whose help and guidance has enabled to make graduate life easier. I would like to sincerely thank the former member of this lab Dr. Stephane Deschamps, for his help, guidance and training which laid the foundation for earlier work carried out during the research. I would

like to express my sincere thanks to the sequencing team, the primer making group and Mounir and his group members for helping in carrying out the research goals smoothly.

I would like to express my heart felt appreciation and sincere thanks to my parents K Kunjummen Verghese and Girly Kunjummen for their priceless and steadfast love, prayers, encouragement, and motivation, which has constantly helped me to do the best and achieve goals. I would like to sincerely thank my elder sister, Dr Susan and her husband, Mathew K George for their constant love, support, prayers and encouragement and my second sister Shylu and her husband, Pradeep Abraham for their constant love, prayers and encouragement and also to all my neices and nephews for their love. I would like to express my sincere thanks to my parents in law A.G Abraham and Chinnamma Abraham and family for their patience and prayers. I would like to express my thanks to my cousin brother Binu C John and family for their love and care and all my relatives, church family and friends both back home and in U.S.A for their prayers, love and encouragement.

I would like to sincerely express my deepest heart felt love and appreciation to my husband Dr.Arnold Abraham for his love, friendship, encouragement and invaluable patience that has enabled me to finish my research successfully.

Table of Contents

ACKNOWLEDGEMENTS…….....iv
TABLE OF CONTENTS……...vi
LIST OF TABLES…….ix
LIST OF FIGURES…..…….....x
LIST OF ABBREVIATIONS…….....xii

ABSTRACT……..xiv
1. INTRODUCTION……...1
1.1 DNA the Genetic Material 1
1.1.1 Historical Background 1
1.1.2. DNA Structure 3
1.2. From Structure to Function 4
1.3. DNA Sequencing Technology 8
1.4. Human Chromosome 22 Cat Eye and DiGeorge Syndrome Region 9
1.4.1. Low Copy Repeat Regions in Human Chromosome 22 10
1.5. Comparative Genomics 13
1.5.1. Significance of Comparison with Both Distantly Related and Closely Related Species 15
1.5.2. Closely and Distantly Related Species Comparison 17
1.5.2.1 Human and chimpanzee (6-8MYA) 17
1.5.2.2 Human and baboon (35-40 MYA) 21
1.5.2.3 Human and cow (85MYA) 23
1.5.2.4 Human and mouse (60/80MYA) 24
1.5.3 Human and Zebrafish (450MYA) 25
1.5.3.1 Zebrafish as the model system for whole mount *in situ* hybridization 25
1.5.3.2 Global gene duplication in zebrafish (350-400MYA) 27
1.5.3.3 Mutation studies in zebrafish 30

2. MATERIALS AND METHODS……..32
2.1 Sequencing of the Chimp, Baboon, Cow, Mouse and Zebrafish Clones Syntenic to Human Chromosome 22 Cat Eye and DiGeorge Region 32
2.1.2 Random Shotgun Sequencing 33
2.1.3 Analysis of the Sequences using Different Tools 36
2.2 Whole Mount *In Situ* Studies using Zebrafish as the Model Organism 38
2.2.1 Isolation of Total RNA 38
2.2.2 RT-PCR and RNA Probe Preparation by *In Vitro* Transcription 39
2.2.3 Single Stranded Exon Specific DNA Probes and Primer Design 41
2.2.4 Whole Mount *In Situ* Hybridization 43

3. RESULTS AND DISCUSSIONS……..46
3.1 Comparative Analysis of Human Chromosome 22 CES-DGCR Region 53
3.1.1 Repetitive Elements 53
3.1.2 Locations of Repeat Element Insertions and Their Significance 57

3.2 Significance of Nucleotide and Amino Acid Changes 62
3.2.1 Substitution Preference between First, Second, and Third Codon Position between Human, Chimpanzee, and Baboon 62
3.2.2 Transitional Changes versus Transversional Changes 64
3.2.3 Synonymous(Ks) versus Nonsynonymous(Ka) Changes 65
3.2.4 Codon Usage Preference between Human, Chimpanzee, and Baboon 67
3.2.5 Substitutions, Insertions and Deletions (indels): Hydrophobic vs Hydrophilic Amino Acid changes 70
3.3 Gene Expression Studies in Zebrafish 78
3.3.1 CES Region Gene Expression Studies 79
3.3.1.1 ENSG00000138860 79
3.3.1.2 Interleukin 17 receptor (IL17R) 81
3.3.1.3 Cat eye syndrome chromosome region, candidate 5 (CECR5) 82
3.3.1.4 Cat eye syndrome chromosome region, candidate 2 (CECR2) 84
3.3.1.5 BCL2-like 13 apoptosis facilitator (BCL2L13) 85
3.3.1.6 O94909/Kiaa0891 87
3.3.1.6.1 Expression of zf1 88
3.3.1.6.1 Expression of zf2 89
3.3.1.6.1 Expression of zf1on (*acerebellar*)*ace* mutant embryo... ……..90
3.3.1.7 Peroxisome biogenesis factor 26 (Q7Z2D7/PEX26) 94
3.3.2 DGCR Region Specific Gene Expressions 95
3.3.2.1 ENSG00000185643 96
3.3.2.2 Solute carrier family 25(mitochondrial carrier; citrate transporter), member 1) (SLC25A1) 97
3.3.2.3 Mitochondrial ribosomal protein L40 (MRPL40) 99
3.3.2.4 T box-1 (TBX1) 100
3.3.2.5 Armadillo repeat gene deletes in velocardiofacial syndrome (ARVCF) 102
3.3.2.6 DiGeorge syndrome critical region gene 8 (DGCR8) 104
3.3.2.7 HpaII tiny fragments locus 9C (Q8IZ69 /HTF9C/ NM_022727).... 106
3.3.2.8 Zinc finger, DHHC-type containing 8 (ZDHHC8) 108
3.3.2.9 ENSG00000182364 110
3.3.2.10 ENSG00000183099 111
3.3.2.11 Scavenger receptor class F, member 2 (SCARF2) 112
3.3.2.12 NM_032775(ENSG00000185214/Q96B68/NP_116164.2/KLH22 (Kelch-like 22)) 114
3.3.2.13 Positive cofactor 2, multiprotein complex) glutamine/Q-rich-associated protein (PCQAP) 117
3.3.2.14 Phosphatidylinositol 4-kinase, catalytic, alpha polypeptide (PIK4CA) 118
3.3. 2.15 v-crk sarcoma virus CT10 oncogene homolog (avian)-like gene (CRKL) 120
3.3.3 PHD finger protein 5A (Phf5a) 122
3.4 Summary 125

4. CONCLUSIONS ……126

4.1 Comparative Analysis between Different Species..... 126
4.2 Genes Postulated to be Involved in Cat-Eye Syndrome (CES) 129
4.3 Genes Postulated to be Involved in DiGeorge Syndrome……...131

5. REFERENCES…….136

6. **APPENDIX……...205**

List of Tables

Chapter 1 ……..... **1**
Table 1.1 Major advances in molecular biology..... 5
Table 1.2 Examples of genes following the DDC model for the maintenance of the extra copies of duplication..... 29

Chapter 3 ……..... **46**
Table 3.1 The chimpanzee BAC clones from the rp43 and ptb libraries syntenic to human chromosome 22 sequenced and their associated marker genes. 47
Table 3.2 The baboon BAC clones from the rp41 library syntenic to human chromosome 22 sequenced and their associated marker genes. 47
Table 3.3 The bovine BAC clones from the rp42 library syntenic to human chromosome 22 sequenced and their associated marker genes. 48
Table 3.4 The mouse BAC clones from the rp21 library syntenic to human chromosome 22 sequenced and their associated marker genes. 48
Table 3.5 BAC The zebrafish BAC clones from the ch211 library syntenic to human chromosomes 22, 5, 11 and 6 sequenced and their associated marker genes.... 48
Table 3.6 List of genes seen in human, chimpanzee and baboon that contain repeat elements in their intronic, 5'UTR, 3'UTR and exonic regions..... 58
Table 3.7 (a,b,c,d,e,f) List of genes showing the insertions and deletions in human, chimpanzee and baboon and also showing the difference in conservation between different species in chimpanzee..... 76
Table 3.8 The table shows the major trends in amino acid substitutions (hydrophilic or hydrophobic) in different genes of human, chimpanzee and baboon..... 77
Table 3.9 TRANSFAC Web Based software used to identify transcription factor binding sites showing conservation of regulatory elements between human and zf1 and human and zf2..... 92

Chapter 4 …….. **126**
Table 4.1 List of genes in the CES region with their expression pattern in zebrafish with reference to earlier studies in human and mouse..... 129
Table 4.2 List of genes grouped based on their expression in relation to the phenotypes noticed in CES syndrome. 130
Table 4.3 List of genes in the DGCR region with their expression pattern in zebrafish. 133
Table 4.4 List of genes grouped based on their expression in relation to the phenotypes noticed in DGCR syndrome……134

Table of Figures

Chapter 3 ……**46**
Figure 3.1 Chimpanzee BACs Studied.. 46
Figure 3.2 Comparison between the nucleotide sequences of the human chromosome 22 CES-DGCR region and syntenic region in chimpanzee. 49
Figure 3.3 Overview of PIP. 50
Figure 3.4 The PIP output highlighting conservation. 51
Figure 3.5 VISTA output highlighting comparison between whole genome shotgun sequence and BAC based sequence. 52
Figure 3.6 Major repeats 54
Figure 3.7 Unique AluY repeats. 55
Figure 3.8 Insertion of SINE repeats in human SNAP29 gene 59
Figure 3.9 PIP output of SNAP29 gene highlighting conservation of exons and deletion in chimpanzee. 60
Figure 3.10 PIP ouptut highlighting unique SINE insertion in human Pex26 gene. .. 61
Figure 3.11 PIP output of Pex26 gene highlighting conservation of exons and deletion in chimpanzee. 61
Figure 3.12 First, second and third position nucleotide changes. 63
Figure 3.13 Substitution rate between human, chimpanzee and baboon. 64
Figure 3.14 Transitional vs Transversional changes 65
Figure 3.15 Codon usage bias in amino acids with six codons. 68
Figure 3.16 Preference for G/C or A/T ending codons 68
Figure 3.17 Phylogenetic tree representation to interpret indels. 70
Figure 3.18 CES region of human chromosome 22..... 79
Figure 3.19 Expression pattern of the ENSG00000138860 gene 80
Figure 3.20 Expression pattern of the IL17R gene. 81
Figure 3.21 Expression pattern of the CECR5 gene. 83
Figure 3.22 Expression pattern of the CECR2 gene. 85
Figure 3.23 Expression pattern for the BCL2L13 gene.. 86
Figure 3.24 zf1 and zf2 gene comparison to human Kiaa0819 gene..... 87
Figure 3.25 Expression pattern of the zf1 gene. 88
Figure 3.26 Expression pattern of the zf2 gene. 89
Figure 3.27 Expression pattern of the zf1 gene in an *ace* mutant 90
Figure 3.28 Expression pattern of the Pex26 gene. 94
Figure 3.29 DGCR region of human chromosome 22. 95
Figure 3.30 Expression pattern of the ENSG00000185643 gene 96
Figure 3.31 Expression pattern of the SLC25A1 gene 98
Figure 3.32 Expression pattern of the MRPL40 gene 100
Figure 3.33 Expression pattern of the TBX1 gene 101
Figure 3.34 Expression pattern of the ARVCF gene 103
Figure 3.35 Expression pattern of the DGCR8 gene 105
Figure 3.36 Expression pattern of the HTF9C gene. 106
Figure 3.37 Expression in the somites and developing myotomes. 107
Figure 3.38 Expression of the ZDHHC8 gene 108

Figure 3.39 Expression for pattern of the ENSG00000182364 gene 110
Figure 3.40 Expression pattern of the ENSG00000183099 gene 112
Figure 3.41 Expression pattern of the SCARF2 gene 113
Figure 3.42 Expression pattern of the NM_032775/ENSG00000185214 gene. 115
Figure 3.43 Enlarged view of the eye photoreceptor cells..... 116
Figure 3.44 Expression pattern of the PCQAP gene..... 117
Figure 3.45 Expression pattern of the PIK4CA gene. 119
Figure 3.46 Expression pattern of the CRKL gene..... 121
Figure 3.47 VISTA output of Phf5a gene. 122
Figure 3.48 PIP output of Phf5a gene. 123
Figure 3.49 Expression pattern of gene Phf5a 124

List of Abbreviations

ALU	-a human DNA repeat element with homology to 7SL RNA
ARVCF	-armadillo repeat gene deletes in velocardiofacial syndrome
B1	-a family of SINEs found in rodents
B1	-a family of SINEs found in rodents
BAC	-Bacterial artificial chromosome
BCL2L13	-BCL2-like 13 (apoptosis facilitator)
BCR	-Breakpoint cluster region
BID	- BH3 interacting domain death agonist
BLAST	-Basic Local Alignment Search Tool
CDC45L	- CDC45 cell division cycle 45-like (S. cerevisiae)
CDCREL-1	- Cell division control related protein 1
CECR1	- cat eye syndrome chromosome region, candidate 1
CECR2	- cat eye syndrome chromosome region, candidate 2
CECR5	- cat eye syndrome chromosome region, candidate 5
CECR6	- cat eye syndrome chromosome region, candidate 6
CES	- Cat Eye Syndrome
CLDN5	- claudin 5 (transmembrane protein deleted in velocardiofacial syndrome)
COMT	- catechol-O-methyltransferase
CRKL	- v-crk sarcoma virus CT10 oncogene homolog (avian)-like
DGCR	- DiGeorge syndrome Critical Region
DGCR14	- DiGeorge syndrome critical region gene 14
DGCR2	- DiGeorge syndrome critical region gene 2
DGCR6L	- DiGeorge syndrome critical region gene 6 like
DGCR8	- DiGeorge syndrome critical region gene 8
ENSG000	-Ensembl Gene name representation
ERV1	- Endogenous retrovirus
GNB1L	- guanine nucleotide binding protein (G protein), beta polypeptide 1-like
GRB2	- growth factor receptor-bound protein 2
HIRA	- HIR histone cell cycle regulation defective homolog A (S. cerevisiae)
HTF9C	- HpaII tiny fragments locus 9C
IL17R	- interleukin 17 receptor
KCNMB3L	- potassium large conductance calcium-activated channel, subfamily M, beta member 3-like
LDLa	- Low Density Lipoprotein Receptor Class A
LINE	- Long Interspersed repeat element, such as L1
LTR	- long terminal repeats
MaLR	- mammalian apparent LTR retrotransposons
MER	- Medium Reiteration frequency Repeat
MICAL3	- microtubule associated monoxygenase, calponin and LIM

	domain containing 3
MIL1	- Bcl-2-like 13 protein (MIL1 protein) (Bcl-rambo).
MIR	-Mammalian-wide Interspersed Repeat
MLT1H	- an old retrovirus-like LTR
mRNA	- messenger ribonucleic acid
MRPL40	- mitochondrial ribosomal protein L40
PCQAP	- PC2 (positive cofactor 2, multiprotein complex) glutamine/Q-rich-associated protein
PCR	-Polymerase chain reaction
PEX26	- peroxisome biogenesis factor 26
Phred/Phrap	- Phil's read editor/ Phil's read assembly program
PIK4CA	- phosphatidylinositol 4-kinase, catalytic, alpha polypeptide
PRODHL	- proline dehydrogenase (oxidase)-1 like
RANBP1	- RAN binding protein 1
RNA	-ribonucleic acid
Rnase	- ribonuclease
SCARF2	- scavenger receptor class F, member 2
SERPIND1	- serpin peptidase inhibitor, clade D (heparin cofactor), member1
SINE	- Short Interspersed Element
SLC25A1	- solute carrier family 25,member 1
SNAP29	- synaptosomal-associated protein, 29kDa
SNP	-Single Nucleotide Polymorphism
STK22A	- serine/threonine kinase 22A
TBX1	- T-box 1
TOMM40	- translocase of outer membrane 40 homolog (yeast)
TUBA8	- tubulin, alpha 8
TXNRD2	- thioredoxin reductase 2
UFD1L	- ubiquitin fusion degradation 1-like
USP18	- ubiquitin specific peptidase 18
UTR	- untranslated region
X-gal	- 5-bromo-4-chloro-indolyl-b-D-galactosidase
ZDHHC8	- zinc finger, DHHC-type containing 8
ZNF74	- zinc finger protein 74

Abstract

Comparative genomic analysis is a powerful tool that can illuminate the genomic sequence features that result in the changes that drive evolution. In this dissertation, the 4.5 Mb region proximal to the centromere of human chromosome 22 that encodes the contiguous Cat Eye Syndrome and DiGeorge-Velocardiofacial Syndrome (CES-DGCR/VCFS) Critical Regions and the orthologous regions from chimpanzee, baboon, cow, mice and zebrafish have been sequenced and compared. Overall the human and chimpanzee sequences were ~98.5% identical and the human-baboon sequences were ~92% identical at the nucleotide level. A high degree of conservation was observed in both the gene order and the coding region sequences for these synteny regions, with a lower degree of conservation in the intronic and intergenic regions. The conserved structural features likely represent conserved functional properties while the observed differences must be responsible for portions of the human and primate specific phenotypes. The region studied was slightly larger in humans than in chimpanzees and baboons, since the human lineage had a higher insertion frequency relative to the other primates (or the other primates have a higher deletion frequency compared to humans). By comparing the sequenced regions of the chimpanzee genome from three different individual chimpanzees, Clint (ch251), Donald (rp43) and Gon (ptb), the first major conclusion from this dissertation research is that these three chimpanzees differ from each other by ~1.2%, almost as much as humans differ from chimpanzees.

The majority of the amino acid substitutions in humans, chimpanzees, baboons and bovines are changes from hydrophilic to hydrophilic amino acids. The observed human and chimpanzee substitution rate was 1.2% and that between humans and baboons was 2.6%, with Ka/Ks ratios for human and chimpanzee at 0.44 and human and baboon at 0.48 . Thus, the second major conclusion from this present work is that at least in the case of humans vs primates, the genes are evolving by purifying selection.

The final series of experiments were based on the earlier observation that 16 genes in the human chromosome 22 CES-DGCR region had reported expression but no detailed expression profiles while 6 others had no known expression profiles. Through the comparative sequencing and subsequent whole mount *in situ* studies reported in this dissertation, expression of these 22 genes was observed to occur during zebrafish development, mainly during early developmental stages followed by either decreased or no expression in later stages in the brain, ear, eyes, heart, pharyngeal arches, liver, and kidney, all organs related to anomalies resulting in phenotype observed in CES-DGCR patients. Therefore, the third major conclusions from this present work is that contrary to prior studies pointing to single gene alterations resulting in these diseases, it now is clear that both CES and DGCR are multigene-based diseases.

1. Introduction

1.1 DNA the Genetic Material

1.1.1 Historical Background

The quest to find the relationship between structure and function in living things started as early as the 17th century with the discovery of the circulatory system by anatomist William Harvey. However, it was not until the late 1830s that Matthias Schleiden and Theodor Schwann proposed the cell theory. In the 1860s it was suggested that hereditary transmission is through sperm and egg, and in 1868, Haeckal postulated that the nucleus was important for heredity because the sperm consisted largely of nuclear material. In 1896, Edmund Beecher Wilson suggested that 'nuclein' was the substance of inheritance, and in 1874, Johanan Fredrich Miescher isolated nuclei from human pus cells and salmon sperms. In 1859, Charles Darwin's publication of the Origin of Species marked the first step in describing how heredity operates in large populations of living things but it lacked an explanation of how traits were transmitted from generation to generation. In 1865, Mendel's work with the garden pea (*Pisum sativum*) provided a theoretical model of inheritance, as he demonstrated that inherited characteristics are carried by discrete units or factors (now known as genes as coined by Johannsen) that are resorted in different ways in each generation. In the first decade of the 20th century, based on Mendel's theory of inheritance and the cell biologist's knowledge of chromosomes, the science of genetics was born. In 1902, Walter Sutton and Boveri independently observed that chromosomes and genes occur as pairs that separated in a similar fashion during gamete formation, and paired again during fertilization. They also suggested that genes are located on the chromosomes.

This chromosomal theory of inheritance postulated by Sutton in 1903 was refined during the period 1910 through 1940 by the classical genetic studies by Thomas Hunt Morgan and his students C.B.Bridges, H.J. Muller and A.H. Sturtevant. In 1910, Morgan's study of the common fruit fly *Drosophila melanogaster* (Morgan, 1910) proved not only the chromosomal theory of inheritance but also helped in understanding the various aspects of meiosis. By the 1940s, the field of molecular genetics began to grow rapidly with the discovery and chromosomal location of several genes. However, little was known about the chemical makeup of the material that constitutes a gene or how genes functioned, although there was significant speculation prior to 1940. Here French geneticist L.H. Cuenot had speculated about gene function as early as 1903 when he proposed that due to the action of different genes, differences in coat color were inherited in mice. Then in 1909, as a result of his studies on the human disease alkaptonuria and other human diseases categorized as "inborn errors of metabolism", Garrod proposed the relationship between gene and enzyme to perform a biochemical reaction. He speculated that the gene controlled the synthesis of the enzyme and a defective gene inherited at birth can disrupt the enzyme function (Garrod, 1908). Later, in 1928, Fred Griffith's transformation experiment with smooth (S) and rough (R) strains of Pneumococcus showed that some "transforming principle" transferred from virulent dead S strain that converted the avirulent R strain to a virulent strain by synthesizing a new polysaccharide coat (Griffith,1928). However, it was not until 1941 that G.W.Beadle and E.L.Tatum clearly showed the correlation between gene and enzyme with their work on X-ray induced mutants of the bread mold *Neurospora crassa* and proposed the "one gene-one enzyme" theory (Beadle and Tatum, 1941). In 1944, Avery, Macleod, and McCarty chemically characterized that the "transforming principle" was DNA through

a series of protease and nuclease digestion experiments (Avery *et al*., 1944) showing that DNA was the genetic material. DNA was confirmed to be the heredity material in 1952, when Hershey and Chase further solidified this concept by demonstrating that DNA is a component of T2 bacteriophage (Hershey and Chase, 1952).

1.1.2. DNA Structure

The biological information in a genome is encoded in the nucleotide sequence of its DNA or RNA molecule. The DNA (deoxyribonucleic acid) of an organism is the hereditary material of all living organisms that makes up its genome (with the exception that viruses have an RNA genome) and the primary unit of this genetic information is the genes. However, it was not until 1953, when Watson and Crick reported that DNA has a double helical structure consisting of two complementary polynucleotide chains that form a right handed helix in an antiparallel fashion, that the field of molecular biology truly began (Watson and Crick, 1953). This work was based on interpreting the x-ray diffraction data from DNA fibers obtained by M.H.F. Wilkins and Rosalind Franklin (Wilkins, 1951) and Erwin Chargaff's observation in 1950 that the number of molecules of adenine equals to the number of thymine and the number of molecules of guanine equals to that of cytosine (Chargaff, 1951). It thus was established that polynucleotide chain consists of four nucleotides, two purines, namely adenine (A) and guanine (G) and two pyrimidines cytosine (C) and thymine (T), and the distance between the adjacent nucleotides is 3.4 Aº (10 nucleotides per turn), with an A base pairing with T and G base pairing with C via 2 and 3 hydrogen bonds respectively. Each nucleotide in a DNA sequence consists of a pentose sugar (deoxyribose) with either a purine or pyrimidine attached at position 1 of the

sugar and a phosphate covalently linked to the sugar by asymmetrical 5'-3'phosphodiester bonds. The direction of the phosphodiester bonds determines the character of the molecules, the sequence from 5'→3' being different from the sequence 3'→5'. DNA sequence by convention is written in the order they are transcribed i.e. from the 5'end (upstream direction) to the 3' end (downstream direction) (Micklos *et al*., 1990; Cooper, 1994).

The DNA content in the haploid genome is a characteristic of each living organism. Genome size (Cavalier-Smith, 1985) is roughly correlated with the complexity and there is high variability not only in the DNA content but also in the number and the size of the chromosomes. There is a lack of correlation between total genome size and organism complexity; this phenomenon, termed the C-value paradox, occurs because of the presence of variable amount of repetitive DNA in different genomes (Gall, 1981, Gregory, 2001). Furthermore, with our present day knowledge of several eukaryotic genomic sequences, it is clear that these genomes mostly consist of noncoding DNA with repeated elements as the major component (Britten and Kohne, 1968).

1.2. From Structure to Function

In 1958, Francis Crick proposed the central dogma of Molecular Biology in which DNA and the subsequent protein sequences are colinear and the DNA directs its own replication as well as its transcription to RNA (RNA which specifies protein synthesis is mRNA) and RNA being translated to proteins. The more recent advances in our understanding of "central dogma", shown in the table 1.1, both confirm and extend the hypothesis that genetic information flows from DNA to RNA to proteins.

Charles Yanofsky and Seymour Benzer (1954)	Mutations in the genes of E.coli and T4 bacteriophage produced parallel changes in amino acid sequence
Zamecnik and Hoagland (1958)	Showed that tRNA was the adaptor molecule responsible to attach to amino acids during translation and order it on the template mRNA
Robert Holley (Holley *et al*., 1965).	Proposed the cloverleaf structure for tRNA
Jacob and Monod (1961)	Model of gene expression in bacteria.
Benjamin Hall and Sol Spiegelman (1964)	Showed specific DNA sequences are transcribed into complementary mRNAs
S.Weiss and J.Hurwitz (1960)	Discovery of DNA dependant RNA polymerase
William (1977)	Showed binding of RNA polymerase to DNA in prokaryotes
Crick and Sydney Brenner (1961)	Triplet genetic code for each amino acid
Nirenberg and Khorana (1966)	Complete cracking of the genetic code
A.Claude (1930) and P.Zamecnik (1950)	Ribosomes as the site of protein synthesis

Table 1.1 Major advances in molecular biology

Transcription, the process of RNA synthesis, is catalyzed by the enzyme RNA polymerase. There are three eukaryotic RNA polymerases: RNA polymerase I, localized in the nucleolus and synthesizes rRNA (where rRNA is processed and assembled into ribosomes), RNA polymerase II, localized in the nucleoplasm and synthesizes heterogeneous nuclear RNA (hnRNA) the precursor for mRNA; and RNA polymerase III, localized in the nucleoplasm and synthesizes tRNAs, 5S rRNAs, 7SL RNA, U6 small nuclear (sn)RNA and other small stable RNAs involved in RNA processing (Hamsey and Reinberg, 1999; Paul and White, 2000). Transcription initiation by RNA polymerases requires several transcription factors that recognize specific consensus sequences located either upstream or downstream of the transcription start site. Hence there are specific consensus sequences in promoter region, such as the TATA or Hogness box, with a consensus sequence of TATAA that is located 25-30 base pair upstream from the transcription start site. The TATA box is recognized by TFIID and facilitates the correct positioning of RNA polymerase II for

transcription initiation. Further upstream from the TATA at -75 is the CAAT box (GGNCAATCT) that plays a role in determining the efficiency of the promoter (Fickett and Hatzigeorgiou, 1997). An additional enhancer sequence found further upstream at -90 is the GC box (GGGCGG) that often is present in multiple copies in the promoter and can occur in either orientation (Werner,1999). One or more of these GC enhancer elements are required for proper functioning of almost all promoters and are used to control gene transcription in both a temporal (developmental stage specific) and a spatial (tissue specific) framework. Two classes of transcription factors are recognized by the core and enhancer promoter elements. The general or basal transcription factors (general transcription factors: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) bind to the core promoter close to the transcription initiation site while gene specific activators (SP1, NF1, Oct-1) binds to the enhancer elements upstream of the transcription start site at varying distances (Wray *et al.*, 2003). Together, these transcription factors and RNA pol II contribute to the assembly of a transcription preinitiation complex of transcription (Dermitzakis and Clark, 2002). The primary transcripts, termed heterologous nuclear RNAs, are formed as a result of transcription, but are not immediately functional in protein synthesis. The production of mature, functional mRNA requires RNA processing at the 5' ends, and polyadenylation, i.e. the addition of approximately 20-250 adenine residues, at the 3' end of the message. All eukaryotic mRNA have a three distinct domains, a 5'untranslated region (5'UTR), a coding region (amino acid coding triplet codons) and a 3'untranslated region (3'UTR). The 5' and 3' UTR of the mRNA play an important role in post-transcriptional regulation of gene expression as they control transport of mRNA from the nucleus to the cytoplasm,

translation efficiency, and the final subcellular localization and mRNA stability (Proudfoot, 2000).

Ribosomes are responsible for the translation of mRNA into proteins. The amino acid sequence of the final product is dictated by the nucleotide sequence of the mRNA whose codon are recognized by tRNA molecules with a specific amino acid esterified to the 3' free hydroxyl group by a specific aminacyl-tRNA synthetase. The anticodon of the tRNA, base pairs with the codons of the mRNA to specify the amino acids to add to the growing peptide chain. Proteins are synthesized from the amino to carboxyl direction in three stages of synthesis namely initiation, elongation and termination. Initially the 40S subunit binds with the initiator aminoacyl tRNA and the initiation factors (IF) at the 5' end cap of the mRNA and slides along till the start codon AUG is recognized and 60S subunit joins to form the 80S ribosome, that then reads the mRNA in the 5'→3' direction. The Kozak sequence GCCRCCAUGG (Kozak, 1989) where R represents two purines that flank AUG, often is necessary for the recognition of the correct AUG initiation codon (Kozak, 1999). Elongation then occurs when the amino acids are incorporated successively into the growing polypeptide chain by the condensation reaction catalyzed by peptidyl-transferase activity (present as a part of large ribosomal RNA between the amino group of the incoming amino acid and the carboxyl group of the last amino acid incorporated), thus forming a peptide bond between successive residues. Translation continues until a termination codon (UAA, UAG, UGA) lacking a cognate tRNA is reached. This results in the binding of release factor followed by disruption of the multi-subunit ribosome structure, the removal of

the N-terminal methionine, and the release of the completed protein into the cytoplasm. (Li, 1997; Saccone and Pesole, 2003)

1.3. DNA Sequencing Technology

Within the past decade, DNA sequencing has been transformed into a convenient, routinely performed technique. The discovery of a large number of type II restriction endonucleases, initially provided the means to isolate small homogenous pieces of DNA for sequence analysis. Later improvements in polyacrylamide gel separation techniques (Sanger and Coulson, 1978) along with the discovery of DNA base specific chemical cleavage methods (Maxam and Gilbert, 1977) and improvements in DNA strand synthesis methods (Sanger *et al.*, 1977), have contributed to the development of rapid DNA sequencing. Since the original Sanger sequencing method was introduced, advancements resulting in improvements in DNA sequencing (Mardis and Roe, 1989) sensitivity and efficiency that include new enzymes for DNA sequencing (Mead *et al.*,1991; Tabor *et al*., 1987; Stenesh and Roe, 1972), optimizing of the fragment set separation (Sanger and Coulson, 1978; Bankier and Barrell, 1989; Brown, 1984), fluorescence labeling detection methods and cycle sequencing (Smith *et al*., 1986; Gaxton, 1991) and improved shotgun sequencing strategy (Deininger *et al*., 1983; Bodenteich *et al.*, 1994), automated instruments for large scale sequencing (McBride *et al.*, 1989; Ansorge *et al*., 1987; Craxton, 1991) and improved sequence assembly and sequence alignment programs (Hyatt and Uberbacher, 2002; Galas and McCormack, 2002). These improvements have lead to increased yields and reproducibility in both the template preparations and sequencing reaction products and

the development of the high throughput DNA sequencing instruments. The resulting increased amount of sequence data now are catalogued in large genomic databases and annotated using both computer based and biological analysis approaches.

1.4. Human Chromosome 22 Cat Eye and DiGeorge Syndrome Region

The DNA is packaged into chromosomes and in the 23 human chromosomes, six are acrocentric (13, 14, 15, 21, 22, and Y), five are metacentric (1, 3, 16, 19 and 10), and the remaining are submetacentric. Human chromosome 22, one of the five human acrocentric chromosomes, is the second smallest and comprises 1.6-1.8% of the genomic DNA. The short arm of chromosome 22p contains multiple rRNA genes but lacks protein coding genes. In contrast, the long arm, 22q, is the gene rich euchromatic region (Dunham *et al.*, 1999). Several congenital anomaly disorders (McDermid *et al*., 2002) are associated with the regions of 22q. These include DiGeorge and Velocardiofacial syndromes (VCFS), diseases that result from microdeletions that occur once in every 4000 births (Wilson *et al*., 1993). These microdeletions affect cardiovascular, neurological, psychiatric, endocrinological, immunological, palatal abnormalities and characteristic facial features. Over 90% of the patients with VCFS typically delete a 3Mb region, and 7% have a 1.5 Mb nested deletion with the proximal breakpoint region similar to that of the 3Mb deletion (Carlson *et al*., 1997). Cat Eye syndrome is caused as a result of an inverted duplication of the proximal 22q11 region leading to a bisatellited supernumerary chromosome and a partial tetrasomy that spans the p-arm and a part of the 22q11 (Footz *et al.*, 2001; McDermid *et al.*, 1986). CES is

characterized by congenital defects including anal atresia, ocular coloboma, preauricular tags /pits, heart and kidney defects, dysmorphic facial features and mental retardation. Derivative 22, or Der 22, is a genetic disease caused by a partial trisomy of both 11q23-qter and 22q11-qter that results in mental retardation and often multiple anomaly disorders (Fraccaro *et al.*, 1980; Zackai *et al*., 1980). Low copy repeats (LCRs) are found in the vicinity of all three 22q11 breakpoint regions with their size ranging from 40-350kb and an identity of 97-98%. These LCRs differ in gene content and organization but often include 11 known genes or pseudogenes that have the potential for recombination leading to duplication or deletion of the above discussed chromosomal region (Shaikh *et al.*, 2001; Edelmann *et al*., 1999).

1.4.1. Low Copy Repeat Regions in Human Chromosome 22

Chromosomal rearrangement (recurrent and nonrecurrent) breakpoints are found mainly in pericentromeric and subtelomeric regions, and often occur within intervals containing LCRs or AT rich palindromes or pericentromic repeats. LCRs constitute 5% of the human genome, 10-500kb in size with a sequence identity of >95% (Stankiewicz and Lupski, 2002) and are involved in meiotic non-allelic homologous recombination (NAHR), the mechanism that underlies genome rearrangements associated diseases. Thus, in most of the human genomic disorders such as Charcot Marie Tooth disease type 1A (Wise *et al.*,1993), neurofibromatosis type 1 (NF1) (Regnier *et al*., 1997), Williams-Beuren (WBS) (Robinson *et al.*, 1996; Francke *et al.*,1999), Smith Magnesis (SMS) (Greenberg *et al.*, 1991; Juyal *et al.*, 1996), dup(17)(p11.2p11.2), Cat Eye and

DiGeorge syndromes, azoospermia and male infertility (as a result of Y chromosome deletion) (Repping *et al*., 2002), are caused by chromosomal deletions or duplications with LCRs acting as recombination hotspots.(Christine and Lupski, 2004) . Reciprocal deletions and duplications are the result of nonallelic homologous recombination (NAHR) in LCRs on the same chromosome in direct orientation, but inversions take place when a NAHR is between LCRs in inverted orientation. These reciprocal translocations often occur as result of NAHR between LCRs on different chromosomes as seen between chromosome 21q21.3-qter and recipients 13 and 18qter (Stankiewicz *et al*., 2003). Since the LCRs resulted from segmental duplications of specific gene rich regions of the genome that then were mutated to gene fragments and/or pseudogenes or repeat gene clusters, they recently have been termed as duplicons (Ji, Y *et al*., 2000).

The main biological outcomes of genomic duplications are that they provide the potential for genetic diseases based on spontaneous chromosomal rearrangements or they can alter gene evolution pathways by the interstitial deletions and reciprocal translocations that occur within them (Ji, Y *et al.*, 2000).

For example the 3Mb typically deleted chromosome 22 DGCR region includes four copies (Edelmann *et al.*, 1999) of chromosome 22 specific LCRs, namely LCR-A ~350kb, LCR-B ~135kb, LCR-C ~ LCR-D ~250kb (Shaikh *et al.*, 2000; Shaikh *et al.*, 2001; Spiteri *et al.*, 2003). LCR-A and LCR-D are near or at the end points of 3Mb typically deleted in DGS/VCFS/CAFS syndromes (Emanuel *et al.*, 1998) and CES duplications localize to especially LCR-A and also LCR-D (McTaggart *et al*.,1998).

Other LCRs also contain repetitive LINEs (long interspersed repetitive elements) and SINEs (short interspersed nuclear elements). The SINE family of

transposable elements include Alu elements that form 10% of the human genome (Chen *et al.*, 2002; Jurka *et al.*, 1993), usually is 280-300 bp long (Lander *et al.,* 2001) and may be transcribed by RNA polymerase III (Duncan *et al*., 1979). Alu repeats are dimeric free left alu monomer and free right alu monomer fusions that originated from a primitive monomeric alu which is a DNA complementary to a 141bp internally deleted processed 7SL RNA. Alu sequences thus have two (G + C) rich similar fragments, that are linked by a variable A-rich region and that terminates with a variable length 3'-poly(A) tail (Quentin, 1992; Jurka and Zuckerkandl, 1991; Ullu and Tschudi, 1984).

Alu elements occur mostly in the introns, often in the 3'UTR rather than the 5'UTR, but rarely occur in the gene coding region (Jun *et al.*, 2004) (Chen *et al.*, 2002). As a result, intron gain or loss is greatly influenced by Alu repeats and depending on the class of Alu element inserted, the insertion is either of recent or distant evolutionary origin. Chromosome 22q11 LCRs have Alu SINEs sequences at the junctions of genes (or pseudogenes) found within them (Babcock *et al*., 2003) and the younger subfamilies AluY and AluS are noticed near or within the junctions of segmental duplications throughout the human genome (Bailey *et al*., 2003). The oldest primate subfamily of alu sequence is AluJ and both LCRs are believed to have evolved 35 million years ago at the same time of the burst of primate Alu retroposon activity (Bailey *et al.*, 2003) and segmental duplications during primate speciation that resulted in multiple LCRs duplications known (Ji, Y *et al.*, 2000). It also has been shown that two pericentric inversion breakpoints in chimpanzee and (4;19) translocation in gorilla are localized in

the LCRs of the orthologous chromosomal regions (Stankiewicz *et al.*, 2001; Locke et al, 2003)..

Thus, there have been numerous studies aimed at understanding and describing the content and evolution of the LCR regions of human chromosome 22q11 that indicate that these LCR regions are predisposed to rearrangements resulting in congenital anomaly disorders, and that the expansion of these regions can cause both genome rearrangements and gene amplification. Orthologs of human chromosome 22 in non-human primates, i.e. chimpanzee, gorilla, and rhesus monkey also have duplications containing elements similar to the chromosome 22 specific LCRs. Studying the variation in both number and organization of these LCR regions between human and primates will help us understand the evolution and predisposition to duplication mediated rearrangements associated with human disease (Babcock *et al.*, 2003; Shaikh *et al.*, 2001, Shaikh *et al.*, 2000; Shaw *et al.*, 2004).

1.5. Comparative Genomics

With the discovery of the double helical structure of DNA, Watson and Crick anticipated that the arrangement of the order of the four DNA bases determined the genetic makeup of all living organisms. This gave rise to the field of molecular biology and saw a growing number of scientists interested in comparing nucleic acids and proteins of one species to another. Interestingly, studies in many different biology disciplines such as anatomy, biochemistry, pharmacology, immunology, and cell biology, are rooted in comparative biology (Nobrega and Pennachio, 2003). Genomics is a recent branch of biology that uses comparison to analyze individual genomic

sequences and provides information into both genomic structure and genome function (Ohtsuka *et al.*, 2003). With the completion of the human genome sequence (Lander *et al.*, 2001; Venter *et al.*, 2001), efforts have been made to completely annotate and understand the coding regions (Batzoglou *et al.*, 2000; Alexandersson *et al.*, 2003; Korf *et al.*, 2001), the functional noncoding (Loots *et al.*, 2000; Stojanovic *et al*, 1999; Wasserman *et al.*, 2000) and the quarter of the genome that contains noncoding intronic regions (Hare and Palumbi, 2003; Williams *et al.*, 2003) as well as intergenic and regulatory regions (Gottgens *et al.*, 2000; Williams *et al.*, 2003; Kent and Zahler, 2000). Comparative analysis is useful for annotating evolutionarily conserved genomic regions because functionally important regions generally are well conserved across different species, and non-functional regions accumulate mutations or identify totally unknown genes that are specific to that species. Functional regions do not accumulate mutation as rapidly as non-functional regions because there is evolutionary pressure to maintain the important encoded biological function (Onyango *et al.*, 2000). Comparative genomics is based on the hypothesis that genomes being compared have a common evolutionary ancestor. Therefore, cross-species comparison spanning wide evolutionarily distant species and closely related species can help distinguish functionally important sequences such as regulatory elements and coding sequences conserved over evolutionary time that have a shared ancestry (Loots *et al.*, 2000; Thomas *et al*., 2003; Rijnkels *et al.*, 2003; Dubchak *et al.*, 2000).

1.5.1. Significance of Comparison with Both Distantly Related and Closely Related Species

Comparing the sequences of distantly related species genomes often can identify functional sequence elements if they have been conserved over evolutionary time. In contrast, comparative genomics of closely related species also provides unique genotype to phenotype correlations. However because of their high sequence similarity in orthologous regions over a relatively short time since they shared a common ancestor, there is very little sequence divergence and only small species specific changes will be observed. This makes it difficult to identify the functional elements (Boffelli *et al*., 2004; O'Brien *et al.*, 1999; Nobrega *et al*., 2003; Thomas *et al.*, 2002). Therefore, for a clear annotation of the human genomic sequence, comparisons with an increasingly wide variety of species between the range of extremely close (primates) will help identify the regions where divergence is tolerated by looking specifically for differences rather than similarities. In evolutionarily distant vertebrates and more distant invertebrate species similarities will help identify highly functionally conserved sequences. For example, the strictly functionally conserved sequences other than sequence conserved by chance were highlighted in studies of 12Mb genomic sequence from 12 species, in region orthologous to a 1.8MB region in human chromosome 7. This 1.8MB region includes the cystic fibrosis (CFTR) gene (Thomas *et al.*, 2003), and comparative analysis of CFTR genes of human, cows and pigs helped to identify intronic regulatory elements that were not possible with the distant human-fugu comparison or the evolutionarily closer human-sheep comparison (Williams *et al*., 2003). Recent comparative genomic studies on analysis of stem cell leukemia SCL loci also indicates the benefits of extending comparison between the phylogenetic distances of the human comparison to chicken (Gottgens *et al*., 2002) as additional functional

enhancers easily could be identified. Earlier studies of the SCL locus in human and mice identified four new genes as well as several conserved regulatory elements (Gottgens *et al*., 2001). In contrast, a human–marsupial comparison was effective in identifying regulatory elements in genes paralogous to SCL and LYL1 (lymphoblastic leukemia derived sequence-1) (Chapman *et al*., 2003) while a human-fish comparison was beneficial in identifying cis-regulatory modules of Homeobox (HOX) genes (Aparicio *et al*.,1995), a human-chicken comparison helped identify the collagen XII gene promoter that further explained the regulation of the expression pattern of the extracellular matrix component (Chiquet *et al*., 1998). In addition, a recent pig-human-mouse comparison analyzed the porcine INS-IGF2-H19 gene cluster, an important Quantitative Trait Locus(QTL) primarily affecting the development of muscles (Amarger *et al*., 2002). Cross species sequence comparison helped analyze and identify the regulatory elements of interleukins 4, 13 and 5 (Loots *et al*., 2000). More recently, an alternative method but similar method for analysis of regulatory modules in closely related species, termed phylogenetic shadowing, was described (Boffelli *et al*., 2003).

1.5.2. Closely and Distantly Related Species Comparison

1.5.2.1 Human and chimpanzee (6-8MYA)

Human and chimpanzee are estimated to share a common ancestor approximately 6-8 MYA (millions years ago) (Chen and Li, 2001; Glazko and Nei, 2003). A comparison of the human and chimpanzee genomes show differences from our closest relative, help to understand human specific diseases (Olson and Varki,

2003), and provide population genetics information (Chen and Li, 2001; Ruvolo, 1997). In their 1975 study, King and Wilson reported that there is 99% similarity between average human and chimpanzee based on DNA hybridization experiments (King and Wilson, 1975). Although highly similar at the molecular level, humans and chimpanzees differ far more than two humans. Protein coding region changes can give rise to new genes by small insertions and deletions and nucleotide substitutions as well as by gene duplications, exon shuffling, retro transposition and gene fusions. The primate genome sequence will provide evidence about the sequences of these events that occurred during the time since humans and chimpanzee shared a common ancestor (Long *et al*., 2003). The availability of sequences for human and chimpanzee genomic sequences already have shown that small sequence changes occur within coding regions and that only 95% of the genomic chimpanzee DNA could be aligned to humans (Britten, 2002) in comparison to earlier estimates (Goodman, 1990). These newer difference estimates result mainly from insertions and deletions, i.e. indels, (3.4%), rather than from single base pair change (1.4%). For example, comparison of human (Major Histocompatibility Complex) MHC class I region to that of chimpanzee showed deletions of 95kb between human (major histocompatability complex class I chain related (MIC-A and B functional transcripts)) MICA and MICB genes when compared to the single hybrid chimpanzee MIC gene (Anzai *et al.*,2003). Other recent studies have observed additional genomic variations that include insertion of human endogenous retrovirus K provirus (HERV-K) in chimpanzee, bonobo and gorilla, but not in human (Barbulescu *et al*., 2001) indicating that chimpanzee, bonobo and gorilla are closer to each other in evolutionary time than they are to humans.

Other differences have been observed in repetitive elements (Bailey *et al.*, 2003; Nergadze, G.S *et al.*, 2004, Hamdi, *et al*.,1999) and retroviral or transposon sequences (Sverdlov, 2000), as well as in multigene families such as the immunoglobulin and olfactory receptor gene families in human that are duplicated and chromosomally rearranged (Nei *et al*.,1997). Differences in cell surface of sialic acid (N-glycolyl-neuraminic acid (Neu5Gc) between human and chimpanzee also has been reported. It has been postulated that the changes in Neu5Gc observed in chimpanzee makes them less susceptible to pathogens or alters how it recognizes pathogens and cellular signals (Muchmore *et al*., 1998; Varki, 2000; Olson and Varki, 2003). Thus, the differences between human and chimpanzee can be seen in chromosomal organization, copy numbers, locations and functional status of a small number of individual genes within multigene families (Nei *et al*., 1997), a few altered coding regions that result in different gene products in both the species (Frazer *et al*., 2003; Hacia, 2001), differences in intron loss or gain (Cho *et al.*, 2004; Coghlan *et al.*,2004; Babenko *et al*., 2004; Fedorov *et al*., 2003), differences in the number and distribution of interspersed repeats (Smit,1999; Kim *et al.*, 1999) and different transposable elements in the human genome (Gagneux and Varki. 2001). Thus, although comparing the primate and human genomes should provide an understanding of the species specific characteristics based on gene structure, further gene expression studies will be needed to fully understand genome function (Karaman *et al.*, 2003; Khaitovich *et al*., 2004). Additional studies into diseases that differ in frequency and severity between chimpanzee and human (AIDS, Alzheimer's, cancer, malaria and reproductive disorders) and understanding the

genetic factors in these differences will be important in developing approaches to control and eventually cure these diseases (Varki, 2000).

Another aspect of human-chimpanzee diversity is variation in the position and sequence of repeats. Segmental duplications constitute about 5% of the human genome (Bailey *et al*., 2001) and either are intrachromosomal (seen in chromosome 15, 16, 17, 21 and 22) or interchromosomal duplications. In human, segmental duplications occur with at least two >90 identical copies of an ~100kb to 200kb per chromosome (Bailey *et al*., 2004). Segmental duplications also are hotspots for chromosomal rearrangements such as a chimpanzee specific inversion ((Locke *et al*., 2003), a human specific inversion of chromosome 18 (Goidts *et al*., 2004) and the human chromosome 2 fusion region (Fan *et al.*, 2002). Segmental duplications have occurred several different times during primate evolution. For example, humans and chimpanzees have two copies of Charcot-Marie-Tooth neuropathy type 1A repeat sequence while gorilla has only one (Keller *et al*., 1999). Similarly human and African apes have 2 copies of creatine transporter SLC6A8 and the adrenoleukodystrophy gene at the Xq28 as well as the HS.135840 gene at 4q24 while only a single copy is present in orangutans and old world monkeys (Courseaux *et al.*, 2001; Lupski *et al.*, 1998, Eichler *et al.*, 2001) .

Information about nucleotide substitutions also has been obtained by sequence comparisons between closely related species (Webster *et al.*, 2003; Hellmann *et al*., 2003a; Hellmann *et al*., 2003), and insertions and deletions (indels) are the commonly reported sequence differences between humans and chimpanzee (Watanabe *et al.*, 2004). Because it has been reported that human chromosome 21 is 9% smaller than chimpanzee chromosome 22 (Frazer *et al.,* 2003), it has been proposed that this

difference is due to insertion of mobile elements, deletion of regions by unequal recombination, or different mutation rates introduced during DNA replication (Ebersberger *et al*., 2002). Except for the human chromosome 21 and the chimpanzee chromosome 22 studies described above, much of the comparative primate genomic sequence data is based on relatively small sequenced regions.

The main differences between human and primates include single nucleotide replacement; insertions, deletions and larger duplications, and differing distribution and abundance of other repetitive elements and endogenous retroviruses. Through the studies described in this dissertation, the comparison of additional genomic regions will allow for a more in depth genomic comparison.

1.5.2.2 Human and baboon (35-40 MYA)

Human and baboon (old world monkey) have been estimated to have shared a common ancestor approximately 35-40 MYA (Stewart and Disotell, 1998; Goodman, M 1999). The sequence information of the genome of more distantly related primate will determine which of the differences between human and chimpanzee occurred due to mutations in the human lineage conserved by natural selection and thus provide an outgroup for human and chimpanzee comparison. It also has been observed that the duplications seen in human happened during the radiation of the ape lineage and since these duplications are not seen in old world monkeys, their sequence will give the evidence to the ancestral segments that lead to human genome. DNA sequences

(Caccone and Powell 1989), specific gene sequences, genetic loci arrangement on chromosomes (Graves *et al*., 1995), wide range of physiological characteristics that will help in the analyses of the interaction between gene-gene and its environment (Blanjero, 1993; Van deBerg and Williams-Blangero, 1996;) neurophysiological function (Carey and Rice, 1996; Kaplan *et al*., 1995; Higley *et al*., 1993) show great similarity between human and baboon, thus indicating genetic similarity and the close evolutionary relationship between baboons and human. The genomes of human and baboon are larger compared to chimpanzee due to the presence of more Alu repeat elements and L1 elements and also due to the increased rate of insertion in baboon compared to chimpanzee (Liu *et al*., 2003). Seven human autosomes have the same loci order as their baboon homologs and a recent evolutionary fusion unique to human gave rise to human chromosome 2 that exists independently in two baboon chromosomes 12 and 13 (Ijdo *et al*., 1991). In contrast, baboon chromosome 3 was formed by the fusion of baboon ortholog of human chromosomes 7 and 21 (Best *et al*., 1998). Similarly, baboon chromosome 7 is homologous to both human chromosome 14 and 15 (van Oorshot and VandeBerg, 1991), while baboon chromosome 10 is a fusion of human chromosome 20 and 22 (Rogers *et al*., 2000). The close evolutionary relationship between human and baboon also has been confirmed by other studies. For example, when a 29,920 human cDNA filter array was used to compare human and baboon bone marrow $CD34^{+}$ cell global expression, the results showed a similar expression pattern and gene abundance with a difference of less than 3% (Gomes *et al.*, 2001). Baboons also have been useful in studies involved in cholesterol metabolism and a region in baboon chromosome 18 has been associated with HDL levels (Cox *et al*., 2002).

Similarly, other experiments have reported the genetic component responsible for an atherogenic response to environmental stimuli (MacCleur *et al*., 1988; Blangero *et al*., 1990), cortical bone thickness, peak bone density (Kammerer *et al.*, 1995) and relative organ weight (Mahaney *et al*., 1993). Atherosclerosis has been studied in baboon as the genes encoding various serum apolipoproteins and the LCAT and LPL genes encoding lipid metabolizing enzymes are 97-98% similar in the coding region between baboon and human (Hixson *et al*., 1993a to b). Baboons also are important models for human osteoporosis (Jerome *et al*., 1986), for human aging (Martin *et al.*, 2002; Jayashankar *et al*., 2003) as well as for the development of vaccines for meningitis, HIV, influenza, tuberculosis, hepatitis B, and several other infectious diseases.

1.5.2.3 Human and cow (85MYA) humchr22: cow BTA17 and 5

It has been estimated that human and cow shared a common ancestor ~85 MYA and because of this recent divergence, the cow has played very important role in studying human diseases (Fries and Ruvinsky, 1999; Sanger, 1959 ; Stahmann *et al*., 1941; Collip, 1925; Wiltbank *et al*., 1961; Evans and Long, 1921). The genomic sequences of bovine-human orthologous regions can help in extrapolating information from veterinary science to human medicine and also in annotating human genome for conserved genes and conserved noncoding regulatory elements (Thomas and Touchman, 2002;Williams *et al.*, 2003).

The first generation comparative genetic maps of human and bovine created by radiation hybrid mapping and EST sequencing identified 768 genes of which 638 had

human orthologs, 105 conserved segments in between the two genomes and the location of 41 translocations and 54 internal rearrangements (Band *et al*., 2000). The second generation map showed 20% improvement to the first map with 1463 genes, 195 conserved segments of which 31 are newly conserved (Wind *et al*., 2004). The comparison between human and cattle was informative in providing insights into mammalian chromosome evolution as seen in the highly conserved gene order of the distal part of bovine chromosome 18 and Human chromosome 19q, which was different in mice due to rearrangements in the proximal part of mouse chr7, locus order between human and bovine in the chromosomal segments containing genes FANCA, CDK10, SPG7, APRT, GALNS and SLC7A5 also was conserved but inverted (Goldammer *et al*., 2002). Complete synteny conservation is seen between four cattle chromosomes and their human orthologs (BTA12 and HSA13, BTA19 and HSA17, BTA24 and HSA 18, and BTAX and HSAX). The cattle human comparative map will help in understanding segment boundaries (Larkin *et al.*, 2003), centromere repositioning and distribution of genes in known breakpoint regions (Pevzner and Tesler, 2003). Although seven human chromosomes (1,3,4,5,10,11,X) have repositioned centromeres in the cattle genome located within large conserved syntenic blocks without any visible gene rearrangements, the centromere of human chromosome 22, repositioned in bovine chromosome17. Thus, of the 29 total cattle autosomes, 20 have either complete homology with human chromosome or homology with p- or q-arm or homology to genes of a single human chromosome (Wind *et al*., 2004).

1.5.2.4 Human and mouse (60/80MYA)

It is estimated that humans and mice shared a common ancestor approximately 80-100 MYA (Li and Gaur, 1991;Kumar and Hedges, 1998). Since most human and mouse coding regions (~1.5% of their genomes) have a high degree of sequence similarity and to some degree in the noncoding regions, sequence comparisons readily can reveal orthologous genes (Touchman *et al*., 2001). Although the noncoding regions usually show great divergence because they are not under selective constraints, approximately 1.5% of the human and mouse genome are conserved in noncoding regions.

Numerous studies have investigated synteny breakpoints between human and mouse and it has been observed that there is an enrichment of segmental duplications between human and mouse genomes (Armengol *et al*., 2003; Bailey *et al.*, 2004). Additional comparison of human and mouse sequences has identified several novel regulatory elements (Hardison *et al.*, 1997) as seen in the analysis of stem cell leukemia (SCL) loci (Gottgens *et al.*, 2001), cystic fibrosis transmembrane conductance regulator gene (CFTR) (Ellsworth *et al*., 2000), Bruton's tyrosine kinase (BTK) (Oeltjen *et al*., 1997) and the α-like and β-like globin genes (Flint *et al*., 2001). Comparative analysis of human (autoimmune regulator (autoimmune polyendocrinopathy candidiasis ectodermal dystrophy)) AIRE gene with its mouse ortholog gene (Aire) also showed high conservation of both gene structure and expression pattern (Blechschmidt *et al.*, 1999). It is well documented that the mouse is a good model for studying human diseases (Bedell *et al*., 1997) and in particular for understanding cancer such as human acute myeloma leukemia. As part of this dissertation research, sequencing and identifying genes involved in radiation induced leukemogenesis in mouse chromosome

2 region orthologous to human chr11p11-13 and human chr15q11-15 (both human chromosomes have putative tumor suppressor genes) was done in an effort to obtain a more detailed understanding of human radiation induced AML (Hayata *et al*., 1983; Rithidech *et al.*, 1997; Rithidech *et al*., 2002).

1.5.3 Human and Zebrafish (450MYA)

1.5.3.1 Zebrafish as the model system for whole mount *in situ* hybridization

The last common ancestor of present day humans and fish is estimated to have lived 400-450MYA (Kumar and Hedges, 1998). This makes fish one of the most distant non-mammalian vertebrates with an available sequence that can be compared to humans to identify conserved functional sequences. Several fish species have been sequenced, two pufferfish genomes, Fugu rubripes and Tetradon nigroviridis and genome of Oryzias latipes (medaka) (Ishikawa, 2000) essentially are completed and a working draft of the Danio rerio (zebrafish) genome (Grunwald, 2002; Vogel, 2000) is available. Comparison of human genome and the compact 365 Mb of Fugu genome has identified 1000 new unidentified human genes, and other conserved sequences that do not correspond to genes (Aparicio *et al*., 2002). The zebrafish is an ideal model system for both comparison with a distant species and also for gene expression studies since most human genes are conserved in fish (Dodd *et al*., 2000; Dooley *et al.*, 2000; Fishman, 2001; Postlethwait *et al.*, 2000). In 1981, George Streisinger introduced zebrafish as a model genetic organism (Streisinger *et al.*, 1981, 1986). This led Kimmel

(Kimmel *et al*., 1989, 1991a, 1991b,1995), Westerfield (Westerfield *et al*., 1990, 1992), Eisen (Eisen *et al*., 1991, 1993), and others to study zebrafish cell fate, lineage, and patterning during early embryonic development and nervous system development. The zebrafish genome is about 2900cM and 1.7x109 bp (Hinegardner *et al*., 1972) or about half the physical size of the human genome. The embryos are small (3-4cm long) and transparent, with one female capable of producing over 200 embryos per week. Embryonic development is rapid (5 days) and external, with the heart beating by the end of the first day and most organs, or at least their primordia, in place by five days after fertilization (Kimmel *et al*., 1995). The availability of the genetic and physical maps (Postlethwait *et al*., 1994; Pijnacker *et al*., 1995; Johnson *et al*., 1995; Knapik *et al*., 1996; Postlethwait *et al*., 1998, Geisler *et al*., 1999; Hukriede *et al*., 2001), cloning resources (Amemiya *et al*., 1999; Collins *et al*., 1995; Woods *et al*., 2000; Hukriede *et al*., 1999; Kelly *et al*. , 2000; Zhong *et al*., 1997), gene expression screens provide a background to understand developmentally regulated genes during zebrafish embryogenesis (Kudoh *et al*., 2001). Recent studies have shown that synteny between the zebrafish and human genome is conserved (Donovan *et al*., 2000) for some gene groups such as MHC class III region (Sultmann *et al*., 2000), but gene order along the syntenic chromosomes often differ and include inverted and/or transposed sequences (Barbazuk *et al*., 2000).

During the past 20 years, development biology, when supplemented with molecular and genetic information, has established that common genetic systems are conserved if they control the formation of organs during embryonic development of different animals. Thus, the genes that control early development in human embryos

can be studied utilizing the information obtained by the detailed analysis of the model genetic organisms such as zebrafish (Haffter *et al.*, 1996) and mouse (Neidhardt *et al*., 2000). The recent availability of 15,000 unique zebrafish EST and microarray data has provided information to control zebrafish embryogenesis by gene coordination in a genetic network and thus help identify many organogenesis specific genes (Lo, J *et al*., 2003) .

1.5.3.2 Global gene duplication in zebrafish(350-400MYA)

A gene duplication occurred during chordate evolution that plays an important role in the evolution of both genomes and organisms, leading to gene function diversification and biological diversity (Ohno 1970; Lynch and Conery, 2000). Zebrafish has two orthologs for many mammalian genes which were introduced by whole genome or chromosomal duplications that occurred in bony fish (including zebrafish and fugu) after their divergence from the tetrapod lineage that includes humans (Van de Peer *et al.*, 2002; Woods *et al*., 2000; Taylor *et al*., 2001).

Two models have been put forward to explain the survival of two or more genes in a genome, the classical model (Fisher, 1935; Haldane, 1933) and the duplication-degeneration-complementation (DDC) model (Force *et al*., 1999; Lynch and Force, 2000). The classical model predicts that one copy of the duplicate will accumulate null or deleterious mutation and usually degenerate to a pseudogene within a few million years and the other duplicate will retain the original function as gene loss is permissible in one duplicate as only one gene is required to maintain function similar to the ancestral single copy. However, this model fails to explain the duplicates present in

different genomes such as the human genome with at least 15% duplicated genes with 5.2% being segmental duplications (Li *et al*., 2001), 30-75 % of which are duplicated in the tetraploid fish lineages (Allendorf *et al.*, 1975; Ferris and Whitt, 1979). Half of these duplicated genes have been maintained for over 30 million years in Xenopus (Hughes and Hughes, 1993) and 20% of the duplicated genes in zebrafish lineage has been retained for over 110 million years (Postlethwait *et al.*, 2000).

In contrast, in the DDC model, the duplicated genes survive by evolving in two phases. In the first phase, one copy of the gene undergoes nonfunctionalization (complete loss of regulatory regions), while the other copy undergoes neofunctionalization, mainly by changes in regulatory regions or subfunctionalization, by complementary changes in regulatory regions. This model is similar to the classical model but differs in that during the second phase, the duplicate genes that acquire neofunctionalization or subfunctionalization are maintained by removal of redundant subfunction (Lynch *et al*., 2001). A few examples of duplicated genes in zebrafish maintained by DDC model are listed in table 1.2.

Hox clusters: single cluster in invetebrates, four clusters in tetrapod vertebrates (mouse ad xenopus), more than four clusters in teleosts	Amores *et al*., 1998; McGinnis and Krumlauf *et al*., 1992; Greer *et al*., 2000; Godsave *et al*., 1994; McClintock *et al*., 2002
Micropthalmia associated transcription factor (Mitf)	Altschmied *et al*., 2002
SOX9 (developmental regulator) One copy in mouse, two copy in zebrafish show subfunctionalization	Yan *et al*., 2002
Duplicated nodal class genes shown by mutants ndr1 (nodal related 1) and Cyclops (cyc)	Sampath *et al*., 1998
Engrailed (four copies)	Amores *et al*., 1998
Msx genes (duplicated)	Ekker *et al*., 1997
Fox1 (three copies)	Solomon *et al*., 2003

Two copies of cytochrome P450 aromatase (Cyp 19) differentially expressed in brain and ovary	Chiang *et al*., 2001
Two copies of JAK family of protein tyrosine kinase. Jak2a expressed in nervous system and developing buds and Jak2b expressed in developing lens and nephritic ducts	Oates *et al*., 1999
Duplicates of nkx genes	Lee *et al*., 1996

Table 1.2 Examples of genes following the DDC model for the maintenance of the extra copies of duplication

Thus, many zebrafish genes and other multigene families, such as the immunoglobulins and T cell receptors in vertebrates, come into existence or expand as a result of tandem duplication and inter chromosomal transpositions (Lewin 2000) and survive as a result of positive selection for both gain of function of each new family member, the classical model, and or partition of function among the members , the DDC model.

1.5.3.3 Mutation studies in zebrafish

In identifying genes involved in a wide variety of developmental processes, zebrafish has become an ideal model for molecular and genetic studies with the recent introduction of large scale mutagenesis screens (Amsterdam *et al*., 1999). Both forward genetic mutagenesis screens (Mullins *et al.*, 1994) and reverse genetic gene expression knock down by morpholino injection (Nasevicius and Ekker, 2000) have been done in the zebrafish. The two main strategies for random saturation mutagenesis are chemical mutagenesis with ethyl-nitrosurea (ENU) that has been to generate over 7000 point mutagenesis (Grunwald and Streisinger, 1992; Mullins *et al*., 1994; Solnica-Krezel *et*

al., 1994) and insertional mutagenesis (Kidwell 1986; Gridley *et al*., 1987) mediated by retroviral insertion (Amsterdam *et al*., 1999; Golling *et al*., 2002; Schier *et al*., 1996; Amsterdam *et al*., 1999; Postlethwait and Talbot, 1997). Some zebrafish organs are morphologically and functionally similar to human; it has been proposed that large scale mutagenic screens could help in understanding the mechanisms and pathways directly relevant to human disease and therapy (Stainier and Fishman, 1994; Stainier *et al.*, 1996; Pack *et al*., 1996; Ransom *et al*., 1996; Weinstein *et al*., 1996; Neuhauss, *et al*., 1996; Whitfield *et al*., 1996; Bennett *et al.*, 2001). The efficiency of the mutant is determined by resemblance of the mutant phenotype to mimic the particular disease. Many human diseases with gene homologs in zebrafish include Huntington's disease gene homolog (Karlovich *et al*., 1998); steroid imbalance associated disease (Lai *et al*.,1998) and Alzheimer's disease (Leimer *et al*., 1999); Hotl-Oram syndrome caused by mutation of the Tbx5 gene (Garrity *et al*., 2002); the *von gogh* mutant formed by mutation in tbx1 that causes defects in ear, pharyngeal arches, absence of thymus, fusion, and loss of neural crest derived pharyngeal cartilages; and reduction in endodermal pouches and aortic arches, a phenotype resembling DiGeorge syndrome (Piotrowski *et al.*, 2003).

2. Materials and Methods

2.1 Sequencing of the Chimp, Baboon, Cow, Mouse and Zebrafish Clones Syntenic to Human Chromosome 22 Cat Eye and DiGeorge Region

BAC clones syntenic to human chromosome 22 sequenced during the course of this dissertation research were from the rp43 (Donald), ch251 (Clint) and ptb (Gon) chimpanzee BAC libraries, rp41 baboon library, rp42 bovine library, ct7 and rp23 mouse libraries and the ch211 zebrafish BAC library. Chimpanzee BACs were selected by locating the BAC end sequences on human chromosome 22 using BLAST (Altschul et al, 1990, 1997) and later confirmed by comparative genomic maps constructed by an alignment between BAC end sequences of chimpanzee with the human genomic sequences (Fujiyama et al, 2002). Clones for mouse were acquired from Jim Lund (mouse chromosome 16) and Kanokporn Rithidech (mouse chromosome 2); baboon and bovine were acquired from Chris Amemiya and Harris Lewin, respectively; and zebrafish clones were chosen from the chori 211 library using gene specific probe

screening of primary and secondary pools of the clones. Ten primary pools of the zebrafish BAC library (76x384 = 29184/10 = 2918 BAC's in each of the 10 pools) were screened by PCR using gene specific primers for the presence of genes of interest. When a pool showed a PCR product, the entire secondary plate was screened via PCR for that particular BAC clone. When a primer pair indicated more than one of the primary pools, due to the presence of 2 to 3 copies of the genes, after the identification of the plate from the positive primary pools, the secondary pool was rescreened by PCR to determine the exact column and row of the individual clone carrying the gene of interest.

2.1.2 Random Shotgun Sequencing

The selected BAC clones were sequenced using the shotgun method (Chissoe *et al*., 1991; http://www.genome.ou.edu; Roe *et al*., 1995, 1996). The target BAC DNA was isolated using a modified cleared lysate procedure (Birnboim and Doly, 1979)) (http://www.genome.ou.edu/DblacetateProcV3D.html) from one liter or 200 ml cultures of each clone harboring the BAC of interest after late log growth in LB medium with appropriate antibiotic at 37°C. A portion of this large scale DNA then was physically sheared by nebulization at 8 psi or hydroshearing into 1.5 kb-4 kb fragments. Shearing generates DNA fragments with single-stranded end, that then are converted to blunt ends using Klenow DNA polymerase and phosphorylated at their 3' blunt ends using T4 polynucleotide kinase at 37°C in a water bath for 30 minutes. Two to four kb DNA fragments then were size selected by electrophoresis on 0.8% low

melting point agarose gel, and extracted from the gel using TE-saturated phenol, TE-saturated phenol-chloroform, chloroform, and finally a water-saturated ether extraction. After ethanol precipitation, the pellet was dried and dissolved in ddH_20. These size selected DNA fragments then were ligated into PUC18/Sma1 vectors (blunt ended and dephosphorylated at 5'ends) using T4 DNA ligase and subsequently, electroporated into E.coli XL1 Blue MRF' electrocompetent cells using a Bio Rad pulser. This transformation mixture containing X-gal (24mg/ml in DMF) for blue and white colony selection was plated onto 24.3 x 24.3 cm petri dishes and incubated at 37°C overnight. The white colonies were picked into 384 microtiter plates using Q-Pix (Genetix) colony picker and incubated in a HiGro oxygenated shaker incubator (Gene Machines Inc.) for 22 hrs at 37°C at 520rpm. The template for sequencing (subclone DNA) then was isolated from individual colony growths using an automated alkaline lysis protocol (modified Birnboim and Doly, 1979) on Zymark (Twister II and Sciclone subunit) and V-prep robots. The cycle sequencing reaction was pipetted on the V-prep robot by adding the isolated subcloned DNA, forward or reverse universal primer, and ABI BigDye or Amersham ET terminator mix (contains fluorescent-labeled ddNTP terminators, dNTPs, Taq DNA polymerase and $MgCl_2$). This was followed by incubation for 60 cycles consisting of 95°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes in a Perkin Elmer Gene Amp PCR system 9600 thermocycler. The resulting nested fragment sets were purified by 95 % ethanol/0.12N sodium acetate, pH 4.5, precipitation that removed the unincorporated dye terminators. Following drying, the DNA pellet was resuspended in 20-50 µl of sterile ddH_2O and electrophoretically loaded onto an ABI 3700 capillary sequencer capillary containing POP-5^{TM} polymer

and electrophoresed for 3 hrs at 6.5 kV. Sequence reads then were transferred to a local SUN workstation where the base sequences were determined by Phred (Ewing *et al*., 1998) followed by sequence reads assembly into large contiguous regions (contigs) with the Phrap assembly program (P. Green, pers. comm.). The consensus sequence from the assembly is created using a program Consed (Gordon *et al*., 1998). An average of 6-8 fold coverage of the shotgun sequence data, with typical read length of from 4-500 bases, from an initial random assembly generated contigs that usually covered 95-99.9% of the cloned sequence (http://www.genome.ou.edu/poisson_calc.html) (Lander and Waterman, 1988). Directed sequencing either with custom synthesized primers off the shotgun sub-clone or by PCR produced template from the BAC clones were sequenced to close either captured or uncaptured gaps. The gaps occurred because a region might have a lower coverage or no coverage at all by shotgun reads, by the presence of low quality or shorter than average reads on the ends of the contigs, or by secondary structures due to high GC content or long repeat regions. Often different nucleotides, such as dGTP or dITP for GC rich gaps, or dRhodamine dye terminators for AT rich gaps, were used to close gaps in especially difficult regions. However, when gaps could not be closed by sequencing from subclones or PCR amplification from the original BAC, the desired region of the target BAC often would be amplified using a PCR mix in which the dGTP was replaced by 7-deazadGTP (GC rich gaps) (Deirick *et al*., 1993; Mizusawa *et al*., 1986). Under these conditions, the resulting template was more easily denatured during the cycle sequencing reaction, since 7-deazadGTP only forms 2-H bonds with dCTP. In addition, 5-10% dimethyl sulfoxide (DMSO) (Winship, 1989), 1M betaine (Henke et al, 1997), or formamaide (Zhang et al,

1991) could be added to the cycle sequencing reaction to improve denaturation of secondary structures and prevent formation of primer dimers. This approach improved sequence quality and read length, using custom synthetic primers and either the Amersham E.T mix, the ABI BigDye mix, the dGTP mix or the dRhodamine mix. In regions where BAC end reads were not found, primers were made specific to the BAC vectors and PCR or MPCR amplified with different sets of primers designed at the ends of the contigs to correct any mis-joining contig-end regions. Typically, a combination of several of the above approaches were needed to close all gaps and reduce the error rate of a sequenced BAC insert to less than one uncertain base per 10,000 bases.

2.1.3 Analysis of the Sequences using Different Tools

Analysis of the resulting DNA sequences was required to locate the biologically relevant information embedded in the sequence. To define the basis of molecular evolution and phylogenetic relationships between organisms, alignment of the sequences were revealed differences and similarities between the sequences in both coding and noncoding regions. For shorter regions with high sequence conservation, Align (Myers and Miller, 1989) or Lalign (Huang and Miller, 1991) (if two sequences are compared) and Clustalw (Higgins *et al*., 1994) or Clustalx (Jeanmougin *et al*., 1998), or T-coffee (Notredame *et al*., 2000) (if more than two sequences were compared) were used. The clustal programs pairwise align the sequences by calculating the best match between sequences such that differences, similarities, and identities become evident, thus defining biological meaning to alignments (Eddy, 1995). For

larger sequences, both local alignment programs that determine high quality alignments between sequences irrespective of their order or orientation and global alignment programs that produce a single optimized comparison on the entire length of the sequences were used. Here two local alignment programs, Pipmaker (two genomic sequences) (Elinitski *et al*., 2002; Schwartz *et al*., 2000) and Multipipmaker (multiple genomic sequences) (Schwartz *et al*., 2003), were used. These programs are based on performing a blastz local alignment and generating a percentage identity plot between the aligned sequences according to their position with reference to one of the sequences that is chosen as the base sequence. In these alignments, only matches 50% and above are shown, and an ungapped alignment is plotted. The global alignment program Vista (Visualization Tool for Alignment) (Mayor et al, 2000) uses the Avid (a global alignment program) to determine the percent identity between one or more sequences using a fixed sliding window. The resulting plot shows the alignments greater than 10% in relation to the position within the sequences. In contrast to multipipmaker, vista shows not only alignments of sequences but also helps identify conserved elements. zPicture**,** another local alignment and visualizing tool based on blastz, can identify more obscure evolutionarily conserved regions such as transcription factor binding sites and other regulatory elements (Ocharenko *et al*., 2004). Repeatmasker compares the input sequence with a database of repetitive elements and can identify these repeats in a DNA sequence (Smith and Waterman, 1981). eShadow can compare both nucleotide and amino acid sequences of closely related species and is useful in predicting conserved functional elements (Ocharenko *et al*, 2004). REPuter can analyze repeats on a genomic level and identify both unique sequences and low copy repeat regions (Kurtz

et al., 2001). The prediction of conserved motifs was done using both ProfileScan (http://www.isrec.isb-sib.ch/software/PFSCAN_form.html) and Motif (http://motif.genome.ad.jp), software that compares amino acid sequence to the protein motifs characterizing protein families and domains in the Prosite and Pfam (Bateman et al, 2002) databases respectively (Gribskov et al, 1987; Bucher et al, 1996). Potential transcription factor binding sites were determined using TFsearch (www.cbrc.jp/research/db/TFsearch .html) and Motif, programs that compare an input sequence with a collection of vertebrate transcritption binding sites. Toucan, another program that can identify regulatory regions, such as known transcription binding regions and predicted binding site motifs, also was useful in visualizing repeated sequences (Aerts *et al*., 2003).

2.2 Whole Mount In Situ *Studies using Zebrafish as the Model Organism*

2.2.1 Isolation of Total RNA

Isolation of total RNA from zebrafish embryos at different developmental stages was done using the Trizol method (Chomczynski and Sacchi, 1987). In this method, the embryos were dechorinated using pronase (30 mg/ml). 1ml of trizol (phenol in saturated buffer 38%, guanidine thiocyanate 0.8 M, ammonium thiocyanate 0.4 M, sodium acetate 0.1 M (pH=5), glycerol) was added to a batch of 100 embryos and vortexed, vigorously to homogenize the embryos. The light red colored supernatant was collected after centrifugation for 10 min at 14000 rpm. The lower dark red layer again was treated with trizol, vortexed, and centrifuged, and the upper lighter red color supernatant solution again was collected. This solution was divided into tubes

containing 1.3 ml each. 0.2 ml of choloroform was added to each tube, followed by centrifugation for 20 min. Then, the upper liquid layer was transferred into a new tube and equal volume of isoproponol was added, centrifuged for 10 min in the cold room. The resulting pellet was washed with 70% ethanol. After a quick spin the ethanol was discarded and the pellet was air dried for 10 min, dissolved in 100 µl of autoclaved DEPC ddH_2O by vortexing, and incubated for 5 min at 60°C. 10 µl of the isolated RNA was checked on a formaldehyde gel and stored at -80°C.

2.2.2 RT-PCR and RNA Probe Preparation by *In Vitro* Transcription

Single-stranded cDNA was synthesized from the total RNA pool isolated from a combined embryos collected at 61, 72 and 77 hours post-fertilization (hpf). A 25 µl RT-PCR reaction contained 5x (avian myeloblastosis virus) (AMV) buffer, 25 mM $MgCl_2$,10 mM dNTP's, individual primer pairs (200 pmoles), 1.5 units of Taq polymerase, 20 units of AMV-RT (Roche) and 2 µg of total embryonic RNA. The first cycle of PCR was incubated at 48°C for 45 min during which the RNA was reverse transcribed by reverse transcriptase AMV-RT into first single stranded cDNA (Sambrook *et al*., 1989) at a site that was determined by the primer (See Appendix table 5 for a listing of PCR primers designed corresponding to the 5'UTR and 3'UTR of the mRNA of interest). This immediately was followed by a denaturation step at 94°C for 2 min for complete inactivation of the transcriptase and a second round of regular PCR of 40 cycles (94°C for 30 sec, 60°C for 45 sec, 68°C for 45 sec and 68°C for 10 min, 10°C for 24 min followed by 4°C) for the second strand cDNA synthesis. Since the first PCR

product sometimes contained multiple length DNAs, they were size fractionated on a 1% agarose gel. The band corresponding to the expected size PCR product was excised, the DNA was eluted (freeze and squeeze method described earlier), and served as the template for a second round of PCR using a pair of nested primers. Typically, this treatment resulted in a single product that then was purified by treating it with shrimp alkaline phosphatase and exonuclease I, followed by a phenol/chloroform extraction, ethanol/acetate precipitation, and a 70% ethanol wash. The pelleted DNA then was dried and resuspended in 30 ml of sterile distilled deionized water. After end repair, the PCR product was cloned into SmaI site of pUC18 and sequence verified. The cloned product then was excised from the pUC18 vector by EcoRI and BamH1 restriction enzyme digestion. It then was sub-cloned into the pBlueScript SK^- at the EcoRI and BamH1 sites. Again, the insert was sequenced verified. Plasmid DNA was isolated by miniprep double-stranded DNA isolation (Roe *et al*.,1995) and purified by passage through a Zymo-spin column (DNA clean and concentrator kit). A single digestion with BamH1 was carried out to linearize the DNA and, after an preparative agarose gel and elution, the insert was purified using the Glass Milk gene clean procedure. After confirming the DNA concentration by agarose gel electrophoresis, a 20 µl of the *in vitro* transcription reaction containing linearized plasmid DNA, T7 RNA polymerase, 10 x transcription buffer, 10 x nucleotide mix with digoxigenin-labeled UTP (Roche), and RNase inhibitor (40 U/µl, cat#300152-51, Stratagene) was incubated at 37°C for 2 hrs followed by RNAase free DNase treatment for 25 min. The RNA probe then was precipitated with 0.5 M EDTA, 5 M LiCl (pH 5.0) and 2 volumes of cold ethanol and the pellet was washed with 70% ethanol. The dried pellet was

dissolved in diethylpyrocarbonate (DEPC) treated H_2O containing 1 µl of RNase inhibitor. The RNA probe was electrophoresed on a freshly prepared 1% agarose gel to confirm the correct size of the DIG labeled RNA probe.

2.2.3 Single Stranded Exon Specific DNA Probes and Primer Design

The expression pattern of the known and unknown human chromosome 22 genes by systematically identifying the expression of zebrafish orthologs in zebrafish embryos. A reciprocal BLAST comparison of the predicted human chromosome 22 genes to the recently available compilation of zebrafish predicted genes from the zebrafish whole genome shotgun assembly gave the sequence of the zebrafish orthologs of the human chromosome 22 genes (exons). Exon specific PCR primers were picked using PrimOU and then crossmatched to the entire zebrafish genome sequences to confirm that the primers were unique. These unique primers (20mers) then were used to amplify the gene of interest using the zebrafish genomic DNA as the PCR template. For this purpose zebrafish genomic DNA was extracted using the protocol described by A. Fritz in ZFIN. After removing the embryo from the chorion, 50 embryo's (7 days old) were transferred into 1.5 ml Eppendroff tubes and rinsed with water. After the water was completely removed, 1ml of extraction buffer (10 mM Tris pH 8.2, 10 mM EDTA, 200 mM NaCl, 0.5% SDS, 200 µg/ml proteinase K) then was added to 50 embryos and incubated overnight at 50°C. This resulted in completely resuspending the embryos in the extraction buffer. The embryonic genomic DNA was extracted by phenol/chloroform extraction followed by ethanol/salt precipitation. The resulting DNA

pellet was dried and dissolved in 50-100 µl of water and stored at -20°C after determining the concentration by measuring the absorbance at 260 nm. The concentration of the genomic DNA (1 A_{260}=50 ng/µl) obtained typically was between the range 100-1000 ng/µl. The PCR reactions were performed with 95°C hold for 5 min, followed by first 10 cycles, denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec and elongation at 72°C for 30 sec and then for 30 cycles, denaturation at 94°C for 1 min, annealing at 60°C for 2 min and elongation at 72°C for 1 min, for a total of 40 cycles. An alternative touch-down cycling PCR program also was used with the following cycling conditions: 94°C hold for 2 min followed by 30 cycles, denaturation at 94°C for 1 min, annealing at range of temperatures 65°C-50°C with an decrement of 2°C (65°C for 2 cycles, 63°C for 3 cycles, 61°C for 4 cycles, 60°C for 4 cycles, 58°C for 4 cycles, 56°C for 4 cycles, 54°C for 4 cycles, 52°C for 3 cycles and 50°C for 2 cycles), elongation at 72°C for 2 min followed by 10 cycles of 94°C for 1 min, 58°C for 1 min and 72°C for 2 min and an additional extension step at 72°C for 5 min (Don *et al*., 1991; Weinholds *et al*., 2003).

In cases where multiple bands were obtained, nested PCR was performed to enrich the specific band of interest using nested primer pairs inside the first set of primers. Amplified PCR products then were loaded onto a 3% low melt agarose gel and after electrophoresis, the specific sized DNA was extracted from the gel using the freeze and squeeze method, precipitated with ethanol/salt, washed with 70% ethanol, dried, and dissolved in ddH2O. The purified PCR product then was used as a template for unidirectional PCR (Fuhrmann et al, 1994, Vanholme *et al*., 2002; Knuchel *et al*., 2000; Kitazawa *et al*., 1999) with a single primer (sense or antisense) at time in the

presence of DIG-dUTP (PCR DIG labeling mix. Roche) to produce the single stranded DNA sense and anitsense probes (Patel and Goodman,1992; Seydoux and Fire, 1995).

The cycling conditions for unidirectional PCR were: hold for 5 min at 95°C, followed by 10 cycles denaturation at 94°C for 1 min, annealing at 58°C for 1 min and elongation at 72°C for 1 min and then 30 cycles denaturation at 94°C for 1 min, annealing at 60°C for 1 min and elongation at 72°C for 1min, for a total of 40 cycles. The DIG labeled single stranded DNA probe were precipitated by ethanol/salt, washed with 70% ethanol, and dissolved in hybridization buffer.

2.2.4 Whole Mount *In Situ* Hybridization

Zebrafish were raised and maintained under standard conditions of 27°C (Westerfield, 1994). Fish embryos from 12, 16, 24, 48 and 72 hpf developmental stages were collected and incubated in 28.5°C (Kimmel et al, 1995) in the presence of PTU (phenylthiourea-0.003%) to prevent melanization of the embryos and then, after dehydration, stored at -20°C in 100% methanol. The mutant embryos, HI1373 (Golling et al, 2002) and Ace mutant (Brand et al, 1996) similarly were isolated and noted for the morphological defects. Prior to hybridization, the dehydrated zebrafish embryos stored in 100% methanol were rehydrated by suspension in 1xPBST (Phosphate Buffered Saline Tween-20) immediately prior to a hybridization experiment. The embryos in 1x PBST were treated with proteinase K (10 mg/ml) by

incubation at 20°C for 1-20 min. Here the short time incubation was for early embryonic stages and longer incubation was for later stages, as previously determined by pilot studies that determined the incubation time needed for maximum permeability with minimal embryo disruption. The embryos then were treated with glycine (25 mg/ml) to inhibit proteinase k and after washing with PBST, were fixed in 4% paraformaldehyde (PFA), washed again with PBST and distributed into separate wells of a flat bottom 96-well microtiter plate. In some instances, heating the embryos in PBST before the prehybridization step reduced the background hybridization to achieve a clearer expression pattern. Pre-hybridization and hybridization was performed at different temperatures ranging from 50 to 70°C (depending on the probe length and binding specificity to reduce the background hybridization). The pre-hybridization was carried out for 1h in hybridization solution (50% formamide, 5X SSC, heparin (50 µg/ml), yeast RNA (500 µg/ml), 0.2% Tween-20, 1M citric acid pH=6) followed by hybridization in 50% hybridization solution with DIG labeled ssDNA probes for 16 h. A series of gradient washes involving high salt (2x SSC), high temperature (ranging from 48-70°C depending on the hybridization temperature used earlier during probe binding). A low salt (0.2x SSC), high temperature (48-70°C) highly stringent wash, followed by a low salt (0.2x SSC), room temperature wash, automated on the Robins Hydra 66 robot, was used to remove any unbound probe and reduce the background. After these washes, the embryos were treated with 2% blocking solution (BSA-100 mg, sheep serum-1 ml, 1X PBST) for 1 hour and then incubated overnight by shaking at 4°C with anti-DIG antibody (1:10,000) (anti-digoxigenin-AP, Fab fragments, Roche). Again after a series of washes with 1x PBST and NTMT (5 M NaCl , 1 M $MgCl_2$, 1 M

Tris pH=9.5, 20% Tween-20) on the Hydra, the color was enhanced by incubating at room temperature in a substrate color reaction mixture (NBT (nitroblue tetrazolium)/BCIP (5-bromo-4-chloro-3-indolyl phosphate). Once a stable color was maintained, the reaction was terminated by washing twice with 1x PBST. The zebrafish then were visualized and photographed on a Leica MZFLIII Stereofluorescence microscope with an attached 35mm camera and Leica DM4500B DigitalMicroscope with an attached high resolution digital CCD color camera (MicroPublisher 3.3 RTV) with high speed real time viewing.

3. Results and Discussion

The 25 chimpanzee, 9 baboon, 6 bovine, 2 mouse and 3 zebrafish BAC clones sequenced, and combined with additional syntenic clones sequenced by others, prior to the analysis and comparison to the human chromosome 22 CES-DGCR region genes, are listed in tables 3.1, 3.2, 3.3, 3.4 and 3.5 below, with the corresponding marker genes annotated for each clone and a pictorial representation of the chimpanzee chromosome 22 clones in the region syntenic to the CES-DGCR region of human chromosome 22 is shown in figure 3.1.

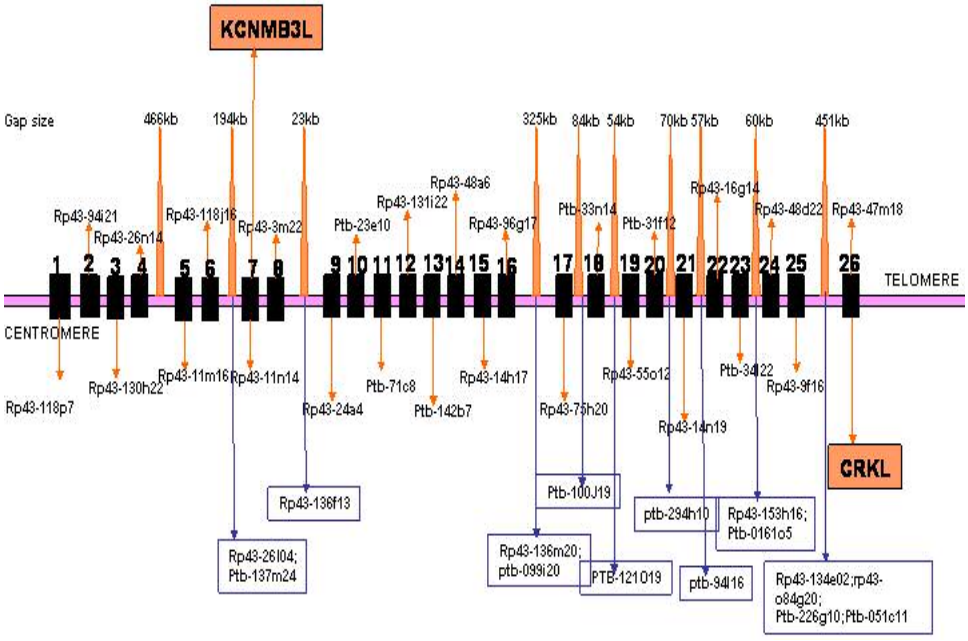


Figure 3.1 Chimpanzee BACs Studied. The black boxes represent the chimpanzee BACs syntenic to human chromosome 22 sequenced during this dissertation research. The cones represent the gaps between the sequenced projects, the blue outlined boxes represent the projects that require sequencing to close the gaps and the numbers above each black rectangle represent the marker genes listed in table 3.1.

	Chimpanzee clones	Marker genes encoded
1	rp43-118p7	TOMM40, L1H 3' homolog, FKSG30

2	rp43-94i21	Repeats similar to human CES genomic DNA
3	rp43-130h22	FKSG30, POTE2A, similar to Nek2 (NimA), similar to NF1
4	rp43-26n14	Similar to NF1, POTE2B, NimA
5	rp43-11m16	Similar to phosphoinositide 3-kinase catalytic gamma polypeptide, FKSG72
6	rp43-11n14	XK related protein 3, similar to immunoglobulin kappa light chain, KCNM3BL, Kiaa1629 Zinc finger protein 532
7	rp43-3m22	GRB2
8	rp43-24a4	IL17R, CECR5, CECR6,CECR1
9	ptb-23e10	CES region specific repeat
10	ptb-71c8	Cecr2, Trf2
11	rp43-131i22	CECR2, Kiaa1740, CECR2, Kiaa1740
12	ptb-142b7	MIL1, BCL2L13, ATP6E1, SLC25A18, CECR2
13	rp43-48a6	MIL1, BID, Kiaa0819, Kiaa1374, Mical-3
14	rp43-14h17	Similar to MICAL-3
15	rp43-96g17	PEX26, TUB8, USP18
16	rp43-75h20	SLC25A1,DGCR14, STK22A,DGCR2, POM121 membrane glycoprotein like 1,
17	ptb-33n14	HIRA
18	rp43-55o12	HIRA, UFD1F, CDC45, CLDN5
19	ptb-31f12	CLDN5, CDC45
20	rp43-14n19	SEPTIN5(CDCREL-1), TBX1, GNB1L, TXNRD2, TR
21	rp43-16g14	ZDHHC8,RANBP1, HpaII9c, T10,ARVCF, COMT,
22	ptb-34l22	T10, DGCR8, HpaII9c
23	rp43-48d22	PRODHL, DGCR6L, USP18, ZNF74, SCARF2, KELCH-like protein 12, Q96B68
24	rp43-9f16	ZNF74, SCARF2
25	rp43-47m18	LZTR-1, CRKL, SNAP-29

Table 3.1 The chimpanzee BAC clones from the rp43 and ptb libraries syntenic to human chromosome 22 sequenced and their associated marker genes**.**

	Baboon Clones	**Marker genes encoded**
1	rp41-88j2	Dgcr2, carbonic anhydrase 15, kiaa0649, HSPC009 protein, DnaJ protein, POM121 kiaa4041, PRODH, DGCR6
2.	rP41-5C11	USP18, XP_372891, RVT,dgcr6,prodh,kiaa4041,POM121 membrane glycoprotein like 1
3	rp41-177K13	XP_372891, USP18, tuba8, Pex26, IS10-RIGHT TRANSPOSASE (Tn10)
4.	rp41-367l20	CECR1, CECR5, CECR6, IL17R
5.	rp41-110d13	XP_372981, Putative 150(LIH3') Dgcr6, Prodh, NgR(Nogo66- reticulon 4 receptor precursor)
6.	rp41-30f24	PRODH, DGCR6, Putative 150(LIH3'),XP_515001(Pan), XP_372891,BCR protein, bcr kiaa3017, Znf74, scarf2
7.	rp41-136c24	CRKL, SNAP29, PIK4CA, SERPIND1
8.	rp41-53o7	PIK4CA, SERPIND1, PIK4CA(shorter), HIC2(hypermethylated in cancer 2 protein-related to gene on chromosome 22, similar to hypothetical protein XP_525533.1(Pan), hypothetical protein-DKFZp43H177.1 (Kiaa1666)
9.	rp41-2d18	Predicted similar to Na/H exchanger 3, Kiaa1418, PCQAP, Kelch like (kelch like variant)

Table 3.2 The baboon BAC clones from the rp41 library syntenic to human chromosome 22 sequenced and their associated marker genes.

	Bovine clones	**Marker genes encoded**

1	Rp42-381a23	ZDHHC8, HpaII9c, RANBP1, RTN4R, PRODH, DGCR8 partial
2	Rp42-382m10	BCL2L13, Kiaa0819, Kiaa1364, MICAL3, MIL1, BID CECR2, ATP6E
3	Rp42-158g13	Reverse transcriptase like protein,BCNT- bucentaur (an Alu-linked repetitive sequence corresponding to 280 aa is expressed in a novel bovine protein, but not its human homolog) , Pregnancy associated glycoprotein,
4	Rp42-400k8	E2L31, Kiaa1666, peptidylpropyl isomerase like 2 isoform, PPIL2 protein, Yipee like protein 1 (DiGeorge syndrome related protein FKSG3), reverse transcriptase like protein, MAPK2,
5	Rp42-155a20	ARVCF ,COMT, TXNDR2, BUCENTAUR, GNB1L, KIAA1645, TBX1, SEPTIN1/5 (CDCREL1)
6	Rp42-266g20	UFD1, NLVCF, HIRA, SLC25A1, DGCR14, STK22A, LDLa, DGCR2, ZNF74

Table 3.3 The bovine BAC clones from the rp42 library syntenic to human chromosome 22 sequenced and their associated marker genes.

	Mouse clones	**Marker genes encoded**
1.	Rp21-598k13	HIRA(histone cell cycle regulation defective homolog A) chromosome 22.
2.	Rp21-594l4	Glutathione-S-transferase, Degradation in ER protein 3(DER1 like protein3)

Table 3.4 The mouse BAC clones from the rp21 library syntenic to human chromosome 22 sequenced and their associated marker genes.

	Zebrafish clones	**Marker genes encoded**
1	Ch211-31g3	Zdhhc8(36-186bp-isoform2), XP_373945, HpaII9c Dgcr8(variant), Zdhhc8(377-709bp), Dgcr8(WW domain), T10 (syntenic to human chromosome 22)
2	Ch211-31g13	Polo like kinase1-2 (PLK2)(serine / threonine protein kinase), cAMP-specific 3',5'-cyclic phosphodiesterase 4D (PDE43) (syntenic to human chromosome 5), Pola2(polymerase alpha2)(syntenic to human chromosome 11),
3	Ch211-31g8	Trace amine associated receptor 9(TAAR9) (syntenic to human chromosome 6)

Table 3.5 The zebrafish BAC clones from the ch211 library syntenic to human chromosomes 22, 5, 11 and 6 sequenced and their associated marker genes.

As shown below in figure 3.2, the chimpanzee clones sequenced covered almost all genes syntenic to human chromosome 22 that were listed in table 3.1 above except for those in the 325 kb region between clones rp43-96g17 and rp43-75h20 that encode the genes DGCR6 and PRODH, the 451 kb region between clones rp43-48d22 and rp43-47m18 encoding the genes PCQAP, PIK4CA and SERPIND1, and the 466 kb region upstream of CES as also shown in figure 3.1.

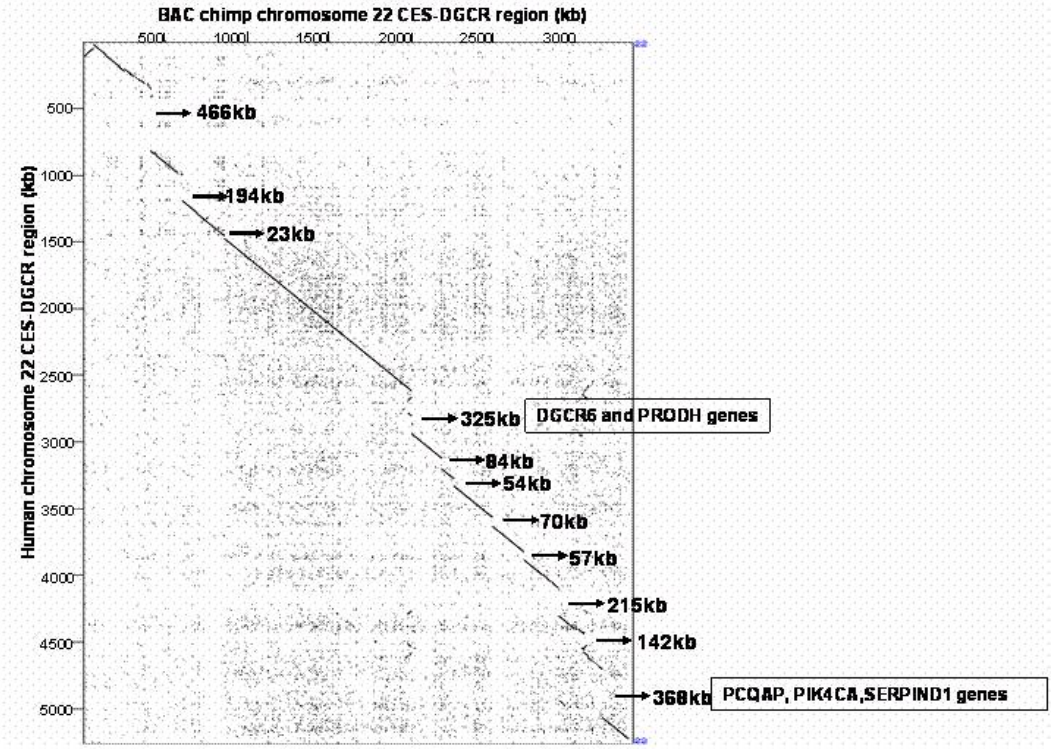


Figure 3.2 Comparison between the nucleotide sequences of the human chromosome 22 CES-DGCR region and syntenic region in chimpanze. The dot plot shows the comparison and match between the nucleotide sequences of the human chromosome 22 CES-DGCR region (Y axis) and the sequenced syntenic region in chimpanzee. Each dot represents match of 50 bases and continuous dots form lines representing significant matches. The numbers in between the lines represent the gap sizes and the genes listed are those missing because of the sequence gap.

Although many of the marker genes present in chimpanzee, baboon, bovine, mouse and zebrafish have conserved synteny when compared to human chromosome 22, each species encodes additional genes and unique repeats such as, for example, the bovine LDLa, pregnancy associated glycoprotein and a unique Alu bucentaur repeat absent in humans. The PIP comparison shown in figures 3.3(a) and 3.4(b) for a 389kb representative region encoding genes TBX1 to ZDHHC8 from 20 different organisms; human, chimpanzee, macaca, mouse, rat, dog, cow, opposum, chicken, frog, zebrafish, fugu, tetradon, ciona, fruitfly, mosquito, honey bee, C. elegans, and yeast clearly shows conservation in the coding region for all species compared. As can be seen in this

figure, the DNA sequence from two chimpanzee BAC libraries (sometimes three BAC libraries, see Methods) also were included in this comparison when it was available.

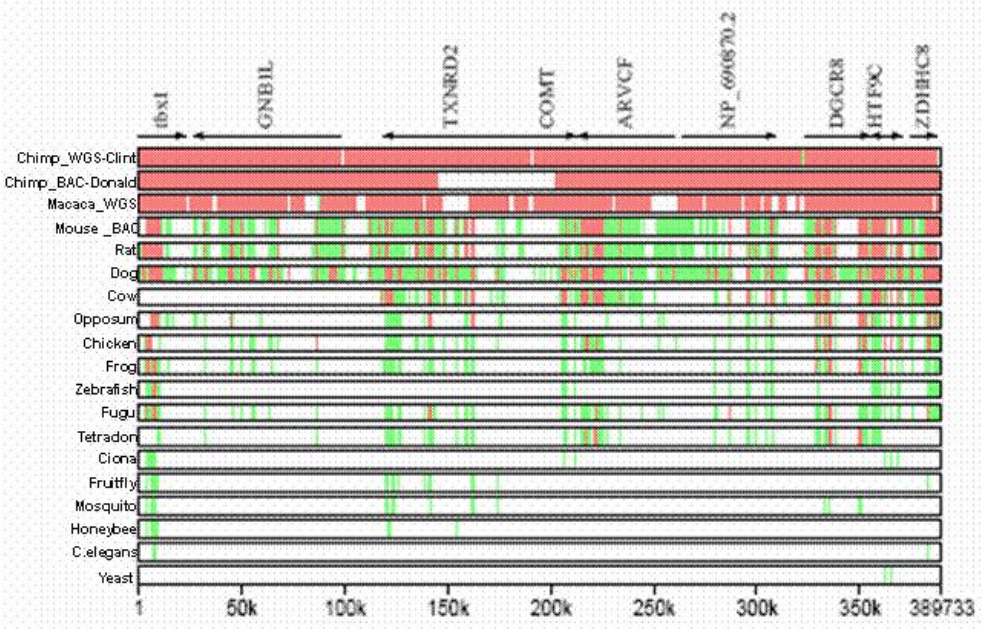


Figure 3.3 Overview of PIP. This figure gives an overview of the PIP in figure 3.4 below. The red color indicates >=70% strong alignment for 100bp without gap and green color indicates lower percentage alignment. A trend in the coloring pattern showing the high degree of alignment at the top for closely related species (red) to decreasing degree of alignment (green) for distant species in the evolutionary time scale**.**

Large conserved areas of the noncoding regions are observed in all mammals with a greater degree of conservation seen in primates. The zebrafish homology with human chromosome 22 is shorter when compared to the mammals and mainly shows a high degree of identity in the exon coding regions. However, it is interesting to note that the gene order over this large region surprisingly is conserved over the species compared considering the large evolutionary distance between common ancestors.

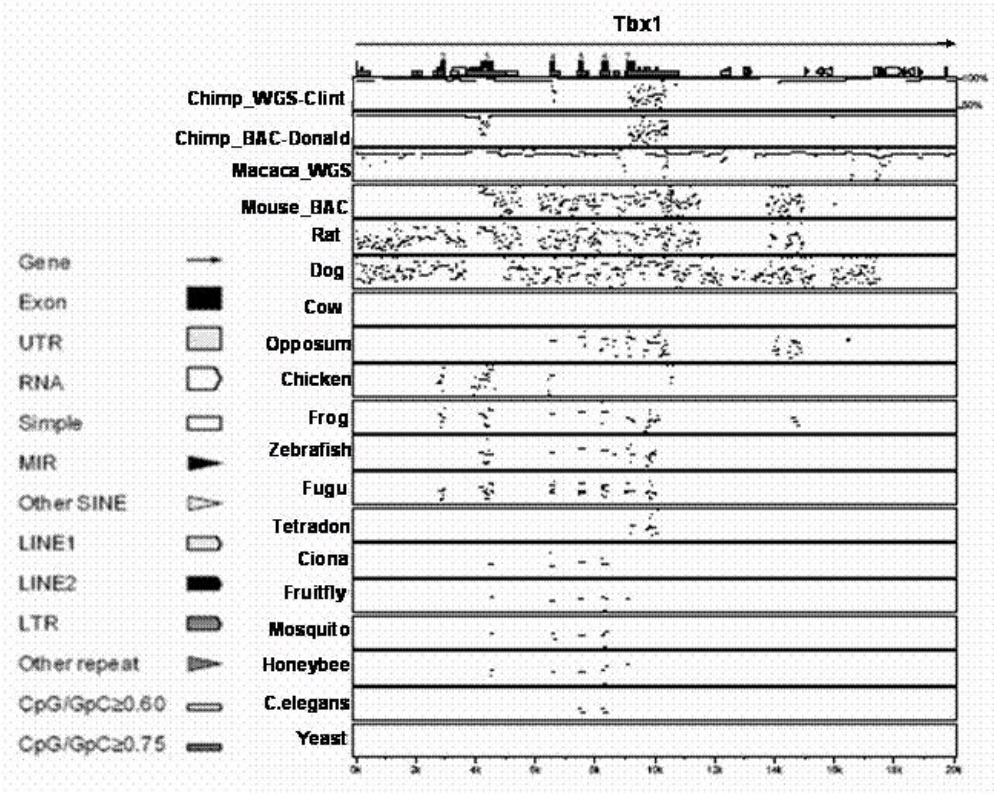


Figure 3.4 The PIP output highlighting conservation. Clear conservation at the coding level between all species compared (cow and yeast missing orthologs for this gene) and conservation at the level of noncoding regions at a greater degree in primates than to other mammals with respect to humans.

A VISTA plot figure of the human, whole genome shotgun chimp-Clint, BAC-based chimp-Donald, whole genome shotgun macaca and BAC-based mouse sequences over the same 389kb region is shown in figure 3.5.

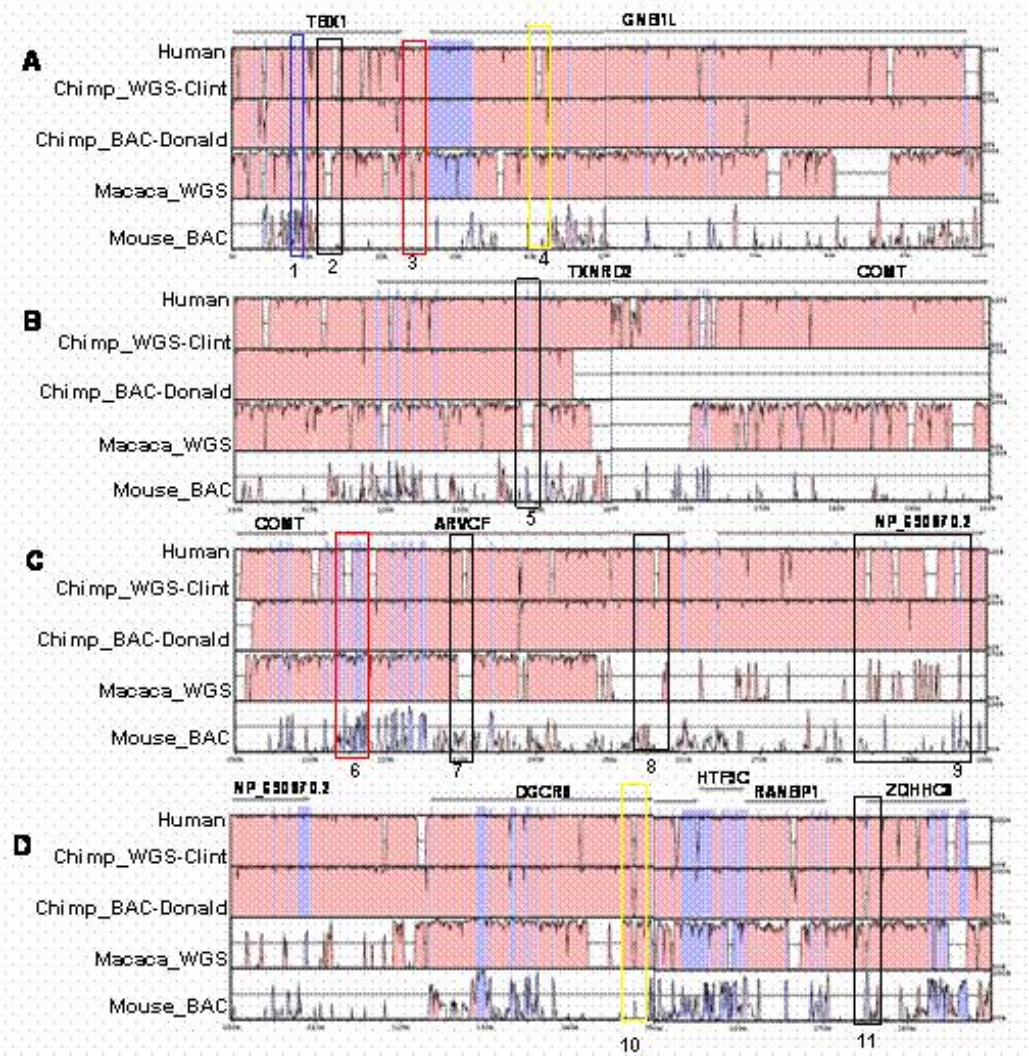


Figure 3.5 VISTA output highlighting comparison between whole genome shotgun sequence and BAC based sequence. The (A) section of VISTA shows three highlighted regions namely (1) which shows a missing exon in macaca while present in human, chimp-Clint, chimp-Donald and mouse and (2) missing noncoding region in chimp-Clint and macaca but present in human and chimp-Donald. (3) shows missing noncoding region in macaca and (4) shows missing noncoding region in chimp-Clint. Section (B) shows missing exon in macaca(5); the missing region in chimp-Donald in this section is due to gap in sequence. Section C shows in region (6) missing exons in chimp-Clint, (7) shows missing noncoding region in Clint and macaca, (8) shows missing noncoding region in Clint and (9) shows missing noncoding and exon in Clint. Section D shows in region (10) missing noncoding region in Clint, Donald and macaca but present in human and mouse. Region (11) shows missing exon in chimp-Donald and macaca but present in human, chimp-Clint and mouse.

Here it can be seen that there is unique conservation of both coding and noncoding regions between the mammalian species compared, although regions are missing in whole genome shotgun data sequence that are present in the BAC based sequences, especially in the noncoding regions. This observation points to one of the major shortcomings of the whole genome shotgun approach when

compared to a BAC by BAC ordered sequencing strategy as the whole genome shotgun approach seems to be lacking sequences that could encode important information further defining the differences seen between human and their close primate relatives that could be significant both at the phenotypic level and in the differences in their susceptibility to diseases.

3.1 Comparative Analysis of Human Chromosome 22 CES-DGCR Region

A comparison between humans and other organisms, in particular chimpanzee and baboon, revealed differences in their repeat elements (both type and number) and in their base composition because of base substitutions that often resulted in differing codon usage.

3.1.1 Repetitive Elements

Previous studies have shown that syntenic regions in species other than chimpanzee and baboon have lesser and shorter repeats than human (Aparcio *et al*., 2002; Mural *et al*., 2002) and that the increase in the repeats in humans mainly is due to the reduced deletion rate efficiency (Dehal *et al.,* 2001). These repeat elements also have been suggested to facilitate the rate at which mutations can take place (Yoshio Miki,1998).

Comparison of the gene repeats (as shown in Appendix table 1 and figure 3.6) between human, chimpanzee and baboon, reveals that human and baboon has a significant increase in the number of retroelement insertions (Liu *et al*., 2003).

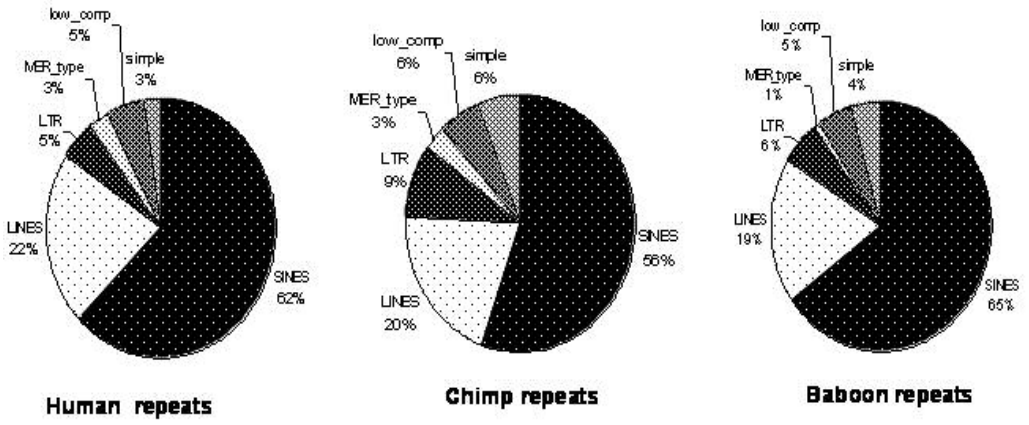


Figure 3.6 Major repeats. Percentage distribution of the major repeats noticed in human, chimpanzee and baboon.

This specially is the case with Alu (SINE) repeats, since chimpanzee has fewer Alu sequences (56%) when compared to both humans (62%) and baboons (65%). It also is evident that the major group of repeats found in humans, chimpanzees, and baboons belong to the Alu, L1 and LTR repeat families. In humans, a majority of the Alu insertions belonged to the young AluY (originating less than 15Myrs ago) subfamilies of recent origin (Dagan *et al*., 2004; Roy *et al*. ,1999), while others belong to the intermediate AluS family (20Myrs), predominately the AluSx subfamily, followed by AluSg, AluSq and AluSc subfamilies. For example, AluYa5 subfamilies are observed in the intron2 of the gene ATP6V1E1 and intron9 of the gene Kiaa0819, 297 and 309bp, respectively. Another human specific Alu repeat is the 300-317 bp AluYb9 repeat seen in the introns of genes SLC25A18, BID, MICAL3 and DGCR8 (Batzer *et al*., 1990, 1991, 1995; Caroll *et al*., 2001; Roy *et al*., 1999; Roy *et al*., 2001; Deninger *et al*., 1999; Matera *et al*., 1990; Jurka *et al*.,1993).

In chimpanzee, the Alu repeat families differed from those of humans (Hedges *et al*., 2004), and commonly belonged to oldest families of Alu repeats, the AluJo and

AluJb that entered the genome 65-40Myrs ago. Chimpanzee Alu repeat families also had unique insertions of recent new AluY subfamilies, AluYg and AluYc3, as observed in intron1 of the BCL2L13 gene, AluYc5 in the intron1 of the BID gene and AluYh9 in intron2 of the DGCR2 gene. In baboon, the repeats belonged to the intermediate family of AluS repeats (AluSx and AluSg). However, the major class of repeats were of the AluY family of repeats, with unique insertions of recent Alu subfamilies such as AluYa5 as observed in the intron 11 of the CECR1 gene and AluYg in intron1 of the Pex26 gene. The unique subfamilies of AluY repeats seen in human, chimpanzee and baboon are summarized in figure 3.7.

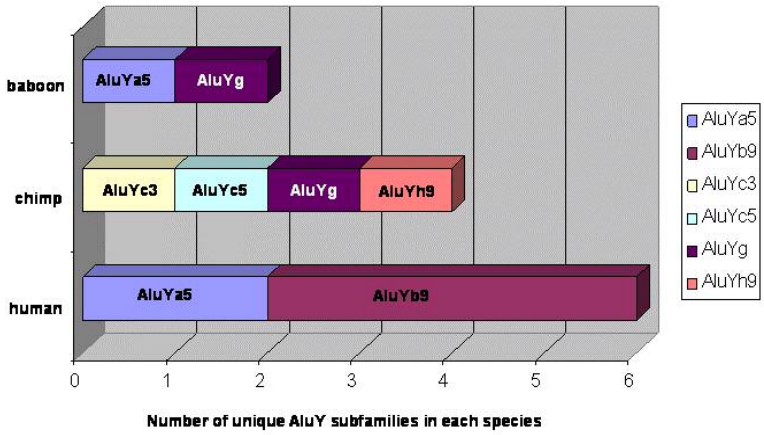


Figure 3.7 Unique AluY repeats. Major classes of unique young AluY family present in human, chimpanzee and baboon.

Other families of "old" Alu sequences, included members of the FLAMC/A and MIR families, that are present in significant numbers in both humans and baboons but is present in lower amounts in chimps. The Alu families present in baboons showed similarity both to humans and chimpanzees, but greatly exceeded in number than those present in humans (Liu *et al*., 2003).

In addition to Alu repeats, other conserved repeats also were observed. In humans, LINEs in the category L2, L1 (active class of LINEs) and L3 category with lesser numbers of L1M, L1PA and HAL1 elements. A similar trend also was noticed in chimpanzees, but in baboons the LINE repeats show major increases in the L1M, HAL1 and L2 elements over those present in humans. Overall, baboons show greater increase in LINE elements similar to those seen with Alu repeats. The number of LINEs present in baboons, humans, and chimpanzees almost correspond to the number of Alus. Earlier studies showed that the correlation between Alus and LINEs may be because Alus are dependant on the reverse transcriptase and endonucleolytic activity of the LINE retroelements, rendering their distribution very similar (Jurka *et al*., 2004). In the category of LTR repeats, the major elements belong to MaLR (LTR, MLT1H) and ERV1 (MER) classes in all three species and are greater in number in baboons and less in humans and chimpanzees (Smit *et al*., 1993; 1999). Other repeats such as the MER1 and MER2 types, occur more often in humans than chimpanzees and baboons. In the category of low complexity repeats, the GC rich repeats are seen more often in humans and chimpanzees while AT rich repeats are seen more often in baboons. Other simple repeats also are significantly increased in baboons and are decreased in both humans and chimpanzees. The major repeats in the bovine genome are of the bucentaur type, a bovine Alu–driven family that encodes a novel 280 amino acids protein in the Alu linked repetitive sequence region (Nobukuni *et al.,* 1997). In the few clones sequenced in mouse and zebrafish in this dissertation, the major repeats seen in mouse mainly belonged to the LINE family both in the exons and introns in comparison to SINEs (mainly belonging to the old Alu subfamilies) and other low complexity and simple

repeats. The major repeats seen in zebrafish belonged to the low complexity and simple repeats followed by SINEs (old Alu subfamilies), LINEs (L2), and few belonging to the MER1, MER2 and also LTR/ERV1 repeats.

3.1.2 Locations of Repeat Element Insertions and Their Significance

Earlier studies have shown that Alu elements play an important role in differential gene regulation by alternate splicing. When present in the 5'UTR and 3'UTR Alu elements can play the role of an enhancer or silencer. Intronic Alus also can inactivate or change the function of a gene product by either creating an alternate splice site or interfering the splicing mechanism. Thus, they can alter the pre-mRNAs processing (for example, exon skipping) (Ricci *et al.,* 2002; Ganguly *et al*., 2003), leading to altered gene products, when present within introns, 5'UTRs, and 3'UTRs (Mitchell *et al*., 1991; Knebelmann *et al.,* 1995; Oh *et al.,* 2001; Groover *et al*., 2003; Hayakawa *et al*., 2001).

In humans, chimpanzees, and baboons, the insertions of most repeats occur in all portions of a protein coding gene, i.e. in introns, 5'UTR, 3'UTR, and exons (Chen *et al*., 2002; Minghetti *et al*., 1993), as listed in the Table 3.6. Low complexity GC rich repeats, Alu, and LINE repeats are observed in these regions.

Locations of repeat insertions	Genes of human chromosome 22 and homologous

	chimpanzee and baboon genes)
Intronic	GRB2, IL17R CECR5, CECR1, CECR2, SLC25A18, ATP6V1E1, DGCR6, PRODH, Q9BYB2, DGCR2, DGCR14, HIRA,CDC45L, TBX1, GN1BL,USP18, TUBA8, PEX26, MICAL3, KIAA0819, Q8N1L1, BID, BCL2L13, TXNRD2, COMT, Q8IZ69, RANBP1, ZDHCC8, PIK4CA, ZNF74, SCARF2, PCQAP, SNAP29, CRKL
5'UTR	CLDN5, STK22B, DGCR6, RANBP1, Q8IZ69, ARVCF, CECR1
3'UTR	CECR1, BCL2L13, PEX26, Q8IW05/T10, RANBP1, CRKL, GNB1L
Exonic	HIRA, SLC25A1, GSCL, Q9BYB2, PRODH (in baboons), PCQAP and SCARF2 (repeat absent in chimpanzee and baboon exon), ZNF74, DGCR6L, ZDHCC8, TXNRD2(repeat absent in chimpanzee exon).

Table 3.6 List of genes seen in human, chimpanzee and baboon that contain repeat elements in their intronic, 5'UTR, 3'UTR and exonic regions**.**

Alu elements also play an important role in genome diversity by facilitating homologous recombination events, resulting in duplication, deletion, or translocation (Dagan *et al*., 2004). The presence of Alu sequences as well as other retrotransposable elements (B1, B2, MIR, LINEs) in pre-mRNA can influence and affect the polyadenylation of transcripts and translation efficiency, as well as the expression patterns (Rowold *et al*., 2000). These changes often result in serious genetic disorders (Wallace *et al*., 1991, Knebelmann *et al*., 1995). As a result, understanding more about the nature of the presence of these repeat elements and their location is very important to understand the evolutionary relationship between repeat elements and genetic disorders. The chimpanzee has two copies of the SNAP29 gene, both of which contain

2 exons. As seen in figure 3.8, copy 1(chimptrans1-SNAP29) encodes exons 1 and 2 of the human gene while copy 2 (chimptrans2-SNAP29) encodes exons 4 and 5.

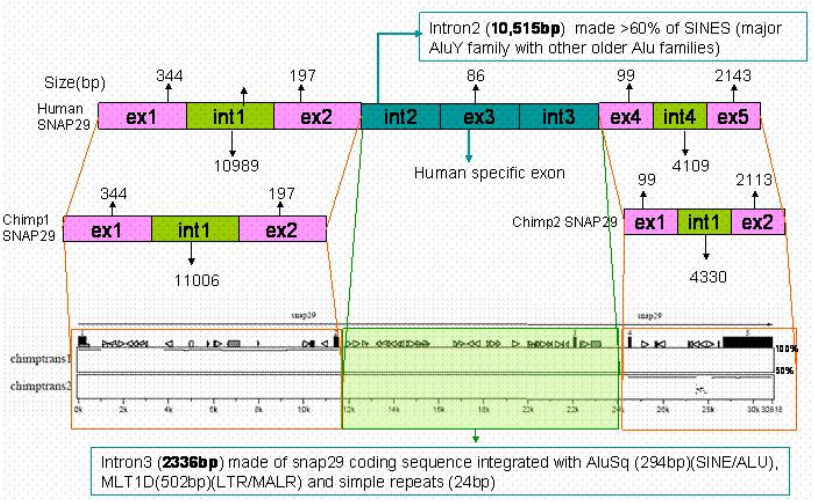


Figure 3.8 Insertion of SINE repeats in human SNAP29 gene. SNAP29 a single copy gene in human originated by insertion of repeats between the two copies present in chimpanzee as seen in the cartoon and the PIP output**.**

The ortholog of the human SNAP29 exon 3 is not present in either of the chimpanzee gene. Intron 2 of the human SNAP29 gene, is made almost entirely of SINEs, and particularly the AluY repeat that is the bridging intron for an additional 83 bp exon that connects the two separate transcripts in chimpanzee to a single transcript in human. The non-repeat regions in this intron showed sequences matching a portion of the cDNA of SNAP29 indicating that the intron was expanded by the addition of an exon likely driven by the presence of SINEs. Interestingly, the human SNAP29 gene organization of a single 5 exon gene in the genome is conserved from human through mouse and zebrafish as shown in the figure 3.9.

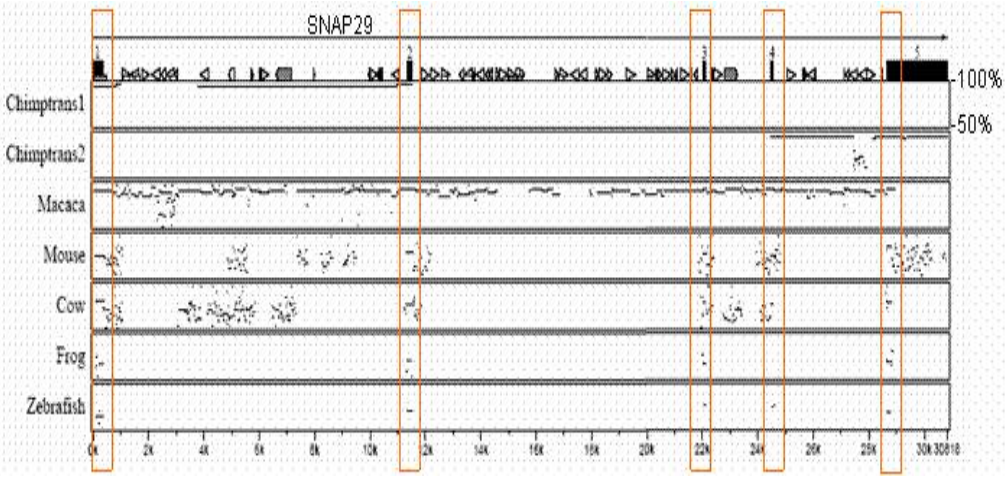


Figure 3.9 PIP output of SNAP29 gene highlighting conservation of exons and deletion in chimpanzee. The PIP output shows conservation of all exons between human, macaca, mouse, cow, frog and zebrafish indicated by red boxes. The missing segment in between the two chimpanzee copies represents the deletion in chimpanzee resulting to two copies.

In comparison to macaca, the intron 2 in human has more AluY and AluSg SINEs insertions. Intron 3 and exon 3 are similar in human and macaca. Thus the species comparison shows that intron 1, intron 2, exon 3 and intron 3 is actually deleted in chimpanzee resulting in the two copies.

Similarly human Pex26 gene has only one copy of the 7 exons, while chimpanzees have two copies, one with 2 exons and the other with 5 exons. Intron 2 in humans is the connecting link between two transcripts in chimpanzee that results in the one transcript in human as shown in figure 3.10. This new intron 2 in human is formed of parts of coding regions of Pex26 with a 142 bp insert of FLAM_C (SINE/Alu) insert.

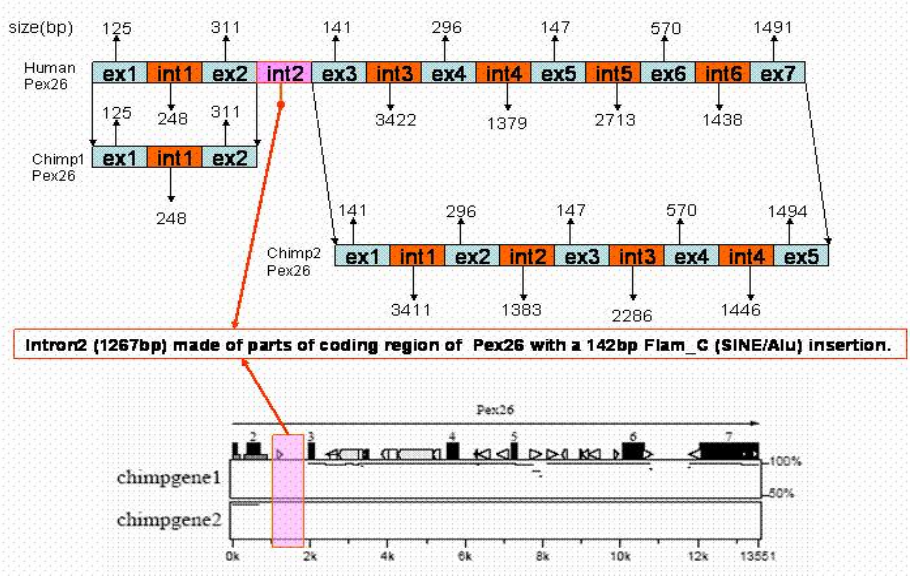


Figure 3.10 PIP ouptut highlighting unique SINE insertion in human Pex26 gene. Pipmaker output shows the unique intron2 only present in human that indicates the fusion of the two copies of Pex26 gene in chimpanzee.

Comparison of PEX26 gene with other species as shown in figure 3.11, shows two copies in mouse.

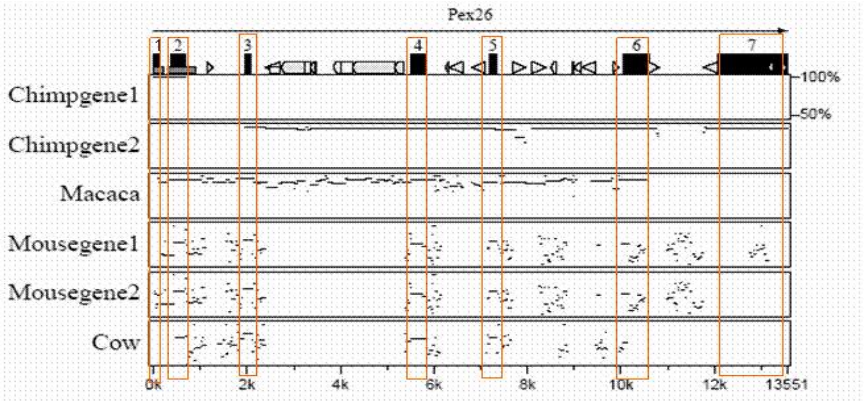


Figure 3.11 PIP output of Pex26 gene highlighting conservation of exons and deletion in chimpanzee. The PIP output shows conservation of all exons between human, and mousegene1 but the last exon 7 is missing in macaca, mousegene2 and cow. The exons conserved are indicated by red boxes. The missing segment in between the two chimpanzee copies represents the deletion in chimpanzee resulting to two copies.

One copy of the mouse gene has all exons conserved similar to the single human gene ortholog. The second copy of the mouse gene has the last exon 7 missing and has a similar gene organization as

seen in macaca and cow. Intron 2 in macaca is different from intron 2 in human and has a AluJo (133 bp) insertion

This comparison between humans, chimpanzees and other species not only shows how the single transcripts in human evolved, but also helps in indicating the susceptibility of the human genome to a variety of different diseases. This can be explained, as an Alu sequence originally has nine potential 5'donor sites and fourteen 3' acceptor sites. The plus strand of the Alu element is known to have only four of the potential splice sites and the remaining is on minus strand. In this case, there may be a possibility of converting an intron into exon containing the Alu repeat, if the Alu element is inserted into the intron in an orientation opposite to the transcriptional direction of the gene in which it resides. The alteration thus caused by Alu insertions into exons may lead to human diseases by disrupting the genes involved in metabolism, signaling, and transport that are known to have a high number of Alu repeats clustered within them. (Mitchell *et al.,* 1991; Kreahling *et al.,*2004; Dagan *et al*., 2004).

3.2 Significance of Nucleotide and Amino Acid Changes

3.2.1 Substitution Preference between First, Second, and Third Codon Position between Human, Chimpanzee, and Baboon

The nucleotide and amino acid changes between humans and other species not only will help in understanding the evolutionary history of the human species, but also will help in finding out the critical changes that differentiates human and other species. Appendix table 2 summarizes the nucleotide positional changes and preferences. It can be seen that 66% of the nucleotide changes observed between humans and chimpanzees

is in the third position, with only 34% of the changes in the second and first positions. In the case of nucleotide changes between human and baboon, it can be seen that >60% of the changes are in the third position, and 39% of the changes are in the second and third positions. In the comparison between chimpanzee and baboon, 64% of the changes are in the third position and 36% of the changes are in the second and first positions.

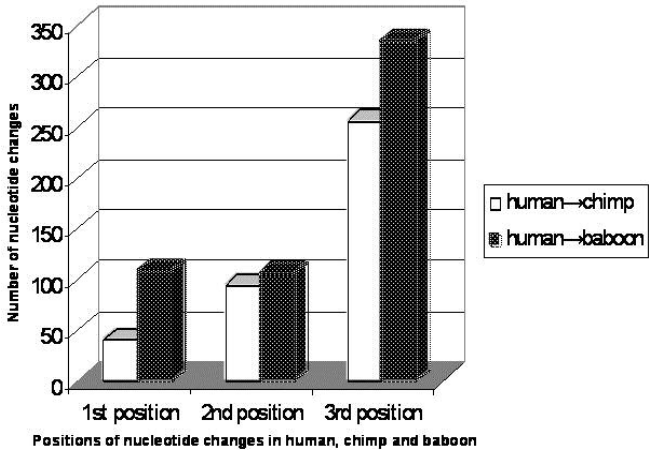


Figure 3.12 First, second and third position nucleotide changes. Graph showing the preferences in position of the nucleotide changes between human and chimpanzee and human and baboon.

Overall, the majority of the changes are conserved changes because they mainly reflect substitutions in the third codon position. Figure 3.12 summarizes the three codon position substitutions between human and chimpanzee, and human and baboon. This data clearly shows that substitutions between human and baboon are greater than those observed for human and chimpanzee especially in the first and third positions. This also is supported by the substitution rate calculated using the program eShadow (Ovcharenko *et al*., 2004) as shown in Appendix table 2, where the average substitution rate (per base) between human and chimpanzee is 1.2% (23 gene set) while

the rate between human and baboon is 2.6% (set of only 11 genes) as shown in figure 3.13.

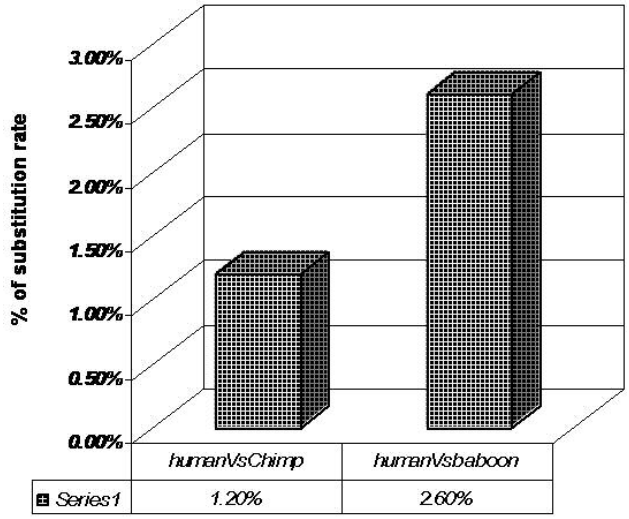


Figure 3.13 Substitution rate between human, chimpanzee and baboon. Graph showing the decreased substitution rate between human and chimpanzee in comparison to human and baboon.

This is similar to earlier observations seen between human and macaque and human and chimpanzee on chromosome 21 (Magness *et al.,* 2005; Watanabe *et al*., 2004). All of these results indicate that purifying selection is playing an important role between human and chimpanzee, and that positive selection is playing a major role in human and baboon in comparison to human and chimpanzee (Kimura, 1977; Hughes and Nei, 1988).

3.2.2 Transitional Changes versus Transversional Changes

As shown in figure 3.14, most of the substitutions observed when comparing the human, chimpanzee and baboon codons are transitional changes (>70%) compared to transversional changes (<30%), a result similar to earlier studies (Anzali *et al*., 2003).

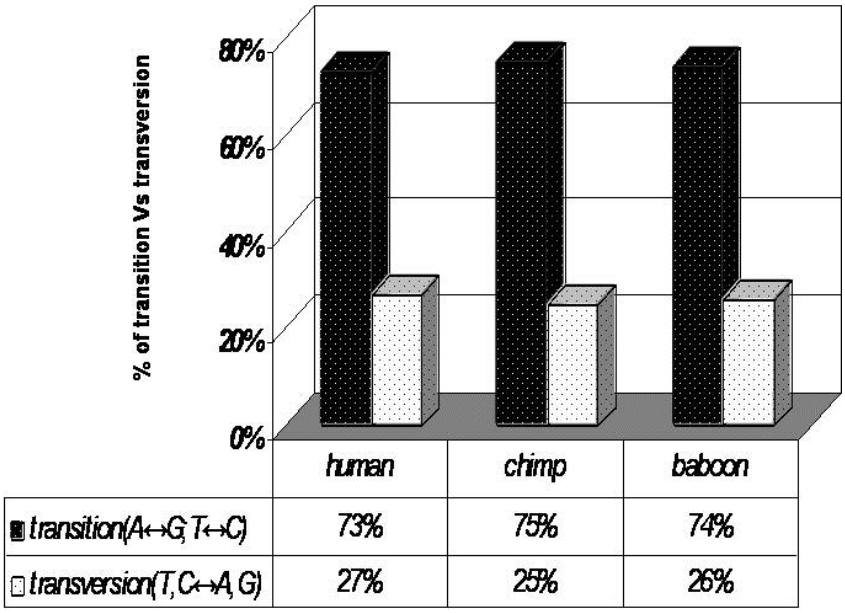


Figure 3.14 Transitional vs Transversional changes. A higher percentage of transitional changes are seen in humans, chimpanzees and baboons in comparison to transversional changes.

In human, G→A and C→T transitional changes are more frequent than A→G and T→C changes. In chimpanzee, G→A and C→T transitional changes are more frequent than A→G and T→C changes and less than that noticed in human. But in baboon A→G and C→T changes are more frequent than G→A and T→C changes. The transversions noticed in all three species show mainly G→C or C→G changes compared to all other transversions. In bovine, it could be observed that the changes also were transitional. Also, dinucleotide changes in codons instead of a single nucleotide change as commonly seen in human, chimpanzee, and baboon.

3.2.3 Synonymous(Ks) versus Nonsynonymous(Ka) Changes

To determine the extent of selective pressure or constraints influencing the coding regions, the ratio of Ka, the number of nucleotide substitutions that change amino acids per nucleotide sites resulting to amino acid changes to that of Ks, the

number of substitutions that do not change amino acids per nucleotide sites was calculated using DNASP (Rozas and Rozas 1999). Usually, when a Ka/Ks value is equal to one, the gene is said to be undergoing neutral changes, when a Ka/Ks value is >1, the gene is said to be undergoing positive selection fixing advantageous mutations but when a Ka/Ks value is <1, the gene is likely undergoing purifying selection against deleterious mutations (Zhang *et al*.,2003). In the comparison between human and chimpanzee, among the 39 genes for which the Ka/Ks ratio was calculated, 35 genes showed Ka/Ks values between 0.0-1.0 and only four genes between 1.0-3.0. Comparison between human and baboon in a set of 14 genes, (the genes compared between human and chimpanzee and human and baboon had genes both inclusive and exclusive of each sets) 13 genes showed Ka/Ks values between 0.0-1.0, only one gene had Ka/Ks value greater than 1 as shown in Appendix table 2. Thus, there is a possibility that the four genes in the set between human and chimpanzee, CECR6 (Ka/Ks=1.02), SLC25A18 (Ka/Ks=1.3), Kiaa0819-zf1 (Ka/Ks=1.05) and zf2 (Ka/Ks=1.14) and one gene in human and baboon comparison CECR6 (Ka/Ks=3.5) showing high Ka/Ks values of 1 or more may have mutations that are beneficial and were maintained by positive selection during evolution. In the overall comparison, among the 39 genes compared between human and chimpanzee, the average Ka/Ks ratio was 0.44. Between human and baboon 14 genes showed Ka/Ks=0.48 and between chimpanzee and baboon among 10 genes the average Ka/Ks=0.40. These results indicate that overall, purifying selection plays a major role in sequence conservation, as it prevents mutations from accumulating in functionally important regions where the

ratio Ka/Ks<1 (Wildman *et al.,* 2003) and also indicate the strong functional constraints in these species (Pagani et al, 2005).

3.2.4 Codon Usage Preference between Human, Chimpanzee, and Baboon

According to the genome hypothesis (Grantham 1980; Grantham *et al*., 1980, 1981), every organism is known to have its own codon usage preference that is selective and has nonrandom usage of synonymous codons for encoding different amino acids but is similar between genes in the same organism. Usually, there is a preference between codons ending with G or C versus A or T. In the comparison of the codon usage between human, chimpanzee and baboon, shown in Appendix table 3, there is an overall preference for codons ending with G or C with correspondingly high relative synonymous codon usage (RSCU) values in the range of 1-4 versus codons ending in A or T ending with RSCU values ranging between 0-2, as in the absence of any codon usage bias, the relative synonymous codon usage (RSCU) value would be 1 (Sharp et al, 1986; Martin-Galiano *et al*., 2004). In amino acids with 6 codons leucine, serine, arginine, codon usage is more biased towards G or C (82%) over A or T (18%) ending codons as seen in figure 3.15.

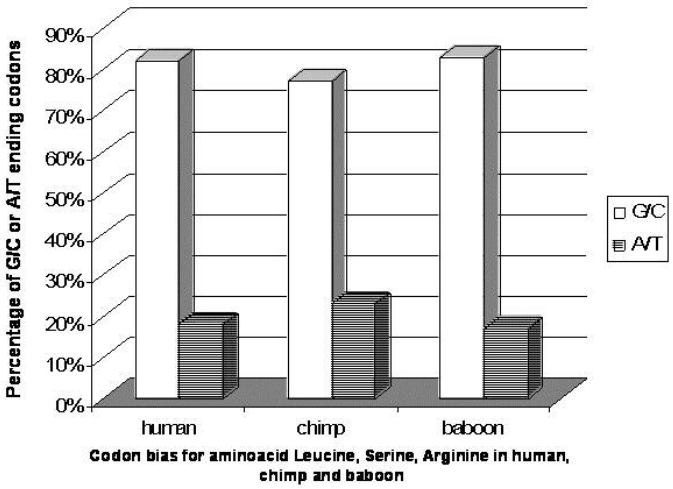


Figure 3.15 Codon usage bias in amino acids with six codons. Amino acids with six codons showing preference for G/C ending codons.

As shown in figure 3.16 (a), in amino acids with four codons such as proline, the codon bias between G or C (56%) versus A or T (44%) is almost equal between humans, chimpanzees, and baboons. However, in the case of threonine as shown in figure 3.16 (b), the baboon has more bias towards G or C when compared to humans and chimps.

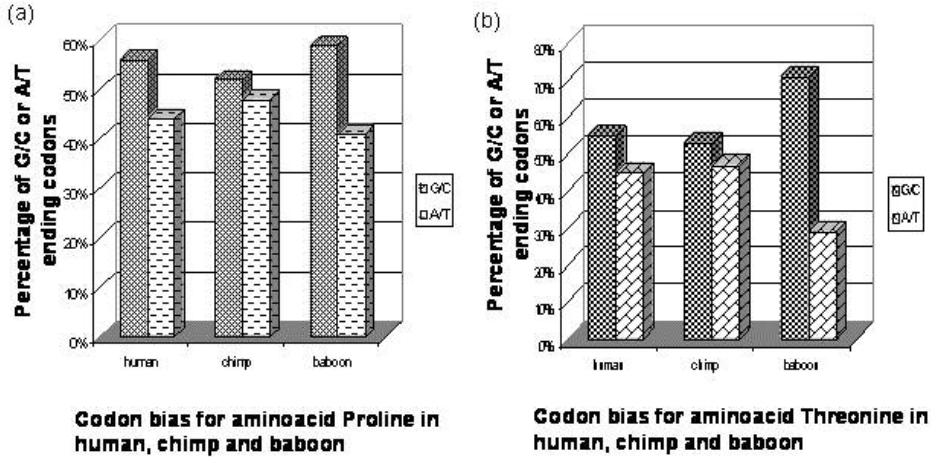


Figure 3.16 Preference for G/C or A/T ending codons. Codon for Proline (a) showing almost equal preference for G/C or A/T ending codons while Threonine (b) shows preference for G/C ending codons.

Glycine, valine, and alanine also are biased towards G or C ending codons. In the category of amino acids with two codons ending in either C or T, human, chimpanzee, and baboon show more bias towards codons ending in C over T, while in the category of amino acids with two codons ending in either G or A, humans, chimpanzees, baboons, and bovines show more bias towards codons ending in G over A. Similar results have been seen in drosophila with preference to G or C ending codons (Comeron *et al*., 1998) in contrast to *Pseudomonous aeruginosa* that has preference to A or T ending codons (Sau *et al*., 2005). These differences reflect the high codon usage bias in highly expressed genes that facilitates efficient translation (Wagner, 2000). Earlier studies have shown that preference for optimal codons over non-optimal codons boosts the translation mechanism as usage of less optimal codons correspond to less abundant tRNA, resulting to missense and nonsense errors during translation. Thus, there is a pressure for positive selection of optimal codons over non-optimal codons as it improves translational efficiency in an organism. In this comparison, the codon is more biased towards G or C ending codon which generally code for the most abundant tRNAs (Moriyama and Powell, 1997; Duret and Mouchiroud, 1999) and implies that the codon usage has been selectively conserved between the closely related species.

3.2.5 Substitutions, Insertions and Deletions : Hydrophobic vs Hydrophilic Amino Acid Changes

In the comparison between the human and chimpanzee, and baboon DNA sequenced, selective pressure plays an important role for the majority of the amino acid substitutions, insertions and deletions. The identification of insertions and deletions seen in human, chimpanzee and baboon in table 3.7 and Appendix table 4 was based on six possibilities as shown in figure 3.17.

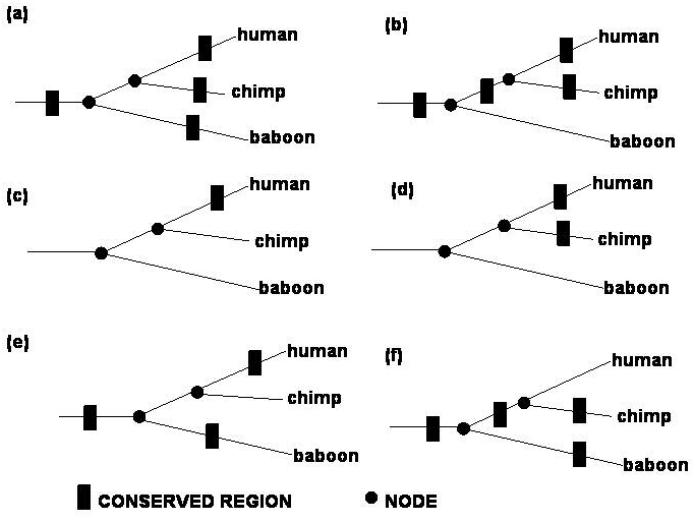


Figure 3.17 Phylogenetic tree representation to interpret indels. The figure represents the different explanations for insertions or deletions as recorded in table 3.8. (a) shows conservation in all species. (b) shows conservation in human, chimpanzee and an outgroup species other than baboon but deleted in baboon. (c) shows insertion unique only to human. (d) shows insertion in chimpanzee and conserved in human. (e) shows conservation in human and baboon and an outgroup species but deleted in chimpanzee. (f) shows deletion in human and conservation in chimpanzee, baboon and outgroup species.

Of the six possibilities shown in figure 3.17, the scenarios labeled a and c seem to occur more often than the others. Thus, there are more insertions than deletions in the human genome than in chimpanzee and baboon as well as differences in the sequences conserved in codon usage in the three sequenced chimpanzee libraries Clint (ch251),

Donald (rp43) and Gon (ptb) in comparison to humans although this evolutionary drift varies depending upon the human genes studied.

(**a**)

(**b**)

Gene	Human	Chimp	Baboon
Q8N7E1/XKRY	aa 197A→206S:aa 423Y→S459 unique in human.(**insertions**)	Absent in chimp	
ENSG00000138860 /GRB2	absent in human	1I→146F unique in chimpanzee (**insertion**)	absent in baboon
	Deleted in human	**deleted** in chimp	314F→322W unique in baboon
	736T→748P (**insertion** in human)	absent in chimp	absent in baboon
	conserved in human	conserved in chimp	aa254C→270V highly diverged
IL17R	aa 184D→199S ; aa252IPA254; aa310PEPI313aa (**insertions in human**)	Absent in chimp	Absent in baboon
CECR6	aa1M→200Y ; aa212G→254G ; aa276L→281A conserved in human and baboon	**Deleted** in chimp	aa1M→200Y; aa212G→254G ;aa276L→281A conserved in human and baboon
CECR5	aa1M→42Q (unique to human)	Conserved with baboon with two aa changes in this region(VH→RA)	Conserved with chimpanzee with two aa changes in this region (RA→VH)

Genes	**Human**	**Chimp**	**Baboon**
CECR1	**Deleted** in human	aa252P→278V conserved in chimpanzee and baboon	aa252P→278V conserved in chimpanzee and baboon
	absent in human	**deleted** in chimp	aa279S→aa390 Q only in baboon
	Absent in human	aa651L→736A (unique to chimpanzee only) **insertion**	absent in baboon
	Deleted in human	737A→835P conserved in chimpanzee and baboon	737A→835P conserved in baboon and chimp
CECR2	absent in human	V421→I499 unique in chimp	
	732V→M759 **insertion** in human	1596V→1660D **unique to chimpanzee ptb clone, absent in rp43 (Donald)**	
		Absent in chimp	
SLC25A18	Aa209G→264R (diverged from chimpanzee insertion, substitution)	1M→56R; 124G→179G (**specific to ptb chimp**)	
	aa209G→220A (specific to human-**insertion**)	221G→264G(unique in chimp)	
BCL2L13	1M→40Q specific to human and missing in ptb chimp(similarity starts at 41G with ptb)	201M→234L ;521→556T (specific to ptb clone)	
BID		573R→592D (unique in chimp)	

(**c**)

(d)

Genes	Human	Chimp	Baboon
O94909(pro5) kiaa0819		aa 1M→163I ;966A→1089P (unique in chimp)	
	aa 598 SPPPPPPPGE617 (in human)	aa598IIIIIIIII617 (in chimp)	
	1138G→1245A (diverged in human by insertion, substitution)	1138G→1245A (diverged in chimpanzee by insertion, deletion, substitution)	
	1246M→1256S (unique in human-**insertions**)		
Q7RTP6/MICAL3	aa739D→758E; aa1066 G→1077I specific to human not in chimpanzee (**insertion**)	aa317V→386L; aa820D→827T; aa854T→944L unique to chimpanzee not in human. (**insertion**)	
	aa 1080→1145 poor conservation between human and chimpanzee showing great divergence		
Q7Z2D7/pex26	aa 157ILLYS161 unique **insertion** in human and Clint	Absent in chimpanzee Donald.	Absent in baboon
	aa 9M→100T similar between human and baboon with two hydrophobic and one hydrophilic change	aa 1→100 in chimpanzee different from human and baboon showing positive selection of changes in this region	aa 9M→100T similar between human and baboon with two hydrophobic and one hydrophilic change
DGCR6L	Absent in human	**Deleted** in chimp	aa221I→404P (present in only baboon)
PRODHL	aa 91L→272A (conserved in human and baboon)	**Deleted** in chimp	aa 91L→272A (conserved in human and baboon)
	Deleted in human	**Deleted** in chimp	aa 411A→450Q (specific to baboon)
		aa 451A→480C (specific to Donald chimp-**insertion**)	
		aa 574M→664A (deleted in **Donald chimp**)	

Genes	**Human**	**Chimp**	**Baboon**
O94909(pro5) kiaa0819		aa 1M→163I ;966A→1089P (unique in chimp)	
	aa 598 SPPPPPPPGE617 (in human)	aa598IIIIIIIII617 (in chimp)	
	1138G→1245A (diverged in human by insertion, substitution)	1138G→1245A (diverged in chimpanzee by insertion, deletion, substitution)	
	1246M→1256S (unique in human-**insertions**)		
Q7RTP6/MICAL3	aa739D→758E; aa1066 G→1077I specific to human not in chimpanzee (**insertion**)	aa317V→386L; aa820D→827T; aa854T→944L unique to chimpanzee not in human. (**insertion**)	
	aa 1080→1145 poor conservation between human and chimpanzee showing great divergence		
Q7Z2D7/pex26	aa 157ILLYS161 unique **insertion** in human and Clint	Absent in chimpanzee Donald.	Absent in baboon
	aa 9M→100T similar between human and baboon with two hydrophobic and one hydrophilic change	aa 1→100 in chimpanzee different from human and baboon showing positive selection of changes in this region	aa 9M→100T similar between human and baboon with two hydrophobic and one hydrophilic change
DGCR6L	Absent in human	**Deleted** in chimp	aa221I→404P (present in only baboon)
PRODHL	aa 91L→272A (conserved in human and baboon)	**Deleted** in chimp	aa 91L→272A (conserved in human and baboon)
	Deleted in human	**Deleted** in chimp	aa 411A→450Q (specific to baboon)
		aa 451A→480C (specific to Donald chimp-**insertion**)	
		aa 574M→664A (deleted in **Donald chimp**)	

(**d**)

As summarized in table 3.8 below and Appendix table 4, the majority of the amino acid substitutions seen in humans, chimpanzees and baboons were changes between the same types of amino acids. The majority of the changes being from hydrophilic to hydrophilic amino acids in comparison to changes from hydrophobic to hydrophobic. However, the minor amino acid substitutions in humans and chimpanzees were changes from hydrophilic to hydrophobic amino acids.

Gene	human	chimp	baboon
	Hydrophilic (philic) /hydrophobic (phobic)	**Hydrophilic(philic) /hydrophobic (phobic)**	**Hydrophilic(philic) /hydrophobic (phobic)**
Q8N7E1/XKRY	philic→ philic	philic→philic	
ENSG00000 138860/GRB2	phobic→ philic	philic →phobic	
IL17R	philic→ philic	phobic→philic	philic→ philic
CECR6	phobic→phobic	phobic→phobic	phobic→phobic
CECR5	phobic→ philic	phobic→philic	philic→phobic
CECR1	philic→philic	philic→ philic	philic→ philic
CECR2	philic→ phobic	philic→ phobic	phobic→philic
SLC25A18	phobic & philic	phobic→phobic ; philic→philic	
BCL2L13	phobic→phobic; phobic→philic	phobic→phobic; philic→phobic	
O94909(pro5) kiaa0819	philic→philic	philic→philic	
Q7RTP6/MICAL3	philic→philic; philic->phobic	philic→philic phobic→philic	
Q7Z2D7/pex26	philic→philic	philic→philic	philic→phobic
Dgcr6(more of dgcr6l)	philic→philic	philic→philic	phobic→phobic; philic->philic
PRODH(more prodhl)	phobic→phobic	phobic→phobic	
DGCR2	phobic→phobic	phobic→phobic	
STK22B	philic→philic	philic→philic	
SLC25A1	philic→philic	philic→philic	
MRPL40	philic→philic	philic→philic	
CDC45L	phobic→phobic	phobic→phobic	
Q99648/ CDCREL-1	philic→philic; philic→phobic	philic→philic; philic→phobic	
GNB1L	phobic→phobic	phobic→phobic	
TXNRD2	philic→philic	philic→philic	
COMT	phobic→ phobic	phobic→ phobic	
ARVCF	phobic→ phobic	phobic→ philic	
Q8IZ69/HTF9C	philic→ philic	philic→ philic	
ZDHHC8	phobic→phobic	phobic→phobic	
ZNF74	philic→philic	philic→philic	philic→philic
SCARF2	philic→philic	philic→philic	philic→philic
PCQAP	philic→phobic	philic→ phobic	phobic→philic
SERPIND1	phobic→phobic	phobic→phobic	phobic→ phobic
SNAP29	philic→phobic	philic→ phobic	phobic→philic

Table 3.8 The table shows the major trends in amino acid substitutions (hydrophilic or hydrophobic) in different genes of human, chimpanzee and baboon.

In baboons, the minor amino acid substitutions were changes from hydrophobic to hydrophilic amino acids. In bovine, a trend similar to baboon is observed in Appendix table 4 showing that the major substitutions are changes from hydrophilic to hydrophilic and the minor substitutions are changes from hydrophobic to hydrophilic amino acids. Thus the majority of the changes are conserved changes between similar types of amino acids and mainly in the hydrophilic type of amino acids which indicates that since these residues are usually exposed, they are more susceptible to evolve at a faster rate in comparison to the hydrophobic amino acids which are usually buried (Yap and Speed, 2005, although the effect of these amino acid substitutions on the biochemical properties of the encoded proteins is unknown.

3.3 Gene Expression Studies in Zebrafish

Gene expression studies of human chromosome 22 (CES-DGCR) gene orthologs in zebrafish sheds light on when and where the genes are expressed as an initial step in investigating the function of their encoded proteins. Although there have been several northern blot, RT-PCR, and microarray studies on selected genes in the ~4.5 Mbp region of human chromosome 22 (Footz *et al*., 2001; Strausberg *et al*.,2000*,* 2002; Fantom *et al*., 2005; Nagase *et al*., 1999, 2000; Kawai J *et al*.,2001; Thierry-Mieg *et al*.,www.aceview.org; Lash *et al*., 2000; Matsumoto *et al*., 2003; Golmuntz *et al*., 1996; Heisterkamp *et al.,* 1995; Galili *et al*., 1997; Funke *et al.,* 1998; Maynard *et al*., 2003; Chieffo *et al*.,1997; Kochilas *et al*.,2003; Sirotkin *et al.,* 1997; Shiohama *et al*., 2003; Shmueli *et al*., 2003; Ishii *et al*., 2002; Lund *et al*., 2000; Berti *et al*., 2001; Schuler *et al*., 1997; Pontius *et al*., 2003; Wong and Cantley,1994; Nakagawa et al.,

1996 ; de Jong *et al*., 1995; Guris *et al*., 2001), detailed expression information often was unavailable for most of the genes in this region. Therefore, in the present studies, the expression profiles of the genes, whose probe size and the location of the probes is listed in the Appendix table 5 and schematically placed on human chromosome 22 in figure 3.18, and figure 3.29 were investigated.

3.3.1 CES Region Gene Expression Studies

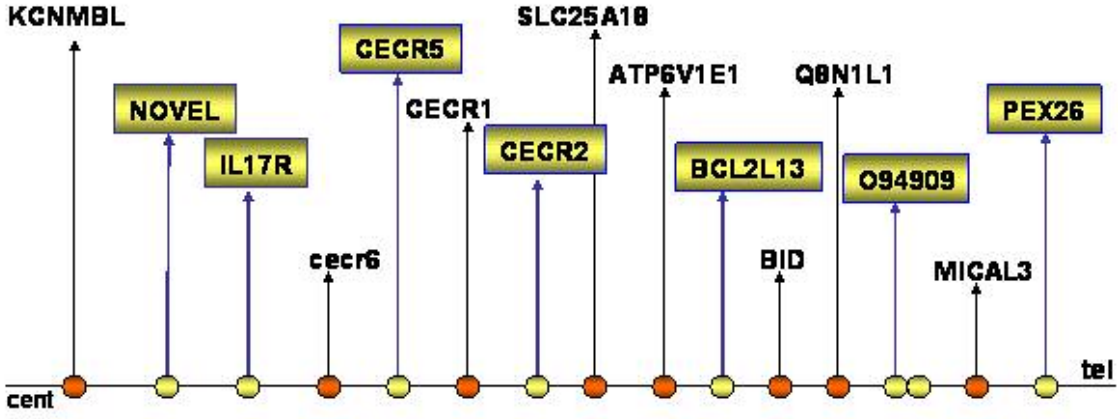


Figure 3.18 CES region of human chromosome 22. This region represents the genes in the CES region of human chromosome 22 and the highlighted genes were studied for their expression pattern in zebrafish model system. The second gene, that is labeled NOVEL, has been assigned an Ensembl gene number of ENSG00000138860.

3.3.1.1 ENSG00000138860

ENSG00000138860 located at position 15,818,177-15,863,443 on human chromosome 22 between genes KCNMB3L and IL17R and encodes for a 639 aa protein with 63% amino acid identity to the 360 aa zebrafish chromosome 1 ortholog. The gene has a pleckstrin-like domain but has no known description of function or expression pattern. Our whole mount *in situ* studies showed expression throughout the body at basal levels and more prominent expression in the tectum, midbrain, eye,

hindbrain, and spinal chord neuron in the early stages of development in zebrafish. Decrease in expression is seen in the brain by 72hpf , with expression mainly in the retina and branchial arches.

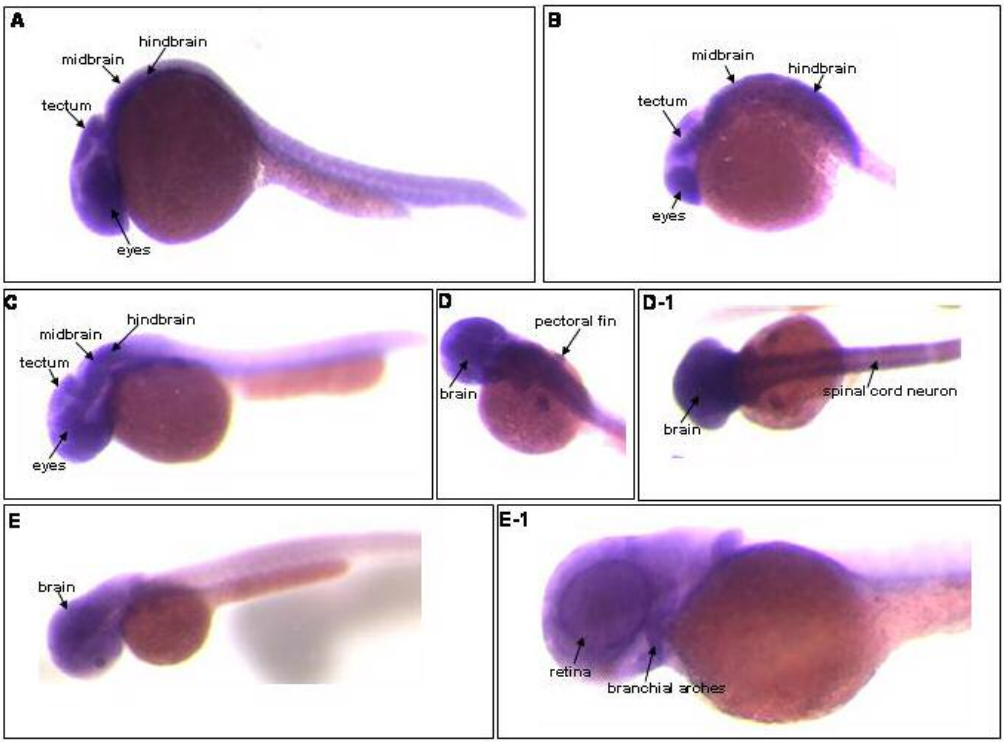


Figure 3.19 Expression pattern of the ENSG00000138860 gene. The whole mount expression pattern of the gene ENSG00000138860 was observed throughout the body. At 22-24hpf (A) and (B) and 35hpf (C) the expression is more concentrated in the eye, tectum, midbrain and hindbrain. At 48pf (D and D-1) the expression mainly is in the brain, pectoral fin and spinal chord neuron. At 72hpf E (top view) and F (side view) the expression mainly is observed in the brain and also in the retina and branchial arches.

3.3.1.2 Interleukin 17 receptor (IL17R)

IL17R, a gene that is located at the position 15,940,412 -15,965,941 in human chromosome 22, has been shown by Northern blot analysis to be widely expressed (Footz *et al*., 2001) and cDNA libraries of early embryo (embryonic day 0.5-4.5-15.5,18.5, postnatal day 1.0,5.0,15.0) of mouse showed expression mainly in brain (Strausberg *et al.,* 2002). This gene has 40% amino acid identity to the zebrafish ortholog with higher conservation in the receptor domain. The expression patterns observed in zebrafish also indicate extensive expression throughout the body, similar to that noted by others above, but in the present studies, expression seems mainly concentrated in the brain region during early stages.

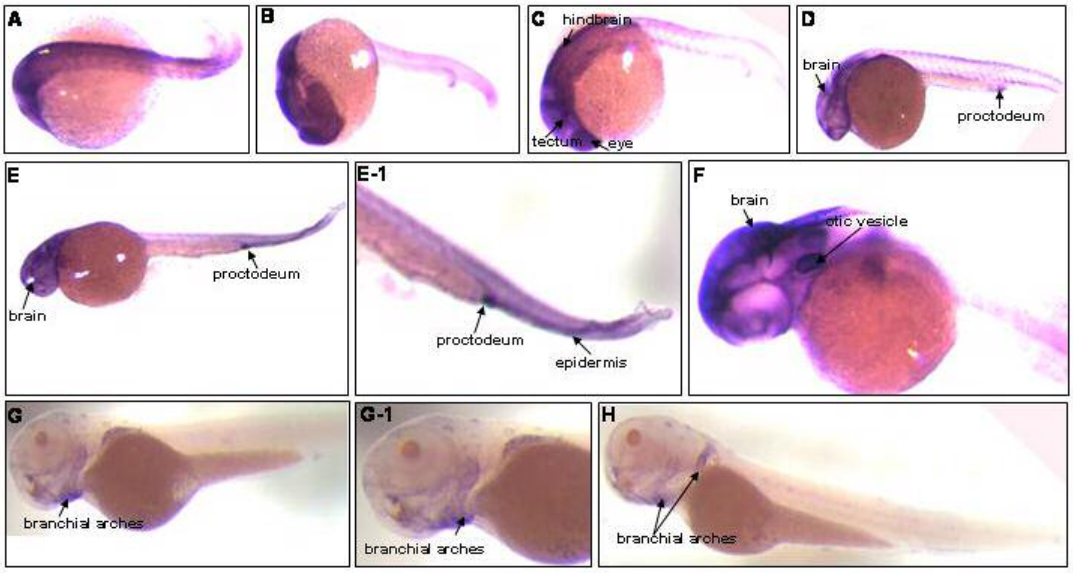


Figure 3.20 Expression pattern of the IL17R gene. The whole mount expression pattern of gene IL17R was observed mainly in the brain, tectum, midbrain and hindbrain and the eyes, at 19-24hpf stages shown in A, B and C expression patterns, with basal levels of expression in other parts of the body. At 35hpf (D) the expression mainly is restricted to the brain and proctodeum. At 42hpf (E) and (E-1) show expression in the brain and proctodeum. At 48hpf (F) the expression mainly is restricted the brain and otic vesicle. At 60hpf (G and G-1) and 72hpf (H) the expression mainly is observed in the branchial arches with no expression in the brain

The major expression occurs in the brain, tectum, midbrain and hindbrain and the eyes, as noticed at 19-24hpf stages, with basal levels of expression in other parts of the body. At stages 35 and 42hpf, the expression mainly is restricted to the brain and proctodeum. At 48hpf the expression mainly is

restricted in the brain and otic vesicle. During later stages (60 and 72hpf) the expression mainly is observed in the branchial arches with no expression in the brain. Since the earlier studies showed that il17r are known to be cytokine receptor belonging to the type 1 membrane protein and involved in signaling (Aggarwal *et al*.,2002; Strausberg R,L *et al*., 2002), the expression in the brain during early stages and the ubiquitous distribution may indicate that this gene is essential as an immune response regulator both in the embryos and adults (Yao *et al*.,1997).

3.3.1.3 Cat Eye Syndrome Chromosome Region, Candidate 5 (CECR5)

CECR5, a gene present at 15,992,955-16,020,731 in human chromosome 22, is the ortholog of a zebrafish gene located on chromosome 23. This gene encodes for a 423 aa with 55% amino acid identity to the 399 aa ortholog in zebrafish. Previous northern blot analyses were done only on adult tissues and showed that expression was present in all tissues studied. cDNA libraries constructed from postnatal mouse embryos show positive expression in brain libraries with additional expression in the retina and kidney (Footz *et al*., 2001; Strausberg R,L *et al.,* 2002; Fantom *et al*., 2005). The present whole mount *in situ* hybridization studies in zebrafish embryos indicated expression throughout the body and major concentration in the brain. At early stages expression is throughout the body with higher expression in the brain, liver, pectoral fin and pronepheric duct. But at later stages, the expression mainly is in the retina and branchial arches with reduction in expression in the brain. The level of expression in the brain is thus at its peak at early embryonic stages and decreases during later stages. CECR5 has a signal peptidase l S26A, domain that has been implicated in serine peptidase activity (Dalbey and Von Heijne, 1993) a function that must be required in the observed tissues at the different developmental expression times.

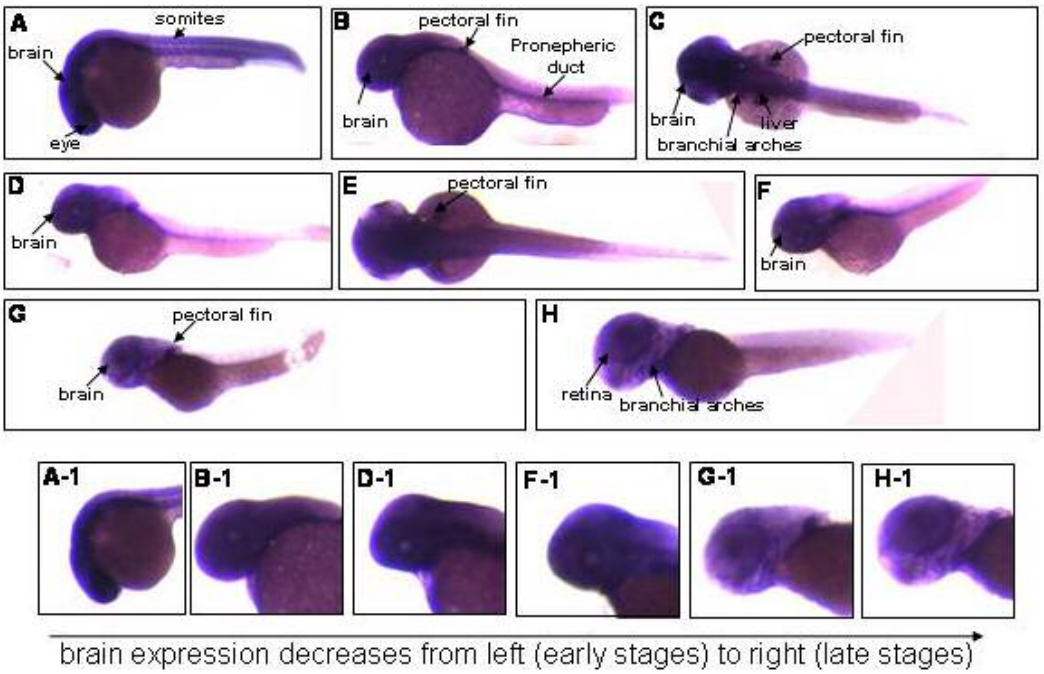


Figure 3.21 Expression pattern of the CECR5 gene. The whole mount expression pattern of the gene CECR5 was observed throughout the body and major concentration is in the brain. At (24hpf)A expression is throughout the body, B (35hpf), C (35hpf –dorsal view) also throughout the body with higher expression in the brain, liver, pectoral fin and pronepheric duct. At 42hpf D and E (dorsal view) the expression is throughout the body with higher expression in the brain, pectoral fins and pronepheric duct. At 48hpf (F) the expression mainly is in the brain, pectoral fins and pronepheric duct. At 60hpf (G) the expression mainly is in the brain and pectoral fins. At 72hpf (H) the expression mainly is in the retina and branchial arches with reduction in expression in the brain. The level of expression in the brain is at its peak at early embryonic stages as seen in A-1(24hpf), B-1(35hpf), D-1(42hpf), a slight reduction can be noticed at F-1(48hpf) and reduced expression in later stages at G-1(60hpf) and H-1(72hpf) is evident.

3.3.1.4 Cat Eye Syndrome Chromosome Region, Candidate 2 (CECR2)

CECR2, present at position 16,331,181-16,408,399 in chromosome 22, has a 60% amino acid identity with its zebrafish ortholog. Northern blot analysis showed that this gene is expressed in all tissues and EST sequences from a cDNA library (kiaa1740) and mouse cDNA (embryonic day 13.5,14.5,16.5,17.5) showed elevated expression in the brain (Nagase *et al*., 2000; Kawai J *et al*.,2001) as well as cDNA in muscles, retina and lungs (Thierry-Mieg *et al*.,www.aceview.org). This gene has an AT hook, a DNA

binding motif as a transcription factor, two bromodomains that are a class of regulatory proteins that mediates protein interactions necessary for transcription activation, and a homology to guanylate binding protein-1, associated with a transcription factor involved in neurulation and chromatin remodeling (Banting *et al*., 2004).

The present whole mount *in situ* studies confirmed the earlier reported elevated expression in brain but extend this observation now to include the otic vesicle, notochord, and pectoral fins. The expression pattern is seen throughout the body at 19 and 22 hpf stages. The expression is present all over the body also at stage 31hpf but higher levels of expression are seen mainly in the brain and notochord. At stages, 35, 42 and 48 hpf, the expression is all over the body with higher concentration in the brain, pectoral fin and otic vesicle. At 72hpf, the expression mainly is concentrated in the otic vesicle, retina and basal expression is seen in the brain and branchial arches.

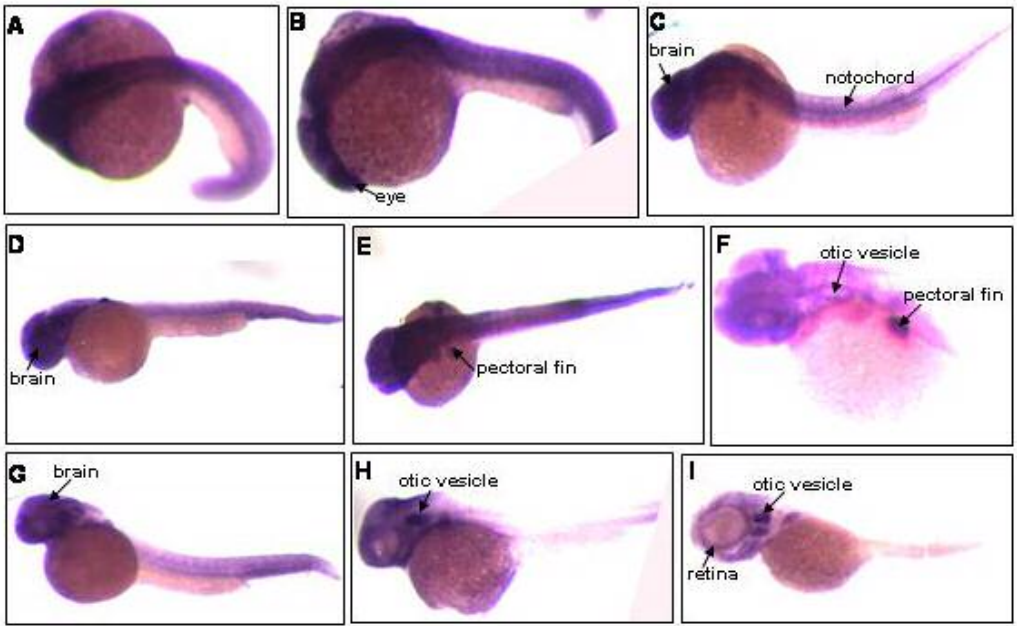


Figure 3.22 Expression pattern of the CECR2 gene. The whole mount expression pattern of gene CECR2 in (A) 19hpf and (B) 22hpf are present throughout the body. At 31hpf (C) the expression is also present throughout the body with higher levels of expression mainly in the brain and notochord. At 35hpf

(D) and (E) (dorsal view) and 42hpf (G) the expression is throughout the body with higher concentration in the brain, pectoral fin and otic vesicle. At the 48hpf (H) the expression mainly is concentrated in the brain, otic vesicle and the pectoral fin. At the 72hpf (I) the expression mainly is concentrated in the otic vesicle, retina and light expression in the branchial arches and brain.

3.3.1.5 BCL2-like 13 apoptosis facilitator (BCL2L13)

BCL2L13 is a gene located at the position 16,486,236-16,587,937 with 47% identity at the amino acid level to the zebrafish ortholog. Earlier Northern blot studies have shown foetal expression to be mainly in brain followed by lung, liver, and kidney, and in all tissues in the adults (Footz *et al*., 2001).

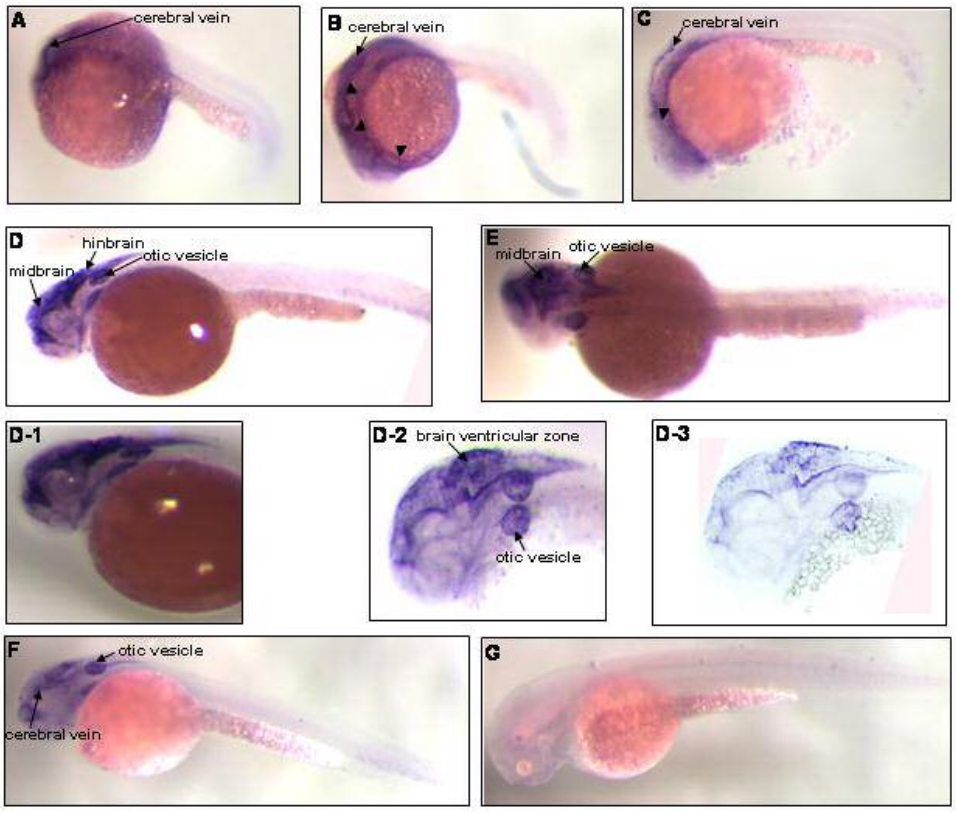


Figure 3.23 Expression pattern of the BCL2L13 gene. At 19hpf (A), 22hpf (B) and 24hpf (C) the expression mainly is in the cerebral vein. At (D) 35hpf and 42hpf (E-dorsal view) the expression is more concentrated in the midbrain, hindbrain and otic vesicle. D-1, D-2 and D-3 (35hpf-42hpf) show the

expression at higher magnification and shows the details of restricted expression in the midbrain and hindbrain region and otic vesicle. A similar expression is also shown at 48hpf (F). No expression is seen at the stage 72hpf (G).

SAGE expression studies showed expression in brain, spinal chord, and muscles (Strausberg R,L *et al.,* 2000, 2002; Lash *et al*., 2000). This gene is known to encode for a apoptosis inhibitor domain belonging to the Bcl2 family (Kataoka *et al*., 2001).Whole mount *in situ* studies showed expression mainly in the forebrain and midbrain compared with the highest expression at 42hpf, and with no expression at 72hpf. In the early stages, 19-24hpf embryos showed basal level expression throughout the anterior part of the embryo with higher expression in the cerebral vein.

3.3.1.6 O94909/Kiaa0891

O94909/Kiaa0891, a chromosome 22 gene (11 exons) located at the position 16,644,972-16,689,189, has two orthologs in zebrafish, one being zf1, present at 307280-316461bp of ctg14067 and the other zf2, present at 107344-119287bp of ctg11065, with 4 and 7 exons respectively. Similarly two copies of this single copy human gene also are observed in chimpanzee.

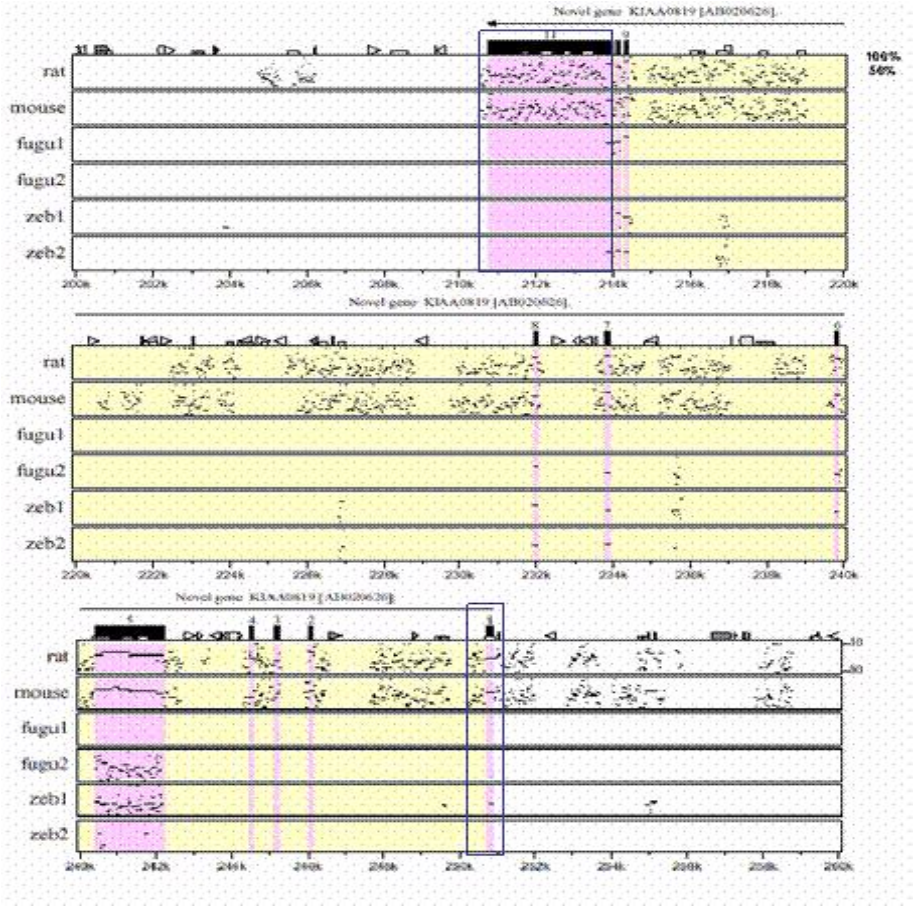


Figure 3.24 zf1 and zf2 gene comparison to human Kiaa0819 gene. The PIP output clearly shows the conservation of specific exons in zf1 and zf2 in comparison to human. The blue boxes show the exons from which the probes were made to test the expression pattern for each of the copies of kiaa0819 human gene. zf1 probe was specifically made from exon1 unique to zf1 and human and zf2 probe was made specifically from exon11 unique to zf2 and human.

There is a unique intron 9 in human made entirely of Alu repeats that joins these two transcripts in human. To investigate this novel gene, a specific zf1 probe was made from the zebrafish exon1 that matches exon1 of the human gene and a specific zf2 probe was made from zebrafish exon4 that matched exon 11 of the human gene as shown in figure 3.24 above.

3.3.1.6.1 Expression of zf1



Figure 3.25 Expression pattern of the zf1 gene. The whole mount expression pattern of gene zf1 at 24hpf (A), 48hpf(B), 72hpf (C), 120hpf(D) shows the expression in the ear. E(48hpf) shows the three cristae and the two maculae in the otic placode.

The whole mount *in situ* hybridization with zf1 sense and antisense ssDNA probes indicated that zf1 the ortholog of human gene O94909 (kiaa0819) shows expression in the otic placode showing expression in the distinct sensory patches-three cristae (associated with a semicircular canal) and two maculae (associated with an otolith) in 48hpf and starts expression by 24hpf and continues until 120hpf as seen in figure 3.25. No expression was observed using the sense probe indicating an antisense, opposite orientation RNA was not presence.

3.3.1.6.2 Expression of zf2

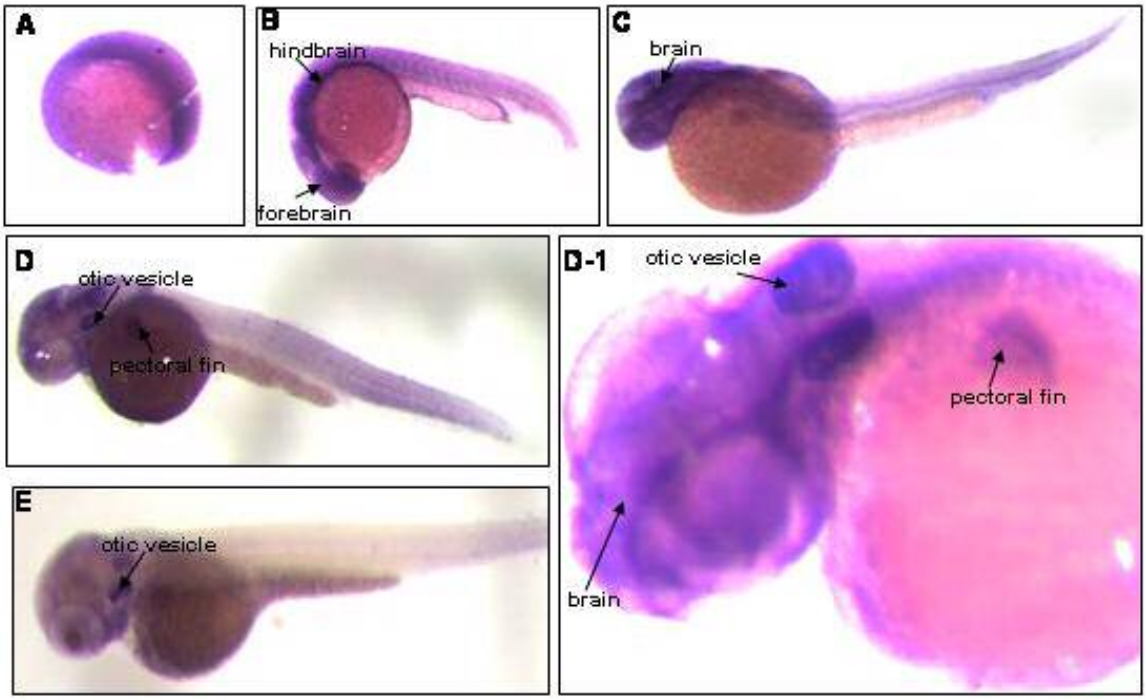


Figure 3.26 Expression pattern of the zf2 gene. The whole mount expression pattern of zf2 in (A) at 12hpf shows expression throughout with higher expression in the brain as also seen at 24hpf(B) and 35hpf (C). During 48hpf (D), the expression is restricted to the otic vesicle and pectoral fins and a enlarged view of the 48hpf expression can be seen in D-1. At 72hpf (E) the expression mainly is in the otic vesicle.

The expression of zf2 mainly is concentrated in the brain and pectoral fin, and in the later stages was mainly in the otic vesicle, and branchial arches with nonoverlapping pattern in comparison to zf1, and no expression was seen during 24hpf in the otic vesicle as shown above in figure 3.26.

3.3.1.6.3 Expression of zf1 in (*acerebellar*) *ace* mutant embryo

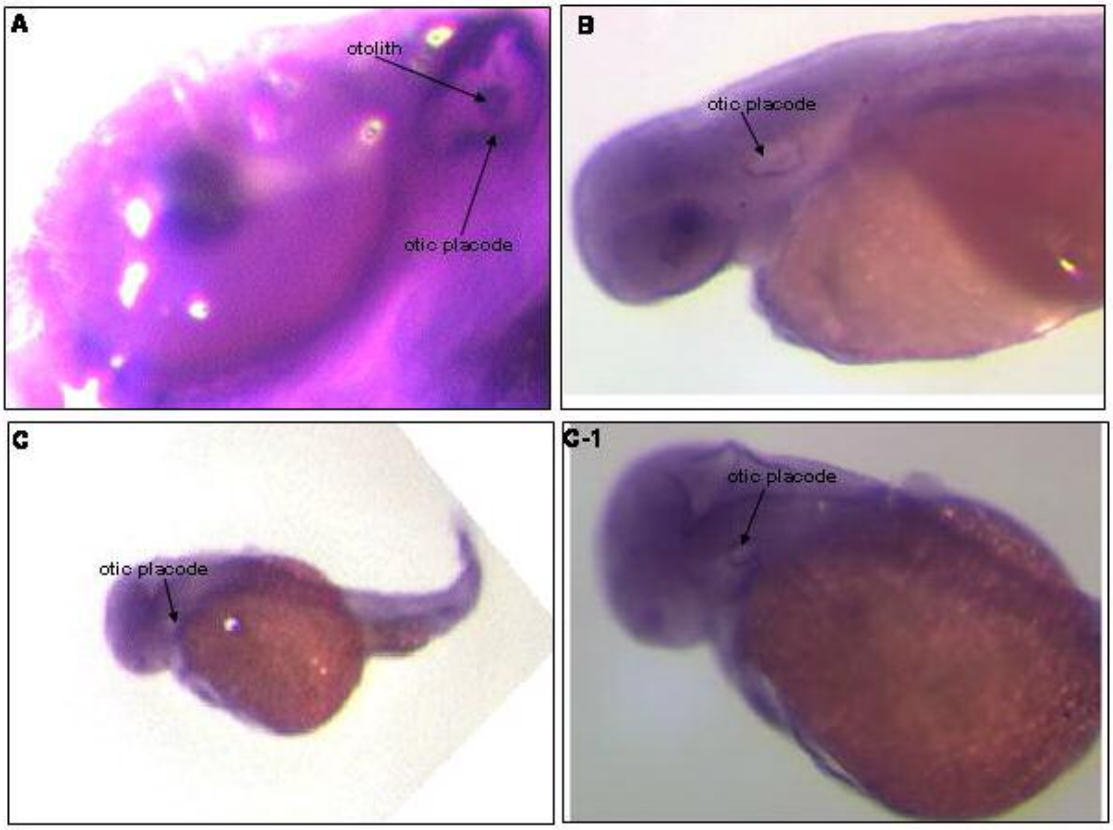


Figure 3.27 Expression pattern of the zf1 gene in an *ace* mutant. The whole mount expression pattern of zf1 gene in an *ace* mutant embryo showed no expression at all in the otic placode as the placode was missing cristae and maculae but in (A) as in case of certain mutants there is one otolith instead of the regular two and the expression can be seen at that particular spot. The light blue is the background in reference to controls. B and C shows abnormal otic palcode which is small and constricted. C-1 is an enlarged view of C showing constricted otic placode.

The expression of zf1 antisense probe in the *ace* mutant embryo (Brand *et al*., 1996) as seen in figure 3.27 showed reduced expression in the otic placode when compared to the wild type counterpart that had expression in the otolith, but no expression in the mutant where both the otolith and cristae were absent. The mutants that retained one otolith showed expression of zf1. The phenotype of this mutant as described by Brand *et al*., 1996 is lack of midbrain-hindbrain boundary, smaller otocyst with one otolith or no otolith and semicircular canal formation is affected and of reduced size.

It should be noted that since expression of the zf1 gene was observed in the otic placode, it is likely involved in ear formation. Interestingly, since one of the abnormalities of DiGeorge Syndrome is ear malformations, the deletion of this gene may be responsible for this observed ear abnormality. Early embryological experiments suggests that in a developing ear, the ear pattern formation depends on continual interaction of the otic vesicle with the hindbrain. Of the genes involved in ear development and function that have been cloned, a whole range of molecular identities is demonstrated from signaling molecules to proteins involved in transcription elongation (Whitfeild *et al*., 2002; Solomon *et al*., 2004). Further support for this correlation is obtained from the analysis of the *ace* mutant. The expression is highly reduced as *ace* mutants have the smaller ear and missing otolith (Brand *et al*., 1996). This indicates that the expression of the gene O94909 (kiaa0819) is specific to the sensory patches in the ear and has a similar function to Fgf8, can be dependent to Fgf8 for its induction, or it is the transcription factor it induces that are expressed by otic placode precursor.

The expression study of the zf2 gene shows that it is expressed in the early stages mainly in the brain as seen in figure 3.26 but in the later stages expression is seen in the otic vesicle, pectoral fin, and hindbrain. The expression differences between zf1 and zf2 show that on the basis of differences in expression patterns of zf1 and zf2 orthologs of human chromosome 22 genes, both zebrafish genes apparently play independent and nonredundant roles. It is well established that the teleost genome underwent a duplication after its divergence from the mammalian lineage (Jaillon *et al*., 2004). This duplication resulted in 2 or more paralogous copies of genes in the

zebrafish genome, each of which is orthologous to a mammalian gene. In some instances these appear to have been maintained through divergence of function so that different sub-functions of an ancestral gene are partitioned between its duplicate descendants. This division of sub-functions may be manifest by different expression patterns (Chiang *et al*., 2001; Whitfield, 2002).

	Transcription factor binding Sites (1000bp upstream of the first exon)
HUMAN	NIT2(2), GATA-1(3), ADR1(4), IK-2, HSF(8), CAP(2), SRY, STRE(2), CF2-II, AP-4, MZF1
ZF1	GATA-1(2), NIT2(2), CAP(5), HSF(15), SRY(3), IK-2, Skn-1,MATa1, c-Ets-,ADR1
ZF2	CREB, GATA-1(2), ADR1(4), HSF(8), CAP**,** GCR1

Table 3. 9 TRANSFAC Web Based software used to identify transcription factor binding sites showing conservation of regulatory elements between human and zf1 and human and zf2.

An analysis of the regulatory elements in the upstream sequences of the human, zf1 and zf2 genes as shown in the table 3.9 above indicate that there are conserved potential binding sites of transcription factors between human and zf1 that include GATA-1 (2), NIT2(2), CAP(5), HSF(15), SRY(3), IK-2, ADR1 domains and between human and zf2 that include GATA-1(2), ADR1(4), HSF(8), and CAP domains. These upstream domains observed for zf1 and zf2 also are observed in the single upstream region of the human gene. These results indicate that the expression of the gene in human would be the cumulative expression of zf1 and zf2 that may indicate duplication from a single gene. The unique domains and transcription binding sites indicate the

divergence. The loss of regulatory modules in each of the duplicates has allowed the survival of the two genes as seen in the case of hoxb1 gene (Prince and Pickett, 2002). Therefore, in zebrafish both genes may have evolved after gene duplication in the fish lineage and present differential transcriptional control and non redundant function, showing functional diversification after the fish-specific genome duplication.

To prove this further, a detailed study must be performed to investigate whether either of these unique and specific expression domains were due to "neofunctionalization" (duplicated genes developing new function and new expression pattern), "sunfunctionalization" (duplicated genes developing different function or different expression pattern) (Lynch and Force, 2000; Ohta, 2003) or were present in the ancient gene before duplication. Further experiments in mouse embryos will help clarify this observation, as there is a lack of expression information from the recent unduplicated tetrapod outgroup.

3.3.1.7 Peroxisome biogenesis factor 26 (Q7Z2D7/PEX26)

Q7Z2D7/PEX26, located at the position 16,935,240-16,948,970, has a 50% amino acid identity with the zebrafish chromosome 3 ortholog. This is an integral perioxisome protein that interacts with PEX1 and PER6 complex via a direct interaction with pex6. This gene has a AT-DNA binding domain (Reeves and Nissen, 1990). Thus its main function is protein binding. Northern blot analysis showed expression in high levels in brain and also expressed in liver, kidney, and skeletal muscles (Matsumoto *et al*., 2003).

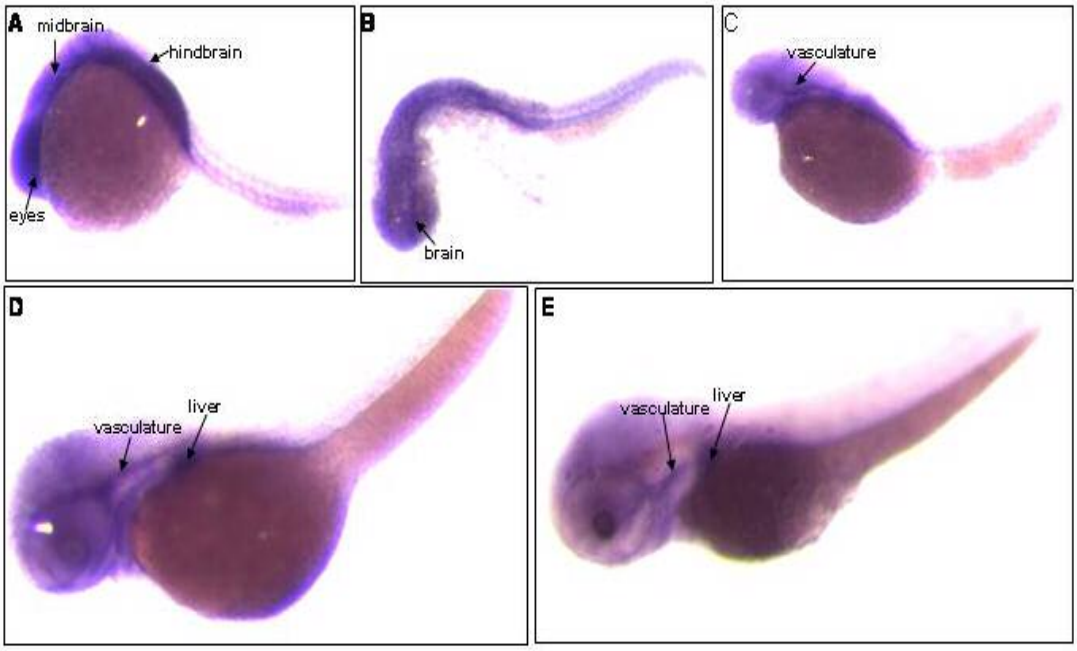


Figure 3.28 Expression pattern of the Pex26 gene. The whole mount expression pattern of Pex26 gene shows expression all throughout the body during early stages at 22-25hpf (A and B) but at higher level in the brain. At 35hpf (C) the expression mainly is in the blood vasculature and brain. At 48hpf (D) and at 72hpf (E) the expression mainly is in the vasculature and liver with lower level in the brain.

cDNA library EST sequences show that this gene is expressed mainly in brain (mouse embryo E12.5-15.5), kidney, pancreas, and liver (postnatal embryos), an observation confirmed by the CGAP (Cancer Genome Anatomy Project) est studies (Strausberg R,L *et al*., 2000, 2002). Our whole mount *in situ* hybridization studies showed similar expression throughout the body during early stages at 22-25hpf but at higher level in the brain. The expression mainly is in the blood vasculature and brain at 35hpf and the expression mainly is in the vasculature and liver with lower level in the brain during 48 and 72hpf.

3.3.2 DGCR Region Specific Gene Expressions

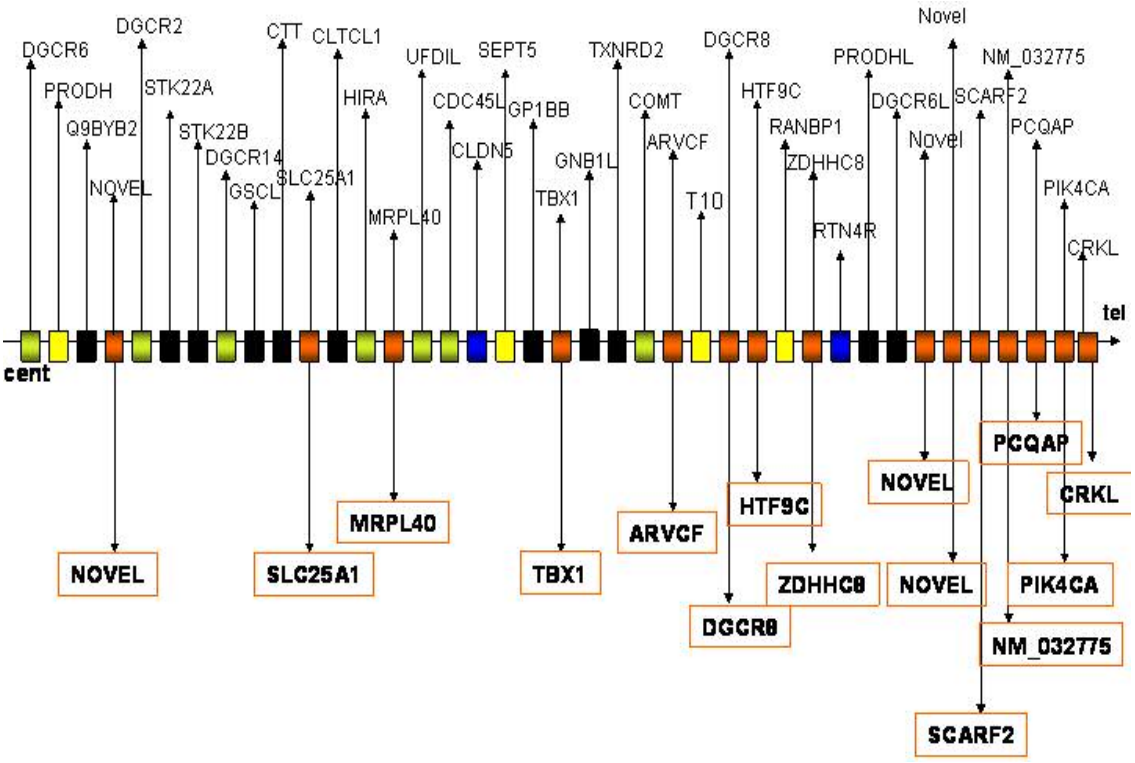


Figure 3.29 DGCR region of human chromosome 22. This region represents the genes in the DGCR region of human chromosome 22.The red boxes represent genes for which expression studies were performed. Green boxes represent genes with no single PCR products or PCR products of a significantly larger or smaller than predicted size . Yellow boxes represent genes with small exons (63-200 bases) where unique primers could not be picked successfully. Black boxes represents genes with no zebrafish orthologs. Blue boxes represent genes with only single exons in zebrafish and likely represented processed pseudogenes with the functional gene elsewhere in the genome. The NOVEL genes have been assigned Ensemble numbers ENSG00000185643, ENSG00000182364 and ENSG00000183099, respectively

3.3.2.1 ENSG00000185643

Gene ENSG00000185643, located at position 17,394,067-17,396,638 on human chromosome 22 between genes prodh and dgcr2, encodes for a 256 amino acid protein with 48% amino acid identity to the 309 amino acid zebrafish chromosome 21 ortholog. The gene has a carbonic anhydrase domain but has no previously studied expression pattern.

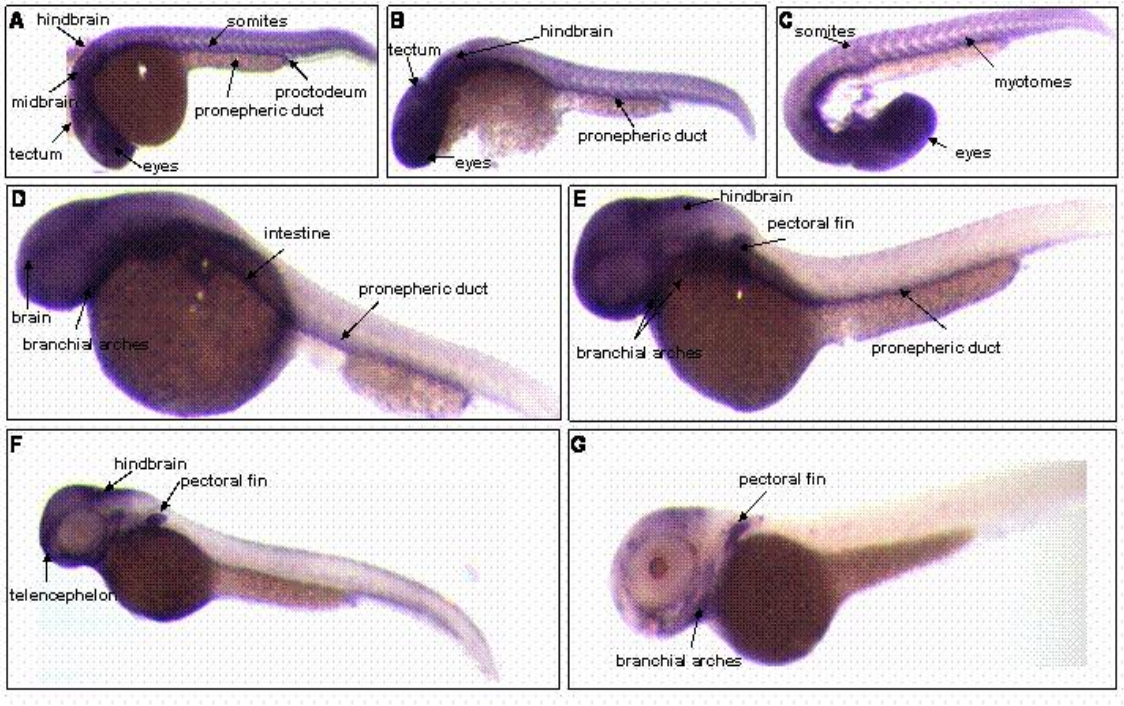


Figure 3.30 Expression pattern of the ENSG00000185643 gene. (A) At 24hpf increased expression in parts of brain in comparison to the other parts also showing expression like somites/ mytotomes, pronepheric duct and proctodeum. (B) and C (yolk sac removed) at 28hpf show expression in the brain, myotomes and pronepheric duct. (D) at stage 31hpf shows more expression in the branchial arches, intestine, pectoral fin and hindbrain. (E) at 42hpf and (F) at 48hpf show increased expression in the hindbrain, branchial arches, pectoral fins and at 48hpf no expression is noticed in the pronepheric duct while the same can be seen in 42hpf. (G) at 72hpf the expression is restricted to branchial arches and pectoral fin.

Our present whole mount *in situ* studies showed expression in midbrain, hindbrain, otic placode, pectoral fins, and mytotomes at early stages of development in zebrafish. At 24hpf an increased expression is seen in parts of brain in comparison to the other parts also showing expression like in somites/ myotomes, pronepheric duct and proctodeum. At 28hpf expression is seen in the brain, myotomes and pronepheric duct and at 31hpf expression is higher in the branchial arches, intestine, pectoral fin, and hindbrain. At stages 42hpf to 48hpf an increased expression is seen in the hindbrain, branchial arches, pectoral fins but at 48hpf no expression is noticed in the pronepheric duct while the same can be seen in 42hpf. It is seen that at 72hpf the

expression is restricted to branchial arches and pectoral fin. This expression pattern indicates that this gene likely is important in early developing brain due to its higher expression in this region.

3.3.2.2 Solute carrier family 25(mitochondrial carrier; citrate transporter), member 1) (SLC25A1)

SLC25A1 gene located at position 17,537,646-17,540,814, is an integral membrane protein that likely is a transmembrane transporter. The gene encodes for a 311 aa protein that has 74% amino acid identity to the zebrafish 304 aa ortholog. This gene has a mitochondrial carrier protein, substrate carrier, and an adenine nucleotide translocator 1 domain. Northern blot analysis showed expression in various adult tissues but not in brain but high expression levels in fetal brain, lung, liver, and kidney (Golmuntz *et al*., 1996; Heisterkamp *et al.,* 1995). RT–PCR on mouse tissues extracts showed expression in brain, limb buds, lungs, liver (Galili *et al*., 1997; FANTOM consortium, 2005). Our whole mount *in situ* studies showed high expression of this gene in the midbrain and hindbrain, frontal region, and in the pectoral fins during stages 24, 48, and 72 hpf even though basal levels of expression was noticed throughout the body.

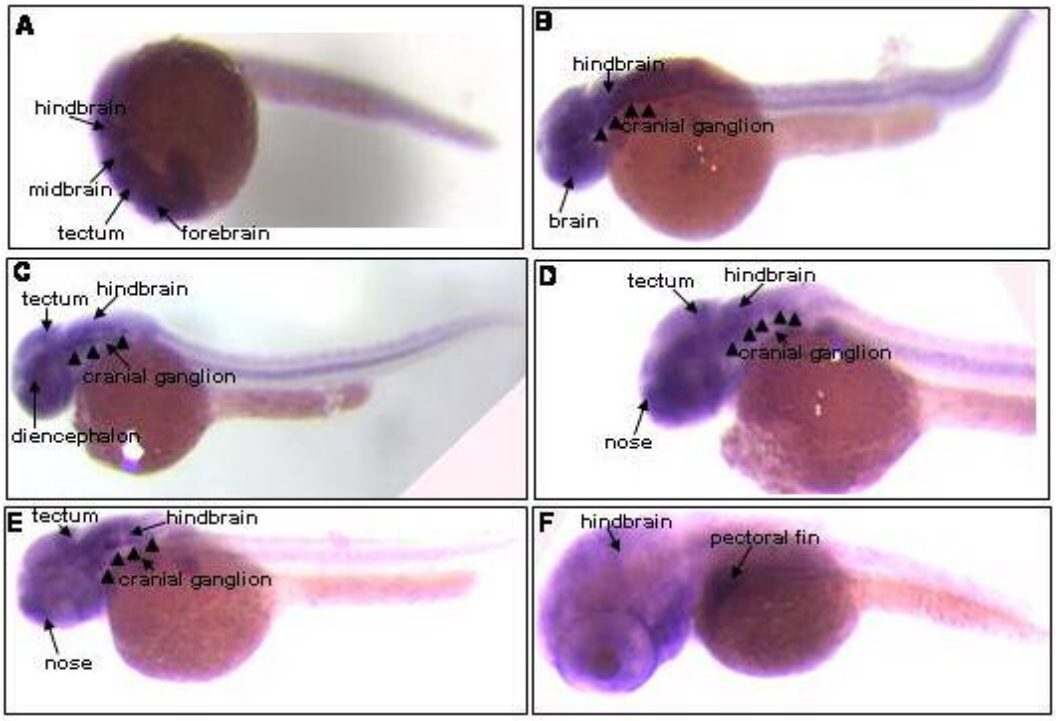


Figure 3.31 Expression pattern of the SLC25A1 gene. The whole mount expression of the gene SLC25A1 shows expression mainly in the brain. At 19-20hpf (A) the expression mainly is in the brain showing specific higher expression in different parts of the brain like forebrain, tectum, midbrain and hindbrain. At 31-35hpf (B and C) the expression mainly is in the brain, tectum, hindbrain and cranial ganglion. At 42-48hpf (D and E) the expression mainly is seen in the nose, tectum, hindbrain and cranial ganglion. At 72hpf (F) the expression mainly is in the hindbrain, retina and pectoral fin.

In the early stages, at 19-20hpf the expression mainly is in the brain showing higher expression in different parts of the brain like forebrain, tectum, midbrain and hindbrain. Expression during stages 31-35hpf mainly is in the brain, tectum, hindbrain and cranial ganglion and at 42-48hpf, the expression mainly is seen in the nose, tectum, hindbrain and cranial ganglion. During the later stages, as seen in 72hpf, the expression mainly is in the hindbrain, retina and pectoral fin.

3.3.2.3 Mitochondrial ribosomal protein L40 (MRPL40)

MRPL40, located at position 17,793,979-17,798,148 in human chromosome 22, has a bipartite nuclear localization signal domain and encodes for a 206 amino acid residue protein that has 51% amino acid identity to its 200 amino acid zebrafish ortholog. This gene has a bipartite nuclear localization signal domain. Earlier studies in E10 mouse using a RT-PCR assay with a probe size of 497bp showed expression in the brain and branchial arches in embryos and in later stages showed expression in brain, heart, lung, kidney, liver (Funke *et al.,* 1998; FANTOM consortium, 2005). Earlier whole mount *in situ* hybridization in mouse showed high expression in brain and the first and second pharyngeal arches (Funke *et al*., 1998; Maynard *et al*., 2003) and our present whole mount *in situ* studies confirm expression in the brain, pectoral fins and branchial arches at 24, 48, and 72hpf. Although the function of this gene is unknown, it is clear that this gene is essential in the embryonic brain and branchial arches development and may be a critical gene resulting in a portion of the velocardiofacial syndrome (VCFS) phenotype.

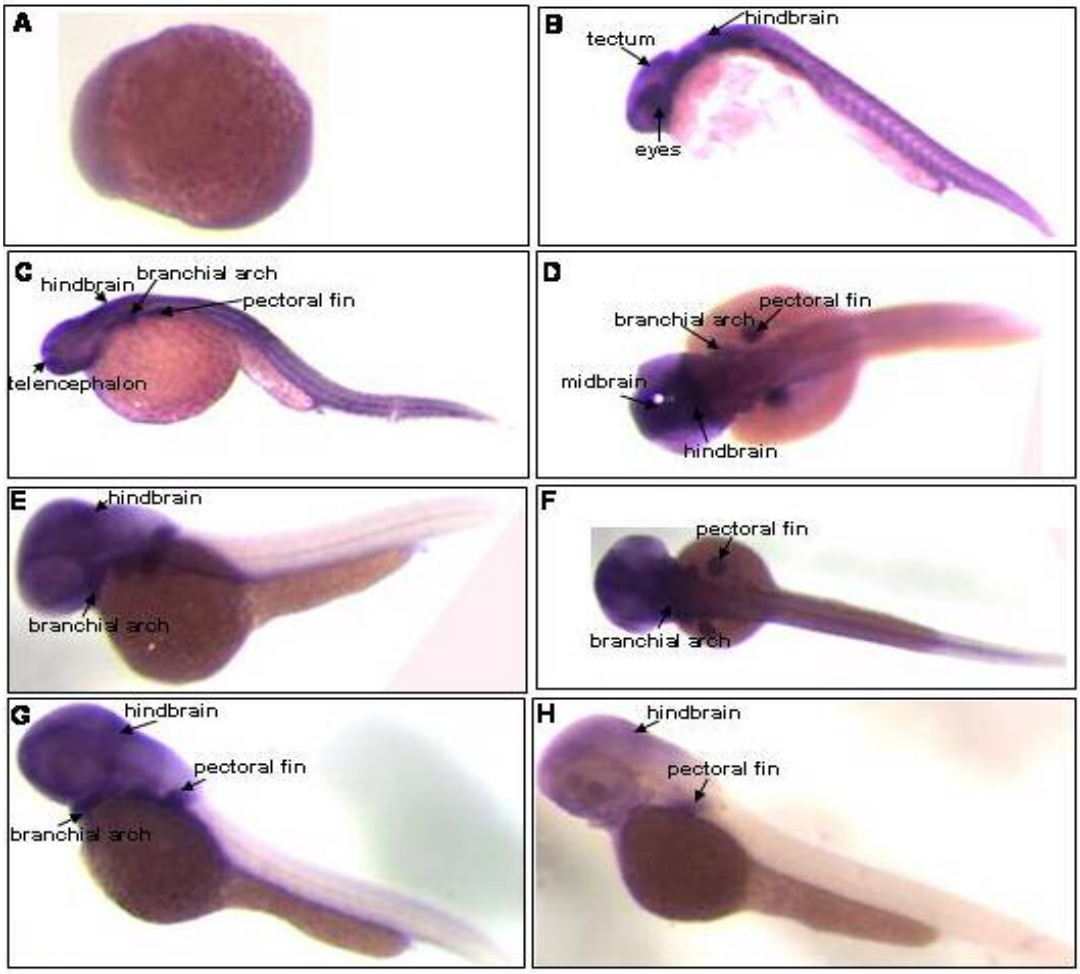


Figure 3.32 Expression pattern of the MRPL40 gene. The whole mount expression pattern of the gene MRPL40 shows expression throughout the body as seen at 12hpf(A), 25hpf (B) and 32hpf C and D (dorsal view) with major expression in the brain, pectoral fin and branchial arches. At 42hpf E and F (dorsal view) and 48hpf (G), the expression mainly is in the hindbrain and branchial arches and also pectoral fins. At 72hpf, the expression is seen only in the hindbrain and pectoral fin.

3.3.2.4 T box-1 (TBX1)

TBX1, a T-box transcription factor, is located at position 18,118,780-18,145,099 on human chromosome 22. This gene encodes for a 398 amino acid protein that has a sequence identity of 76.2% with the 460 amino acid zebrafish ortholog. This gene is a transcription factor and has a T box DNA binding domain involved in heart

formation (Flageman et al, 2005). Northern blot analysis showed Tbx1 expression in mouse during early embryogenesis is in the pharyngeal arches, pouches, and otic vesicle and later during development in the vertebral column and tooth bud (Chieffo *et al*.,1997). RT-PCR studies in mouse (E8.5-12.5) show expression in otic placode, otocyst, head, mandibular, and maxillary process (Raft *et al*., 2004).



Figure 3.33 Expression pattern of the TBX1 gene. The whole mount *in situ* expression of the gene TBX1 is throughout the body but concentrated mainly in the brain. At 12hpf (A), 18hpf (B), 31hpf (C and C-1) the expression is observed in the brain, but at 35hpf (D) the expression also can be seen in the mouth and heart. At 42hpf (E and E-1) the expression is seen in the brain, heart and pectoral fin. At 48hpf (F and F-1) the expression is in the mouth and branchial arches while at 72hpf (G and G-1) the expression mainly is in the mouth, retina and branchial arches, heart and proctodeum.

PCR array expression studies showed foetal and postnatal brain expression of Tbx1 (Maynard *et al.,*2003). Other *in situ* studies have shown expression in the heart, pharyngeal arches, and otic vesicle (Kochilas *et al*.,2003). In our present whole mount

in situ studies expression is seen throughout the body but concentrated mainly in the brain in the early stages but at later stages expression is also seen in the heart. At stages 12, 18, 31hpf the expression is observed mainly in the brain and at 35, 42 and 48 hpf the expression is seen in the mouth, heart, brain and pectoral fin. At 48hpf the expression is also seen in the branchial arches while at 72hpf the expression mainly is in the mouth, retina, branchial arches, heart and proctodeum.

3.3.2.5 Armadillo repeat gene deleted in velocardiofacial syndrome (ARVCF)

ARVCF, a gene located in position 18,331,975-18,378,863 on human chromosome 22 and has a 50% identity with zebrafish ortholog. This gene has a bipartite nuclear localization signal domain as well as Armadillo and HEAT elements that have been implicated in signaling and intracellular transport. Earlier studies showed that this gene is ubiquitously present in foetal and adult tissues (Sirotkin *et al.,* 1997), as well as present in brain, heart, branchial arches, and forelimb buds (Maynard *et al*., 2003). SAGE analysis (CGAP- Strausberg R,L *et al*., 2000; Lash *et al*., 2000) showed expression in the brain and kidney.

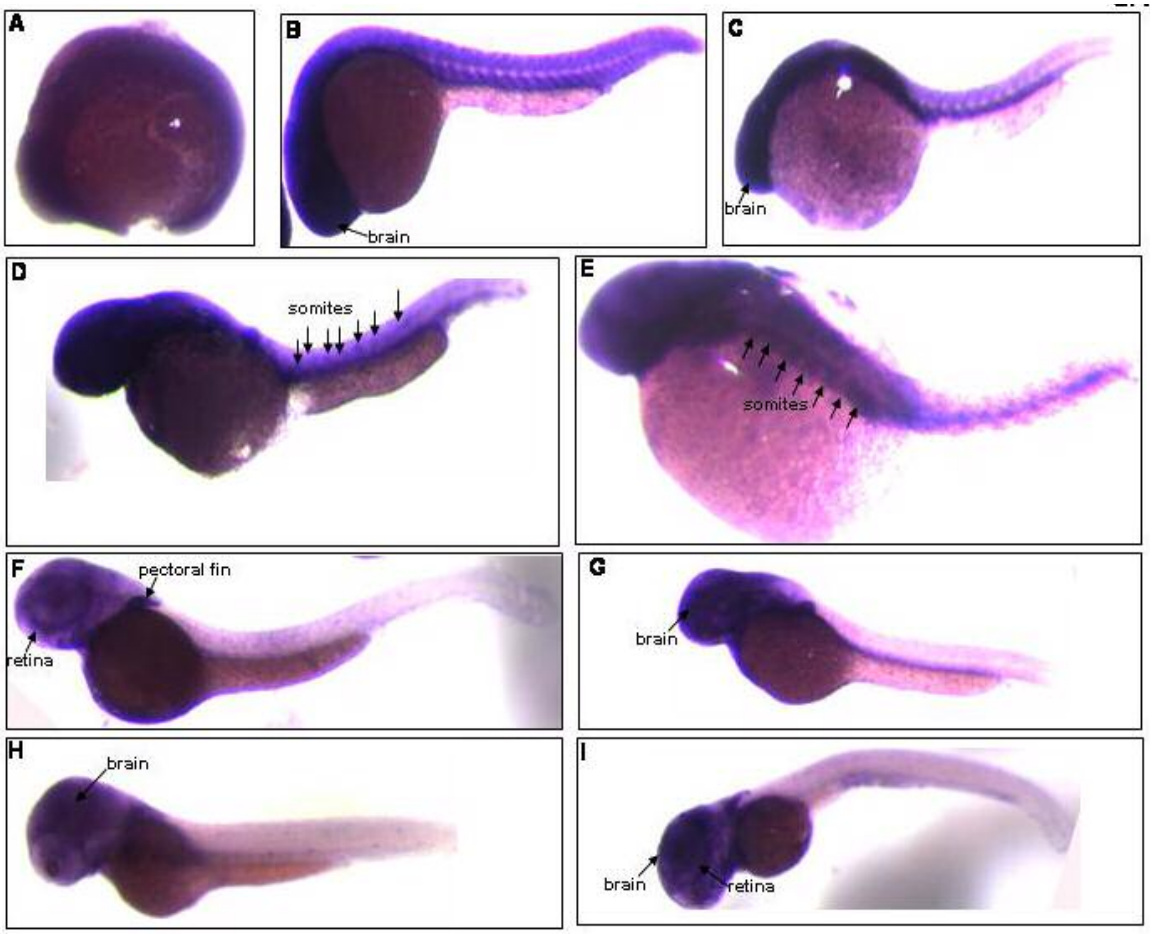


Figure 3.34 Expression pattern of the ARVCF gene. The whole mount expression of the gene ARVCF mainly is present all over the body as seen at 12hpf(A), 24hpf (B), 31hpf (C) and 35hpf (D and E) and also specifically in the somites in 35hpf (D and E). At 42hpf (F) expression is seen in the pectoral fin and retina with a reduction in expression in the brain, slight increase in expression in brain is noticed at 48hpf (G), and again reduction in expression is noticed in 60hpf (H) and 72hpf (I).

Our present whole mount *in situ* studies show expression all over the body as seen at 12 , 24, 31, and 35hpf and also specifically in the somites at 35hpf. Expression at 42hpf is seen mainly in the pectoral fin and retina. The brain expression shows a reduction in expression at 42hpf, slight increase at 48hpf, and again reduction is noticed at later stages at 60 and 72hpf.

3.3.2.6 DiGeorge syndrome critical region gene 8 (DGCR8)

DGCR8, located at the position 18,442,360-18,473,950 and encodes a 777 amino acid protein with 50% identity to the zebrafish ortholog on chromosome 5. DGCR8 has both a double stranded RNA binding and WW40 domain. RNA *in situ* studies show expression in the forebrain, heart, branchial arches and arterial system (Shiohama *et al*., 2003). RT-PCR studies indicated expression in the brain, heart, forelimb bud and branchial arches in the embryonic stages in mouse but in adult stages expression was observed in the brain, heart, spinal chord, liver, kidney, and spleen (Maynard *et al*., 2003; FANTOM Consortium, 2005). Our present whole mount *in situ* zebrafish embryo hybridization studies are consistent with these earlier studies as it showed expression in hindbrain, branchial arches, heart, and pectoral fins and no expression in the heart in later stages. At stages 12 and 22 hpf the expression is throughout the body with slightly higher expression in the eye, midbrain and hindbrain. At stages 31, 42 and 48hpf the expression mainly is in the heart, branchial arches and otic vesicle. But in the later stages at 60 and 72hpf the expression is no longer in the heart but in the branchial arches, otic vesicle and in the pectoral fins.

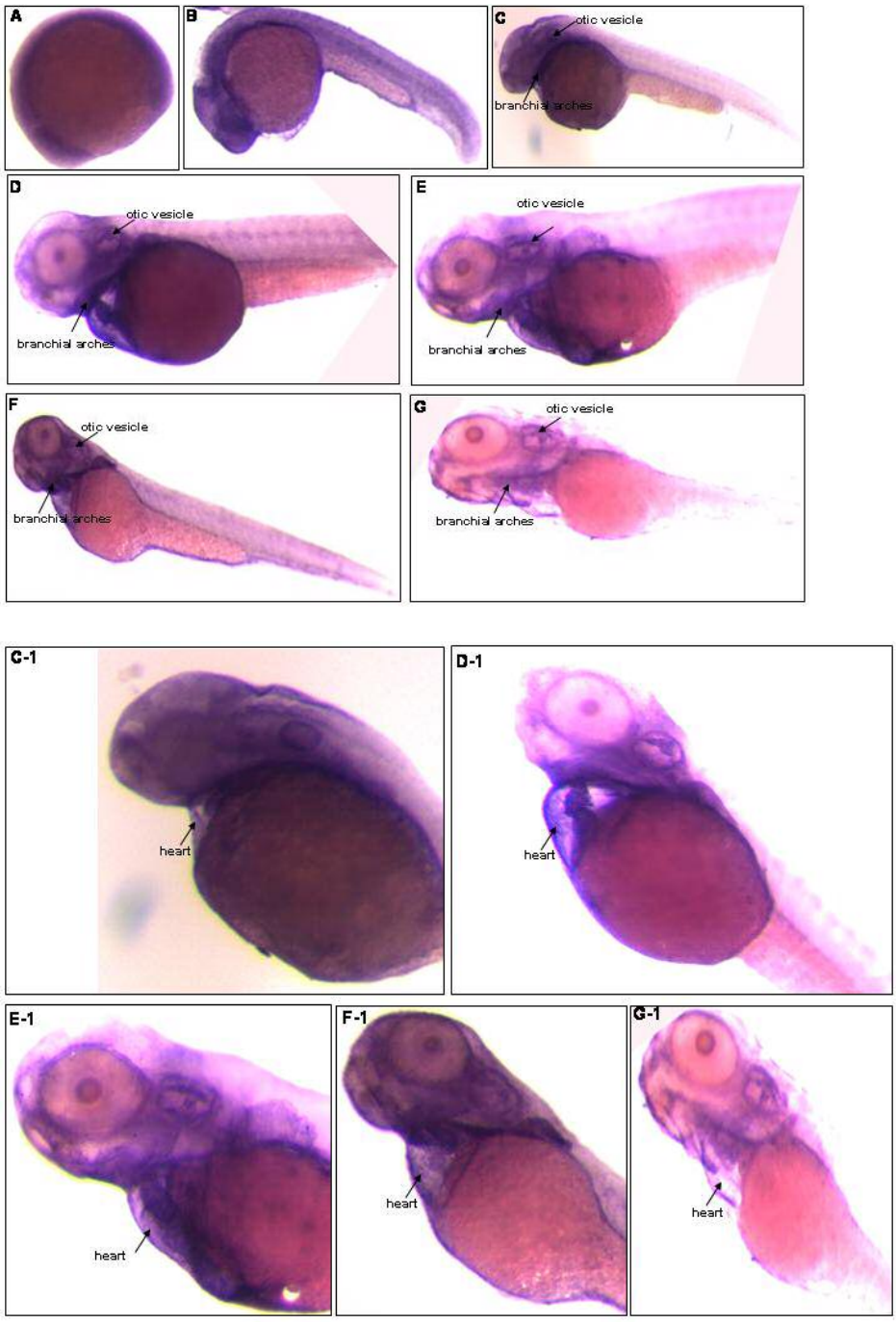


Figure 3.35 Expression pattern of the DGCR8 gene. The whole mount expression pattern of the gene DGCR8 shows expression mainly in the brain and heart during early stages of development but no expression in the heart in later stages. At 12hpf (A) and 22hpf (B) the expression is all over the body with slightly higher expression in the eye, midbrain and hindbrain. At 31hpf (C), 42hpf (D) and 48hpf (E) the expression mainly is in the heart, branchial arches and otic vesicle. But at 60hpf (F) and 72hpf (G) the expression is no longer in the heart but more in the branchial arches and also in the otic vesicle and in the pectoral fins.

3.3.2.7 HpaII tiny fragments locus 9C (Q8IZ69 /HTF9C/ NM_022727)

Q8IZ69/HTF9C/NM_022727 gene located at the position 18,473,952-18,479,258 in human chromosome 22 has 56% amino acid identity with the zebrafish ortholog. The present whole mount *in situ* studies showed expression mainly in the somites as shown in figure 3.30 and 3.31, with 7 above the yolk cell, 10 above the yolk extension, and 13 posterior to the anus which will later gives rise to myotomes and dermatome.



Figure 3.36 Expression pattern of the HTF9C gene. The whole mount expression pattern of gene HTF9C showed expression mainly in the somites and developing myotomes. At 19hpf (A) and 22hpf (B), the expression is very low in the somites. At 25hpf (C) the expression shows slight increase in the somites. The peak of expression in the somites and myotomes mainly is at 48hpf as seen in D (side view), E (dorsal view), E-1 and E-2. At 72hpf (F), there is no expression seen in the mytotomes and slight expression is seen in the retina.

The expression of this gene could be seen at 20-24hpf, higher expression at 48hpf and no expression at 72hpf. Numerous studies have shown that many transcription factors and signaling factors are involved in the formation of somites (van Eeden *et al.,* 1996; Heather *et al*., 2000). This gene has protein domains similar to the SAM (S-adenosylmethionine-dependent methyltransferase), Uracil5-methyltransferase, and RNA binding domains. The nucleotide binding domain of the methyltranferase makes up the major domain that contributes to DNA methylation, that effects transcription factor binding and is an important epigenetic factor (Martin *et al.,* 1999).

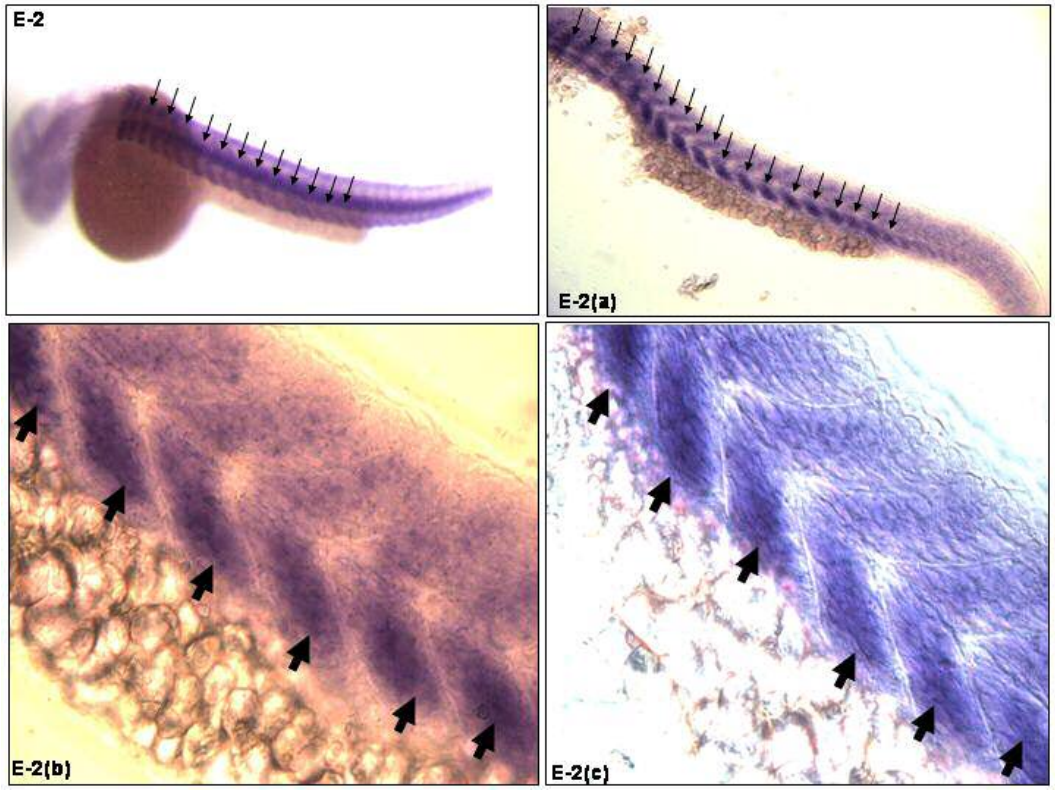
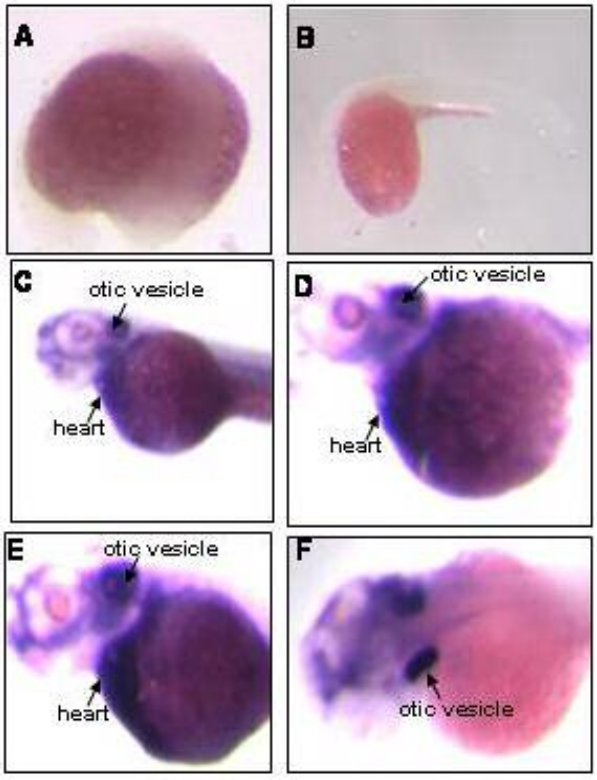


Figure 3.37 Expression in the somites and developing myotomes. Detailed and enlarged view of expression in the somites and developing myotomes are shown as marked by arrows.

Since DNA methyltransferases mRNA are present in differentiating somites, this gene likely is involved in similar processes in early developmental stages and depletes at later stages in zebrafish development.

3.3.2.8 Zinc finger, DHHC-type containing 8 (ZDHHC8)

ZDHHC8, located at the position 18,493,918-18,508,512 in human chromosome 22, has a putative transmembrane palmitoyltransferase zinc finger DHHC domain containing protein 8 precursor. This gene encodes for a 765 amino acid protein that has 56% amino acid identity to its 716 amino acid zebrafish ortholog. RT-PCR studies showed that this gene in mouse (E 10.5) is present in the branchial arch, heart, forebrain, forelimb, frontonasal region, and arterial system (Maynard *et al*., 2003).



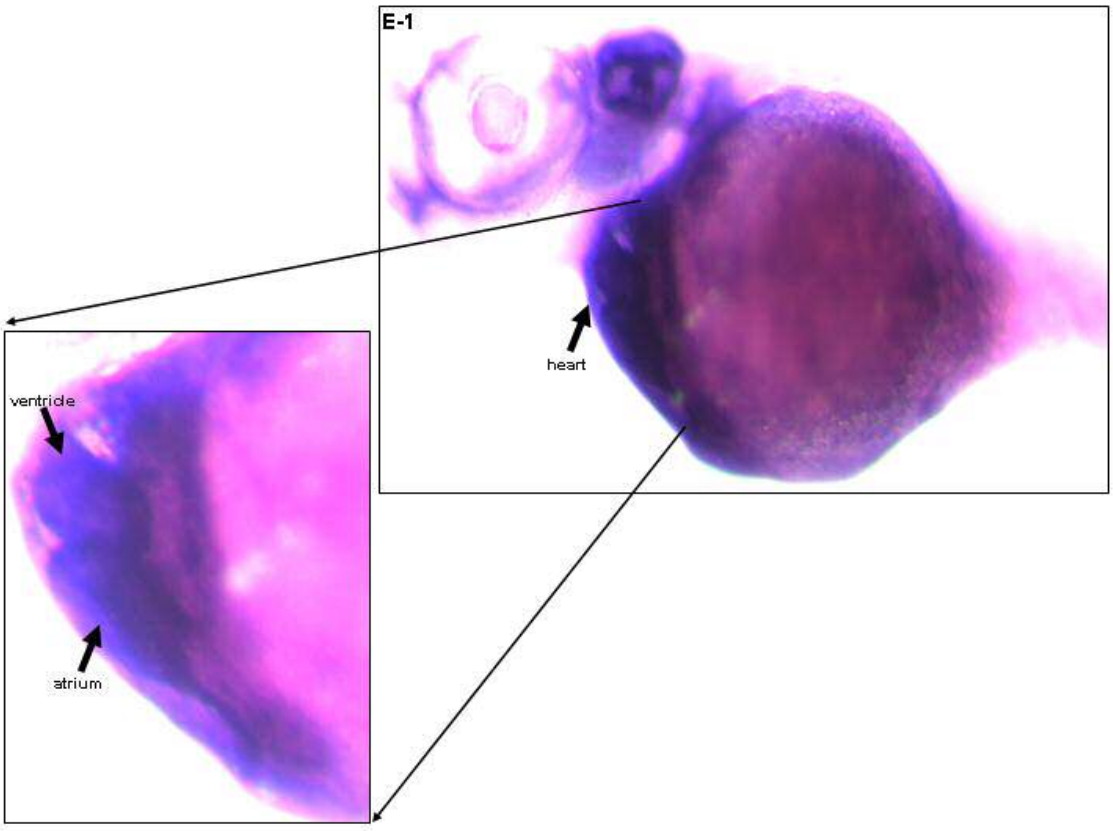


Figure 3.38 Expression pattern of the ZDHHC8 gene. The expression pattern of the gene ZDHHC8 shows expression in the otic vesicle and heart showing expression in the ventricle and atrium**.** At 12hpf (A) and 22hpf (B) the expression of this gene is not seen. But at stages from 35-48hpf as seen in C, D, E the expression mainly is at a higher level in the heart and the otic vesicle. The enlarged picture of the heart (E-1) shows expression in the pericardial cavity, ventricle and atrium. At 72hpf (F) the expression mainly is in the otic vesicle relative to the basal expression in the brain.

SAGE and CGAP studies showed expression in the brain, spinal chord, and heart (Strausberg R,L *et al*., 2000; Lash *et al*., 2000) while other cDNA studies showed expression in the brain (Nagase *et al*., 1999). Our present zebrafish whole mount *in situ* hybridization studies confirm expression in the brain, ear, distal part of tail, branchial arches, and heart showing expression in the ventricle and atrium. At stages 12 and 22 hpf the expression of this gene is not seen. But at stages from 35-48hpf the expression mainly is at a higher level in the heart and the otic vesicle. At 72hpf, a basal expression is seen in the brain, with higher expression in the otic vesicle.

3.3. 2. 9 ENSG00000182364

Gene ENSG00000182364, located at position 18,758,069-18,802,360 on human chromosome 22, encodes for a 256 amino acid protein with 33% amino acid identity to zebrafish chromosome 2 ortholog.

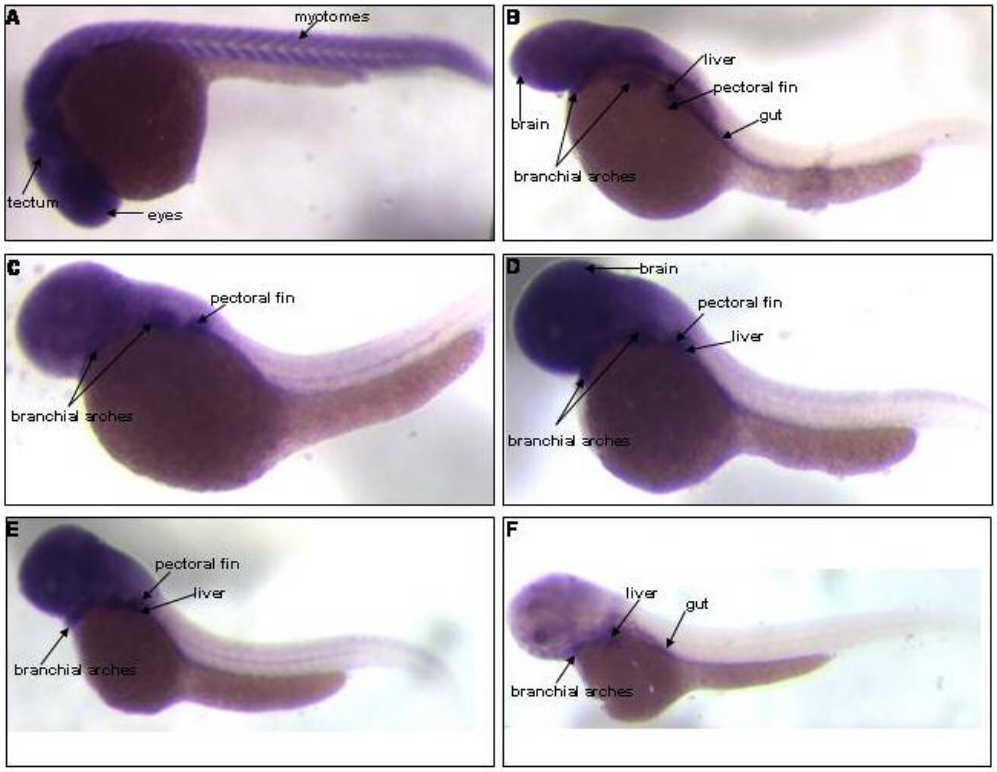


Figure 3.39 Expression pattern of the ENSG00000182364 gene. (A) At 24hpf the expression is throughout the body with slightly higher and defined expression in the brain mainly eyes, tectum, hindbrain and myotomes. (B) and (C) at 31-35hpf shows concentration of expression mainly in the brain followed by branchial arches, pectoral fins, liver (primodium) and gut. (D) and (E) at 42-48hpf shows increased expression in brain and branchial arches and also expression in pectoral fins and liver. (F) At stage 72hpf, a basal level expression is noticed in the brain, with reduced expression in branchial arches, liver and gut

The gene is expressed in all normal human tissues based on Genenote analysis (Shmueli *et al*., 2003) but with expression mainly in the brain. Our present whole mount *in situ* studies indicate that at 24hpf the expression is throughout the body with slightly higher and defined expression in the brain mainly eyes, tectum, hindbrain and myotomes. At 31-35hpf the expression mainly is in the brain followed by branchial

arches, pectoral fins, liver (primodium) and gut. At 42-48hpf there is increased expression in brain and branchial arches and also expression in pectoral fins and liver. At 72hpf, a basal level expression is noticed in the brain, with reduced expression in branchial arches, liver and gut.

3.3.2.10 ENSG00000183099

Gene ENSG00000183099, located at position 18,935,803-18,940,755 on human chromosome 22, encodes for a 256 amino acid protein with 61% amino acid identity to the 489 amino acid zebrafish chromosome 16 ortholog and encodes a gamma-glutamyl transpeptidase domain. Whole genome expression profiles in normal human tissues based on Genenote analysis (Shmueli *et al*., 2003) showed expression in all tissues with notable expression in the kidney, prostate, liver, lung, and pancreas. Our present whole mount *in situ* studies show expression throughout the embryo at 24hpf mainly in the brain region and also the myotomes and proctodeum but more defined expression at 31hpf in the pronephric duct, liver, branchial arches, brain, somites, myotomes. At 42hpf expression is seen in the brain (more in the hindbrain), branchial arches, liver and pectoral fins and at 48hpf there is an increase in expression in comparison to 42hpf. A gradual decrease in expression at 72hpf in the zebrafish embryos. This expression pattern indicates that this gene is important in early likely is involved in the formation of the kidney, lungs and liver.

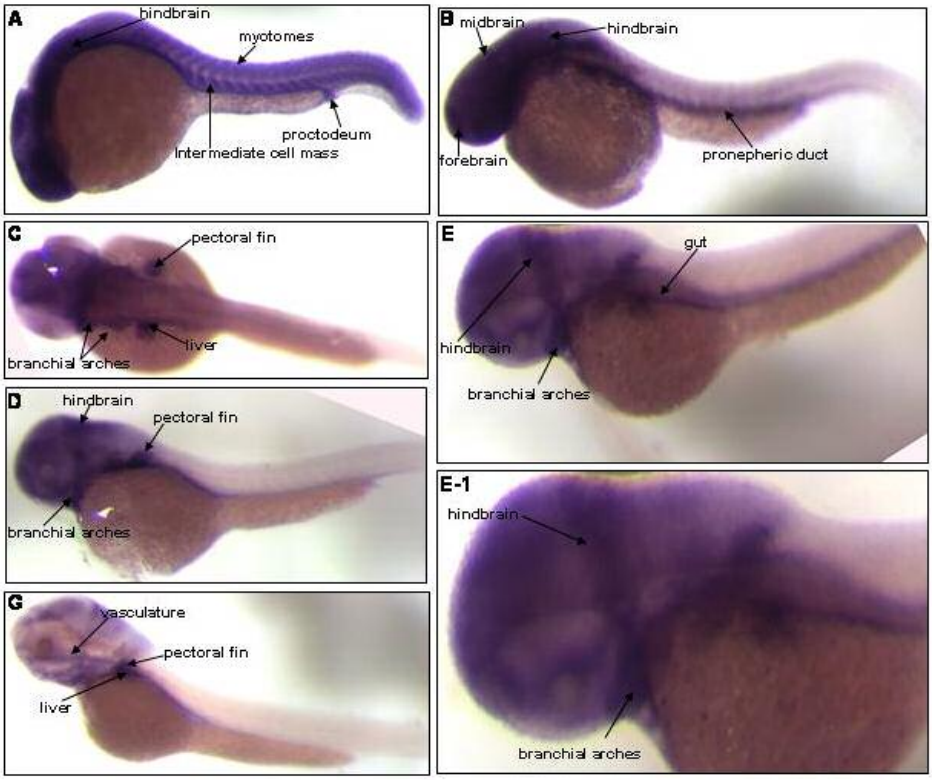


Figure 3.40 Expression pattern of the ENSG00000183099 gene. At 24hpf (A), the expression is noticed mainly in the brain region and also the myotomes and proctodeum. At 31hpf (B) localized and increased expression is observed in the brain and pronepheric duct. C (dorsal view) and D (side view) at 42hpf shows expression in brain (more in the hindbrain), branchial arches, liver and pectoral fins. At 48hpf E and E-1 show increase in expression in the hindbrain, branchial arches, pectoral fin, liver and gut in comparison to 42hpf. At 72hpf (F), expression is observed in blood vessel, pectoral fin and liver.

3.3.2.11 Scavenger receptor class F, member 2 (SCARF2)

SCARF2, located at position 19,103,429-19,116,700 in human chromosome 22, is known to have both a protein binding and a receptor activity. It contains EGF-like, Laminin type EGF-like, and ATP/GTP binding site motifs as well as A (P loop) and Dopamine D4 receptor domains that are known to have interactions with growth factors. This gene encodes for a 868 amino acid protein with 52% amino acid identity with the 854 amino acid zebrafish ortholog. RT-PCR in mouse indicated expression in the eyes, lungs, and urogenital organs (Lund *et al*., 2000; FANTOM Consortium 2005).

Northern blot analysis showed high expression in heart, and expression in the lung, ovary, and placenta (Ishii *et al*., 2002).

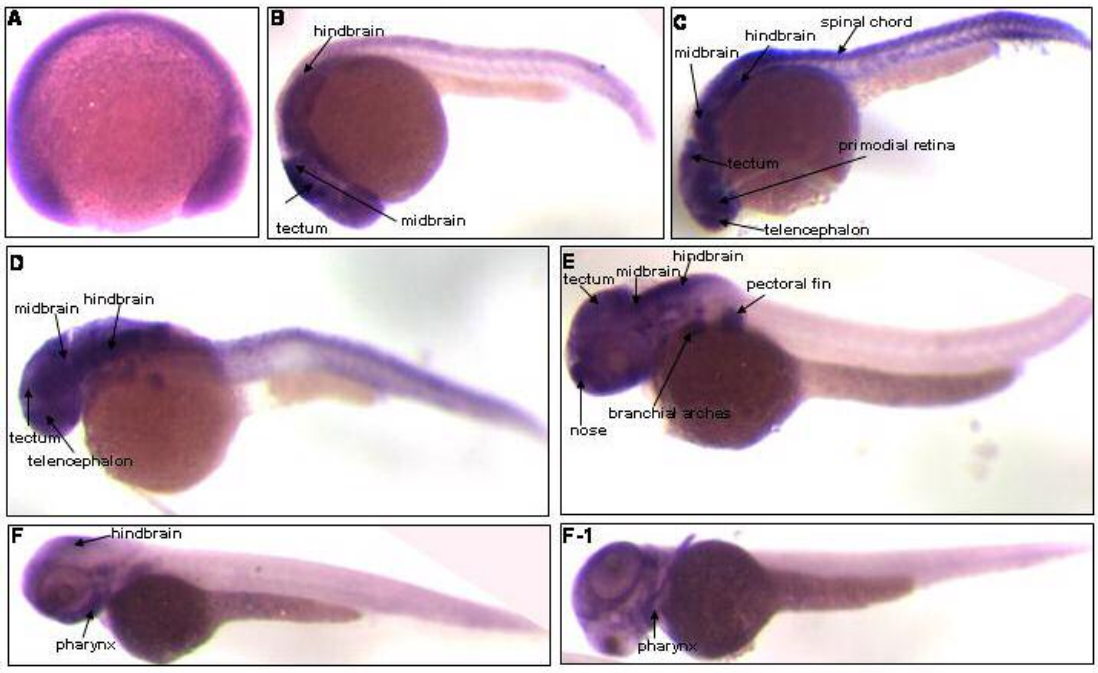


Figure 3.41 Expression pattern of the SCARF2 gene. The whole mount expression pattern shows expression mainly in the brain and in the later stages in the pharynx. At 12hpf (A) the expression is at its basal level and throughout the body. At 24hpf (B) the expression mainly is in the midbrain, tectum. At 32hpf (C) the expression mainly is in the retina, tectum, telencephalon, midbrain and hindbrain. At 35hpf (D) the expression is in the telencephalon, tectum, midbrain and hindbrain. At 48hpf (E) the expression is more in the anterior half of the body especially in the nose, midbrain, hindbrain, tectum, branchial arches, otic vesicle and pectoral fins. At 72hpf (F) and (ventral view, F-1) the expression mainly is in the hindbrain and pharynx.

Our present whole mount *in situ* studies also show high expression mainly in the brain and in the later stages in the pharynx. At 12hpf a basal level of expression is seen all over the body. At 24hpf the expression mainly is in the midbrain, tectum and at 32hpf the expression mainly is in the retina, tectum, telencephalon, midbrain and hindbrain. At 35hpf the expression is also in the telencephalon, tectum, midbrain and hindbrain. But at 48hpf the expression is more in the anterior half of the body

especially in the nose, midbrain, hindbrain, tectum, branchial arches, otic vesicle and pectoral fins. By the stage of 72hpf the expression mainly is restricted in the hindbrain and pharynx.

3.3.2.12 NM_032775 (ENSG00000185214 /Q96B68/NP_116164.2/KLHL22 (Kelch-like 22))

NM_032775, located at position 19,120,360-19,174,676 in human chromosome 22 encodes for a 634 amino acid protein that is 61% identical to the 585 amino acid zebrafish ortholog and contains kelch-like and BTB /POZ (BR-C, ttk and bab) (Pox virus and Zinc finger) domains. Kelch- like domains seem to occur in genes expressed in both skeletal muscle and heart (Wu *et al*., 2004) although this particular gene has no known previous expression pattern. Our present whole mount *in situ* hybridization studies showed basal expression all over the body mainly in the epidermis and photoreceptors in the eye in the early stages of development.

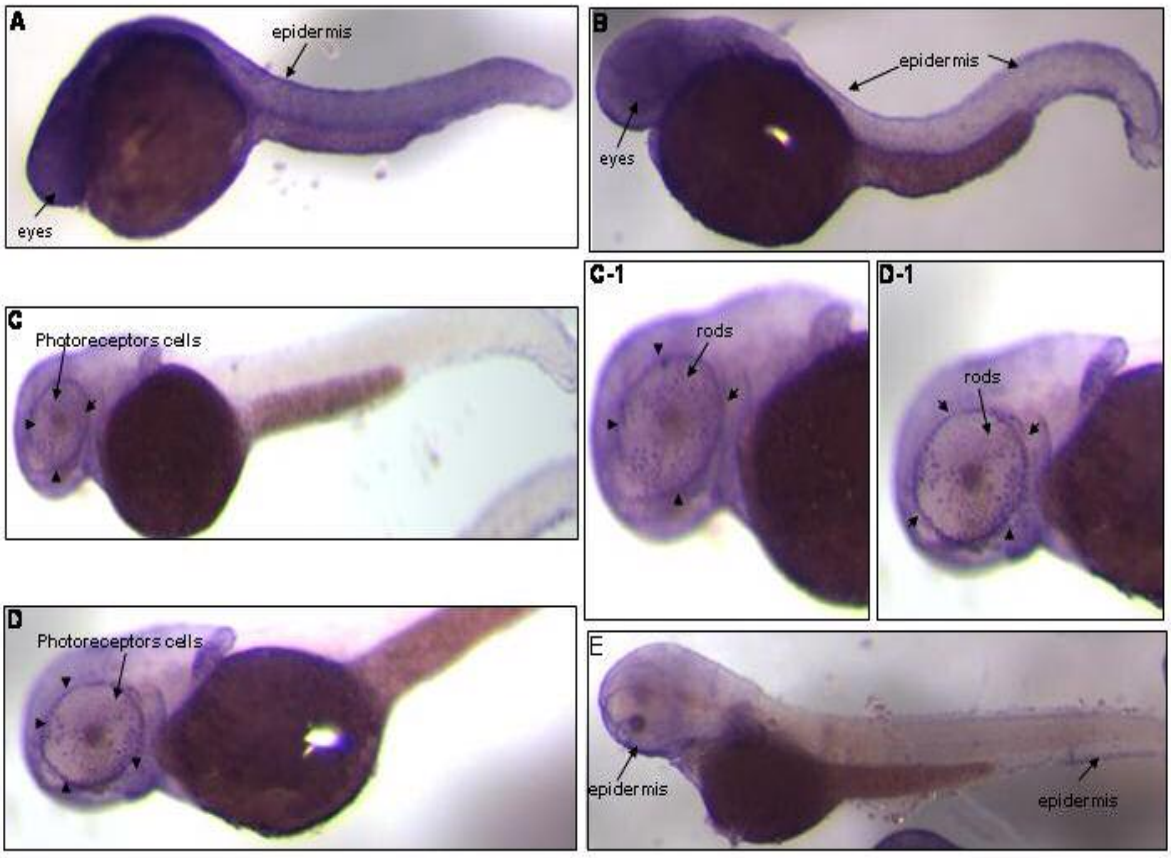


Figure 3.42 Expression of the NM_032775/ENSG00000185214 gene. The whole mount expression of the ENSG00000185214 gene shows expression all over the body mainly in the epidermis especially in the eye in the early stages of development. At 22hpf (A) and 32pf (B) the expression is all over the body and the epidermal layer. At 48hpf (C) and 60hpf (D) the expression mainly is in the photoreceptor cells of the eye (mainly rods) and is more concentrated at 48hpf than at 60hpf. At 72hpf (E) the expression in the photoreceptor cells are very less compared to the earlier stages.

At 22 and 32 hpf the expression is observed throughout the body and the epidermal layer. At 48 and 60 hpf expression also is seen in the photoreceptors in the eyes, mainly in the rods. Experiments have shown that rods accumulate faster as a ventral patch initially and rods outside of this region accumulate by slowly radiating unevenly from the center of the retina (Raymond et al., 1995).

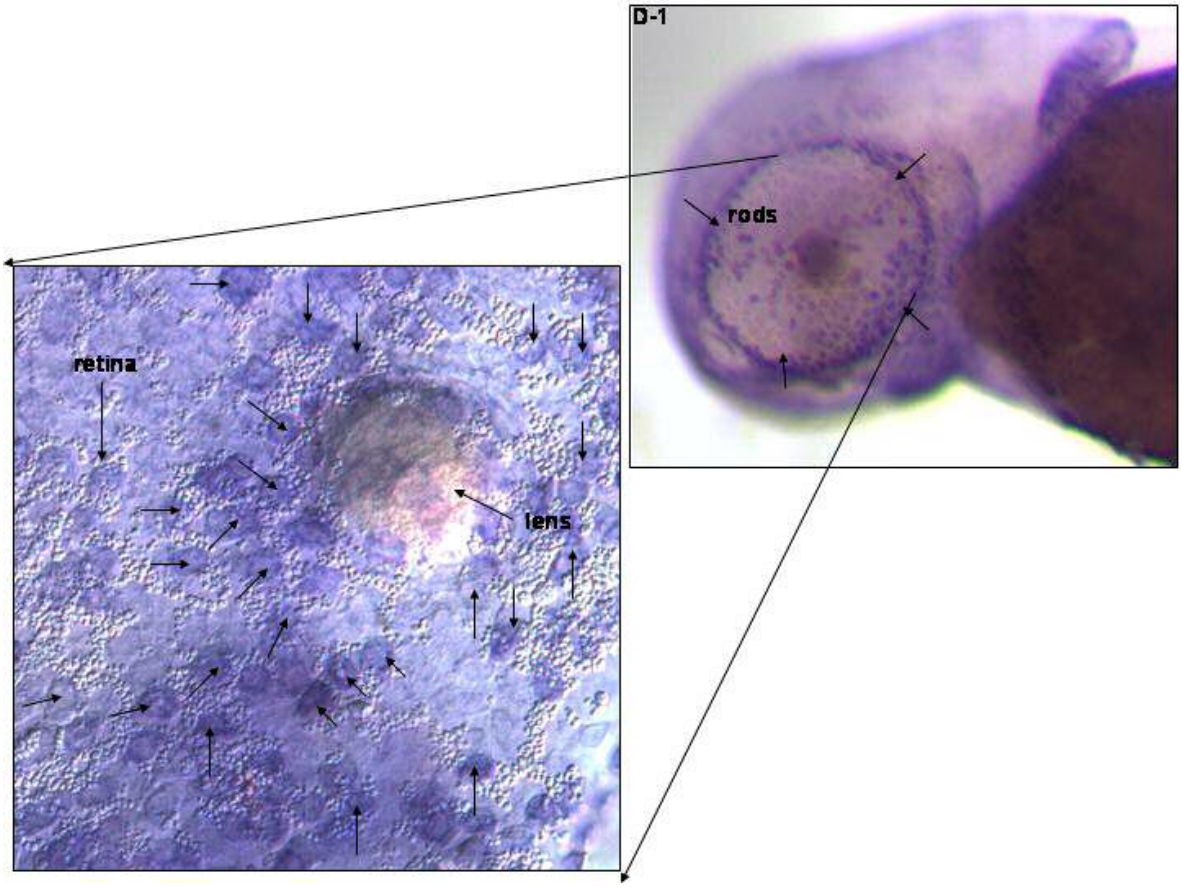


Figure 3.43 Enlarged view of the eye photoreceptor cells . The enlarged view of the eye clearly showing expression in the photoreceptor cells (mainly rods) of the retina and no expression in the lens.

A similar distribution is observed in the expression pattern as shown in figure 3.43. Thus, as proposed for Drosophila (Cagan and Zipursky, 1992), this process likely requires cell-cell communication tempting the speculation that this NM_032775/ENSG00000185214 gene may function as a signaling molecule or be involved as one of the signal transduction components in zebrafish and may be even humans. At later stages (72hpf) the expression in the photoreceptors and epidermis is very less compared to the earlier stages.

3.3.2.13 Positive cofactor 2, multiprotein complex) glutamine/Q-rich-associated protein (PCQAP)

PCQAP, is located at position 19,186,443-19,266,458 in human chromosome 22. This gene encodes for a 788 amino acid protein that has a 60% identity to the 802 amino acid protein encoded by the zebrafish chromosome 21 ortholog. This gene has three significant domains, namely a proline rich domain, an extensin like protein, and a bipartite nuclear localization signal domain.

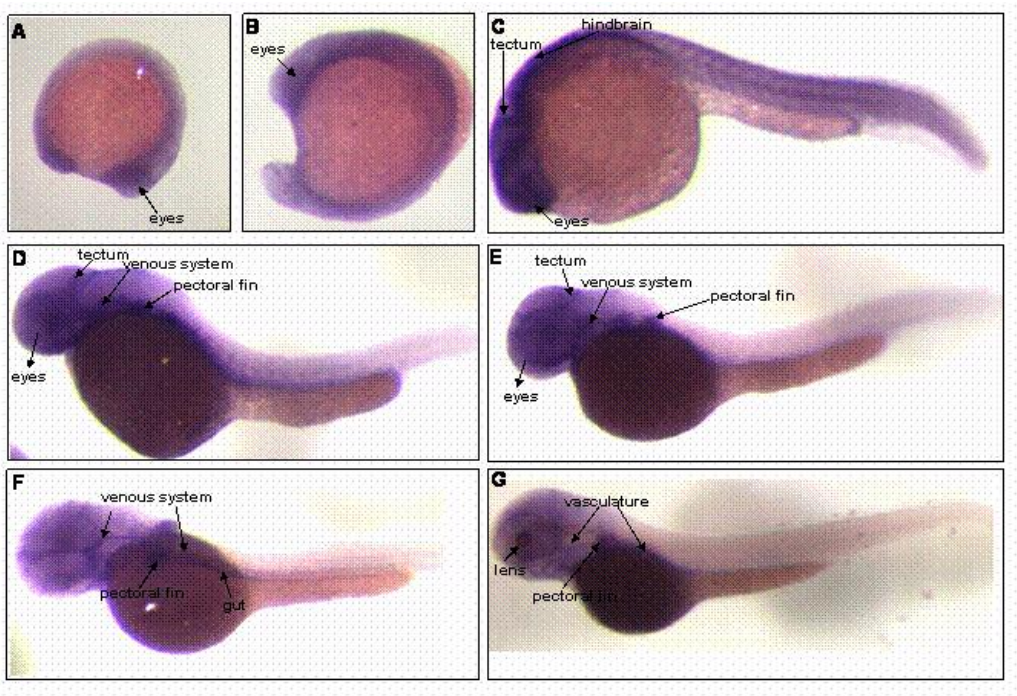


Figure 3.44 Expression pattern of the PCQAP gene. The whole mount expression pattern of the gene PCQAP shows expression mainly throughout the body, but especially in the venous system, gut and pectoral fin. At 12hpf (A) and 16hpf (B) there is basal expression throughout the body but a relatively higher expression is found in the eye and anterior region of the embryo. At 22hpf (C) the expression mainly is in the eyes, midbrain and hindbrain. At 32hpf (D), 48hpf (E) and 60hpf (E) the expression mainly is in the eyes, tectum, venous system, gut and pectoral fin. At 72hpf (G) the expression mainly is in the borders of the lens, vasculature and pectoral fin.

Although mouse *in situ* hybridization experiments have shown almost ubiquitous expression, high expression levels were observed in mouse embryonic frontonasal region, limbs, and pharyngeal arches (Berti *et al*., 2001). Unigene and

SAGE data also showed high expression in the spinal chord and brain (Schuler *et al*., 1997; Pontius *et al*., 2003; Strausberg R, L *et al*., 2000). Our present whole mount *in situ* hybridization studies show expression pattern mainly throughout the body, but especially in the venous system, gut and pectoral fin. At 12hpf-16hpf there is basal expression throughout the body but a relatively higher expression is found in the eye and anterior region of the embryo. The expression mainly is in the eyes, midbrain and hindbrain at 22hpf. At stages 32, 48 and 60 hpf, the expression mainly is in the eyes, tectum, venous system, gut and pectoral fin. In the later stages at 72hpf, the expression mainly is in the borders of the lens, vasculature and pectoral fin.

3.3.2.14 Phosphatidylinositol 4-kinase, catalytic, alpha polypeptide (PIK4CA)

PIK4CA, a member of the phosphatidylinositol 4-kinase family, is located at position 19,386,545-19,517,555 in human chromosome 22. This gene encodes for a 2044 amino acid protein that has a 70% identity to the 1427 amino acid zebrafish ortholog. This protein encodes for a basic leucine zipper transcription factor domain, a phosphatidylinositol 3- and 4-kinase catalytic domain and a bipartite nuclear localization domain.

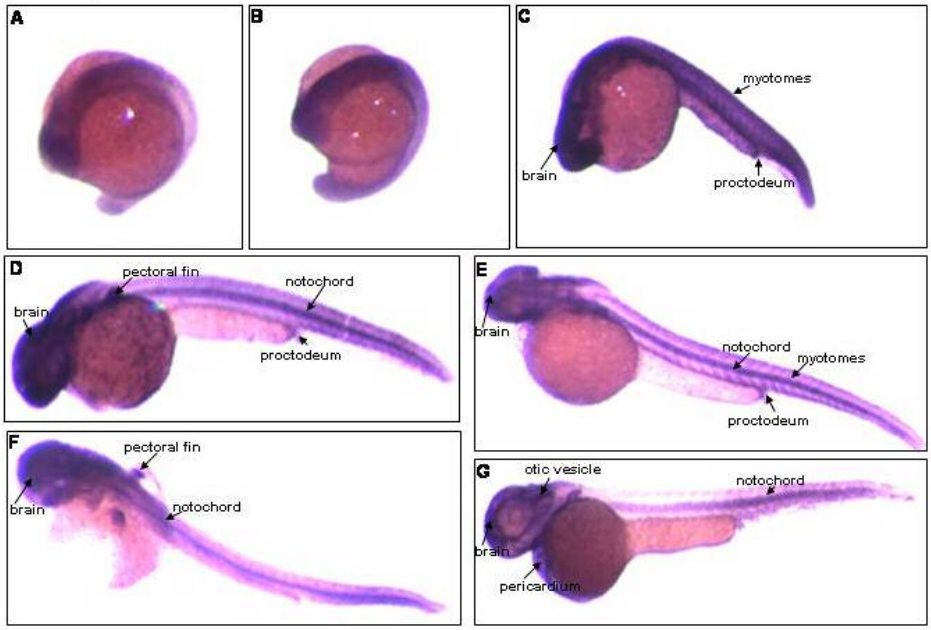


Figure 3.45 Expression pattern of the PIK4CA gene. The whole mount expression pattern of PIK4CA gene shows expression throughout the body with higher levels of expression in the brain. At 16hpf(A) and 18hpf(B) shows a higher expression in the presumptive brain region compared to rest of the body. At 22hpf (C) the expression mainly is in the brain and also in the proctodeum and myotomes. At 32hpf (D) the expression in mainly concentrated in the brain, pectoral fins and also in the notochord and protcodeum. At 42-48hpf (E and F) the expression mainly is in the brain, notochord, and proctodeum. The expression is reduced with respect to earlier stages. At 60-72hpf (G), the expression in the brain is still reduced and expression can also be seen in the otic vesicle, pericardium and notochord.

Earlier Northern blot studies have reported high expression in the brain and moderate expression in the placenta (Wong and Cantley,1994). Whole mount in situ studies in rodents showed high expression in fetal brains compared to adult brains (Nakagawa et al., 1996). Our present whole mount *in situ* studies in zebrafish confirms the expression in the brain and indicates basal levels of expression in the pectoral fins. At stages 16hpf-18hpf shows a higher expression in the presumptive brain region compared to rest of the body. At 22hpf the expression mainly is in the brain and also in the proctodeum and myotomes. At 32hpf the expression in mainly concentrated in the brain, pectoral fins and also in the notochord and protcodeum. At 42-48hpf the expression mainly is in the brain, notochord, and proctodeum and in comparison to

earlier stages, the expression is reduced. At 60-72hpf, expression can also be seen in the otic vesicle, pericardium and notochord in addition to the reduced expression in the brain. Thus this **gene therefore likely is important in brain development, and it may be related to various reported brain disorders (Saito *et al*., 2003).**

3.3.2.15 v-crk sarcoma virus CT10 oncogene homolog (avian)-like gene (CRKL)

CRKL gene, located at position 19,596,268-19,632,588 on human chromosome 22, encodes for a 303 amino acid protein with 82.1% amino acid identity to the 305 amino acid zebrafish chromosome 21 ortholog. Each gene consists of three exons and it contains SH2-SH3-SH3 domains similar to those observed in tyrosine kinases (de Jong, 1995; Uemura, 1997). RT-PCR studies in mouse showed high expression in the embryonic brain (mouse E10-E18 stages) compared to expression in muscles, kidney, liver, while in adults high levels of expression were observed in the brain and in several other tissues (Maynard *et al*., 2003; de Jong *et al*., 1995). A null mutation of this gene in mice caused craniofacial and cardiac defects (Guris *et al*., 2001).

Our present whole mount *in situ* hybridization studies showed expression all over the body during early stages but during late stage is restricted to the otic vesicle and pectoral fin. At stages 22-24hpf the expression is all over the body especially in the brain, notochord and choroid neural hinge and tail bud.

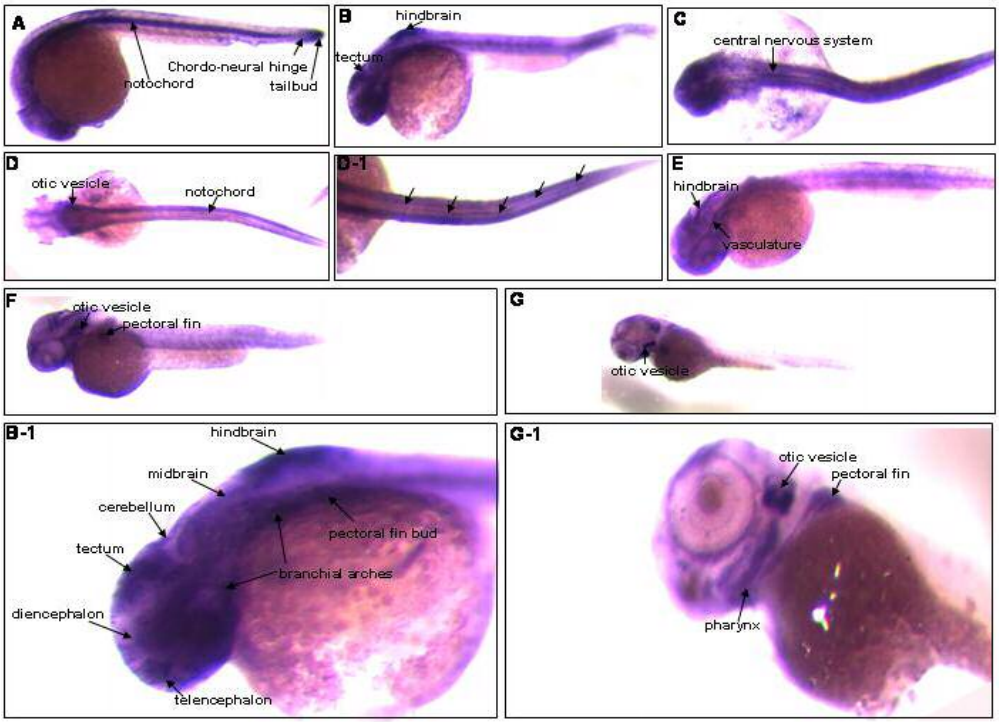


Figure 3.46 Expression for pattern of the CRKL gene. The whole mount expression for CRKL shows expression throughout the body during early stages but during late stages is restricted to the otic vesicle and pectoral fin**.** (A) At 22-24hpf the expression is throughout especially in the brain, notochord and choroid neural hinge and tail bud. At 32hpf B (whole side view), C (dorsal view) and B-1 (side enlarged view) the expression also is throughout the body but higher in the midbrain, cerebellum, dienceplalon, telencephalon, hindbrain, tectum, branchial arches and pectoral fin bud. At 35hpf D (dorsal view), D-1 (enlarged view of the tail) showing the expression in the notochord. At 42hpf (G) the expression mainly is in the hindbrain and vasculature. At 48hpf (F) the expression mainly is in the otic vesicle and pectoral fin. At 72hpf (G and G-1 side view) the expression mainly is in the otic vesicle, pectoral fin and pharynx.

At 32hpf, the expression is also all over the body but higher in midbrain, cerebellum, dienceplalon, telencephalon, hindbrain, tectum, branchial arches and pectoral fin bud. At 35hpf the expression is in the notochord. At 42hpf the expression

mainly is in the hindbrain and vasculature. At 48hpf the expression mainly is in the otic vesicle and pectoral fin. At 72hpf the expression mainly is in the otic vesicle, pectoral fin and pharynx.

3.3.3 PHD finger protein 5A (Phf5a)

PHF5A, a PHD finger like domain containing protein, is present at position 40,180,222-40,189,191 in human chromosome 22q13.2, and has a 100% identity with the zebrafish chromosome 12 ortholog. The VISTA figure 3.47 and the pipmaker output in figure 3.48 show the high conservation of all four exons in PHF5a gene between human and zebrafish.

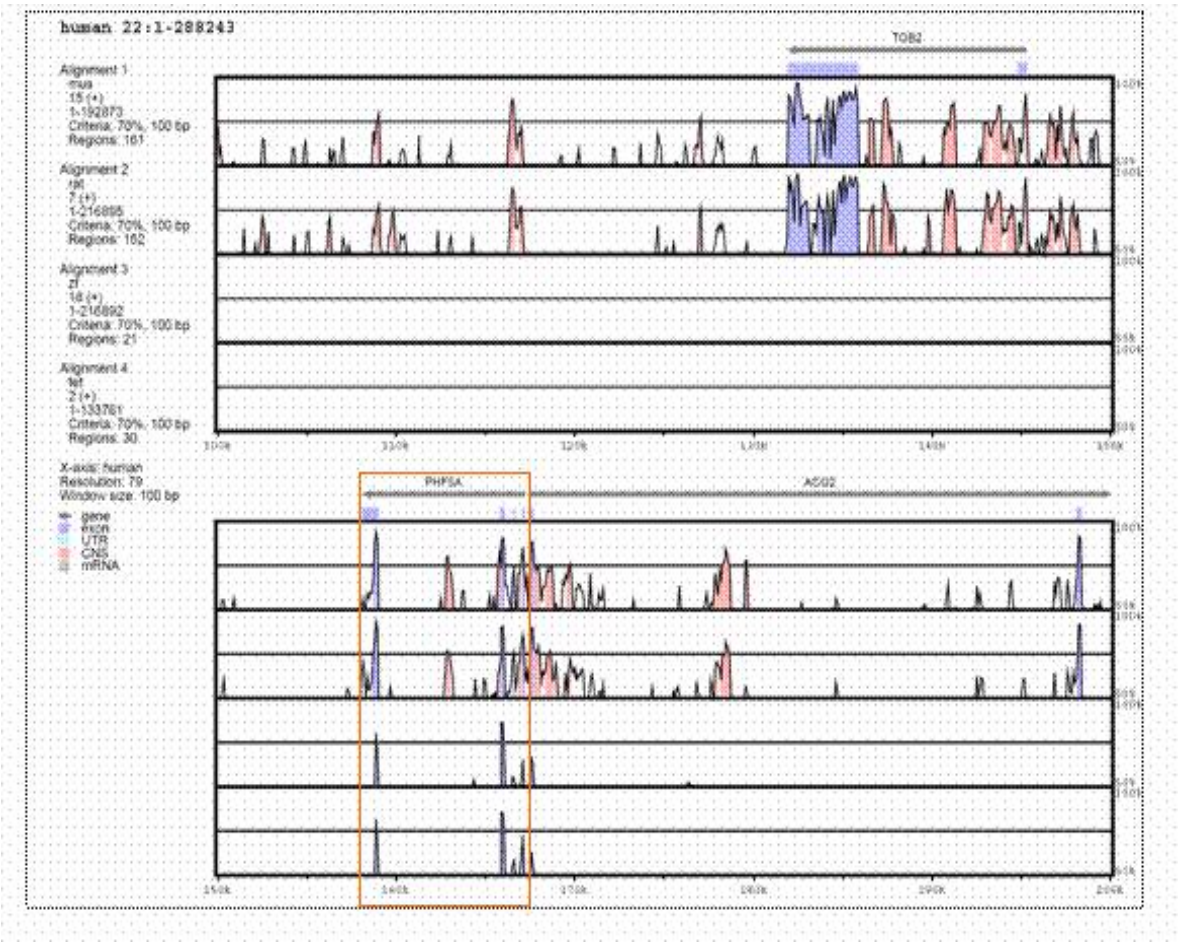


Figure 3.47 VISTA output of Phf5a gene. VISTA output shows the conservation of all four exons of Phf5a gene in all species compared.

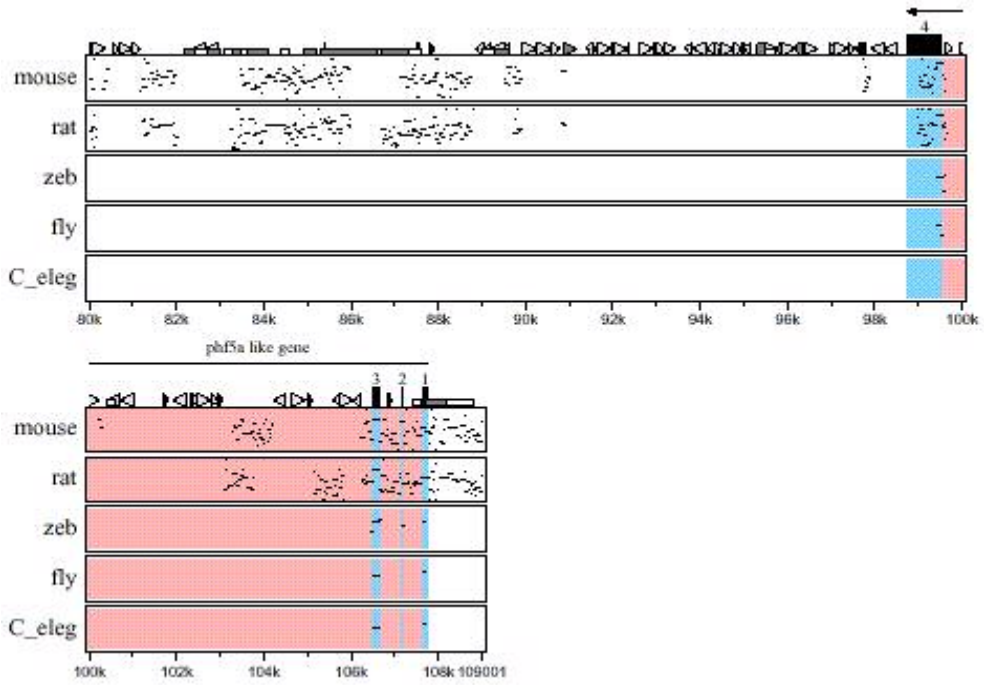


Figure 3.48 PIP output of Phf5a gene. The PIP output also shows conservation of all four exons of Phf5a gene in the different species**.**

Earlier studies in *C. elegans* (Trappe *et al*., 2002) and mouse (Trappe *et al*., 2002) indicate that this gene is essential for cell viability and morphogenetic development. Our whole mount *in situ* studies using a 280bp RNA probe in zebrafish embryos showed expression in the brain at 12hpf, 24hpf, 48hpf, with high levels of expression at 55-72hpf, and slight reduction at 96, 120, and 168hpf as seen in figure 3.49 (A, B, C, D, E, I, J, K). To further understand the importance of this gene in development, whole mount *in situ* hybridization was carried out in a recessive mutant hi1373 (Golling *et al*., 2002). This mutant has a proviral insertion in intron 1 of the PHF5A gene, causing the formation of a truncated PHF5A protein. The mutants showed both phenotypic and expression changes in comparison to the wild type. The mutants have a curved body and constricted yolk sac extension as seen in figure 3.49

(F, G, and H) compared to the wild type embryos and die within 3 to 6 days of age. Expression results in mutants with reference to wild type showed extremely reduced expression at 96, 120, and 168hpf.

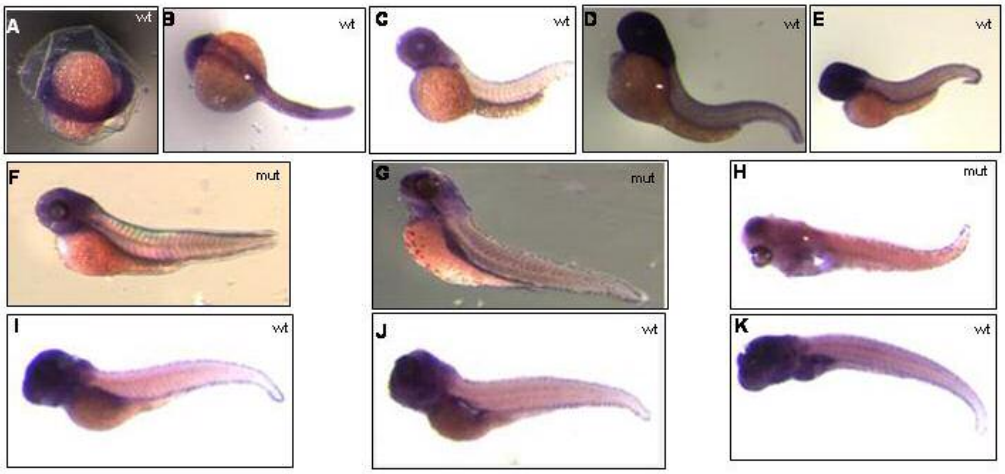


Figure 3.49 Expression pattern of gene Phf5a. The whole mount expression pattern of gene Phf5a in both wild type and mutant (HI1373) zebrafish embryos. In the wild type embryos (A) at 12hpf shows expression throughout the neural plate. (B) at 24hpf and (C) at 48hpf , expression can be seen in the brain. (D) at 55hpf and (E) 72hpf shows peak level of expression in brain in comparison to earlier stages. (I) at 96hpf,(J) at 120hpf and (K) at 168hpf still shows high level of expression in brain but slightly reduced in comparison to 55 and 72hpf stages. When compared to the wild type stages I (96hpf), J (120hpf), K (168hpf), the mutants (F) at 96hpf, (G) at 120hpf and (H) at 168 hpf showed both phenotypic abnormalities (constricted yolk sac, bend tail) and expression defects showing a drastic decrease in expression in the brain. Very few embryos lived past 168hpf, as the mutation was lethal in mature zebrafish.

These results show that PHF5A plays an important role in the survival of the embryo as the mutants show phenotypic defects as a consequence of reduced expression, ultimately leading to death. This gene likely could be functioning as a transcription factor or chromatin modulating or chromatin mediated transcriptional regulator, since the gene has a highly conserved PHD finger like domain in all the orthologs.

3.4 Summary

Throughout the present expression studies of human chromosome 22 CES-DGCR region orthologs in zebrafish, we observed that the majority of these genes were expressed in the brain in the early embryonic stages, and that most of the genes contain domains related to transcription factors or signaling factors. The genes with well-defined expression data in mouse also showed similar expression in zebrafish, thereby confirming the conservation of their expression during evolution.

4. Conclusions

4.1 Comparative Analysis between Different Species

Comparative analysis of the genes in the 4.5Mb region in human chromosome 22 with orthologous regions of the chimpanzee, baboon, bovine, mouse and zebrafish genomes facilitates our understanding the evolutionary relationships between species that shared a common ancestor at various times during the past 400 million years. These comparisons reveal that the distribution and types of repeats in the human, chimpanzee and baboon genomes favor the expansion of the genome with new exons, introns or UTR regions and an increased vulnerability to genetic disorders. Lineage specific repeat sequences inserted into the human genome include AluYa5 and AluYb9, while AluYg, AluYc3, AluYc5 are observed in chimpanzee, AluYa5 and AluYg occurs in baboon and bucentaur are bovine specific. Greater than 60% of the major repeats in human belong to the AluY family while they represent only 56% in the orthologous regions of chimpanzee. This observation suggests that human chromosome 22 is more susceptible to recombination than chimpanzee chromosome 22 and thus we humans are more prone to a variety of different diseases than chimpanzees. In humans, the frequency of insertions is more than that of deletions but in the case of chimpanzee there were more deletions observed than insertions while baboon was more similar to humans with more insertions than deletions. This likely indicates the influence of genomic expansion and the vulnerability for both beneficial and deleterious insertional changes. Sequence comparisons of genes sequenced from all three chimp libraries

(Clint,ch251; Donald, rp43; and Gon, ptb) revealed that they differed from each other to the same extent as humans and chimpanzees differ.

The majority of the amino acid substitutions in human, chimpanzee, baboon and bovine were changes from hydrophilic to hydrophilic amino acids with only a few examples of changes from hydrophobic to hydrophobic amino acids. From an evolutionary perspective, these changes show the selective constraint of favoring substitutions between similar classes of amino acids for species that are closer in evolutionary time. In addition, since hydrophilic amino acid residues usually are exposed to the aqueous environment, they often evolve at a faster rate in comparison to the hydrophobic amino acids that usually are buried within the protein three-dimensional structure. Since it also is known that changes in the hydrophilic solvent exposed environment to hydrophobic residues are less destabilizing and changes in the hydrophobic core of the protein are more destabilizing (Schwehm *et al*., 1998). Similarly, the minor substitutions observed in human and chimpanzee are less destabilizing as they are substitutions from hydrophilic to hydrophobic and the minor substitutions in baboon and bovine that were from hydrophobic to hydrophilic residues are more destabilizing. But the vast majority of the hydrophilic and hydrophobic amino acid substitutions are neutral and are context dependent.

The nucleotide and amino acid changes observed in the codon distribution in the regions compared from human, chimpanzee and baboon showed that >60% of the changes occurred in the third position, while the first and second position changes occurred slightly more often between humans and baboons (39%) than between humans and chimpanzees (34%). However, these differences alone are not sufficient to

describe the spectrum of differences in the primate phenotypes. The substitution rate comparison between human and chimp (1.2%) and between human and baboon (2.6%) and the Ka/Ks ratio (human and chimp (0.44); human and baboon (0.48)) all point to the selective pressure or constraints influencing the coding region that results from overall purifying selection especially between human and chimp in comparison to human and baboon. The few genes that showed Ka/Ks>1 occurred in the CES region that is known for its duplication in CES patients and maintaining these mutations through positive selection may have been beneficial for the evolutionary survival of their specific gene product functions. Codon usage was more biased towards codons ending in G or C (RSCU-relative codon usage value ranging between 1-4) as their tRNA are more abundant and >70% of the nucleotide changes in primates were transitional mononucleotide substitutions, while in the bovine genome the transitional changes often involved dinucleotide substitutions.

Through comparing the genes in the 4.5 MB human chromosome 22 CES-DGCR region with chimpanzee, and evaluating the changes in the repeats, base substitution rates, amino acid codon usage and nucleotide changes, highlight the differences between the two species. Comparison with other species have revealed the evolutionary conservation and changes present in human at the coding level. However, the biological and genetic significance of these differences and the genetic basis for the CES, DGCR, schizophrenia and other genetic diseases related to this region of human chromosome 22 are yet to be determined. The effect of the observed differences in the regulation of genes will be interesting and will help to further our understanding of the genetic basis for these syndromes as recently was observed (Gilad *et al*., 2006).

Scanning for regions highly diverged and highly conserved (on the basis of SNP's), determining the critical copy number of the genes present with respect to LCR's in the entire chimpanzee chromosome 22 with respect to human chromosome 22 also would be enlightening.

4.2 Genes Postulated to be Involved in Cat-Eye Syndrome (CES)

The expression profiling of genes in the cat eye syndrome region using zebrafish model system extends the earlier studies carried out to understand this rare developmental disorder. Both known and predicted genes in this region showed interesting expression patterns, many of which matched the altered phenotype in CES patients.

Genes	**Human studies /northern blot /RT-PCR**	**Mouse model studies**	**Zebrafish model studies by whole mount *in situ* studies**
ENSG00000138860			Expressed in the tectum, midbrain, eye, hindbrain and spinal chord neuron in early stages and shows decreased expression in later stages.
IL17R	All over the body		Expression majority in the brain and all over the embryo.
CECR5	All over the body	Expression in brain libraries with notable expression in the retina and kidney.	All over the body especially in the brain, pectoral fin, retina and pronepheric duct
CECR2		EST's sequence from a mouse cDNA library of brain, muscles, eye and lungs	Expression in the brain, ear and eye.
BCL2L13	In all adult and foetal tissues tested especially in brain, spinal chord and muscles.		Expression in the brain especially midbrain and hind brain, muscles in the early stages of development and after 48hpf no expression is seen.
O94909/kiaa0819			Expression mainly in the ear
Pex26	Expression in brain , liver, kidney, and skeletal muscles	EST's sequence from a mouse cDNA library mainly in brain and kidney, pancreas and liver.	Expression in the brain, notochord, liver, otic placode and pectoral fins

Table 4.1 List of genes in the CES region with their expression pattern in zebrafish with reference to earlier studies in human and mouse.(Footz *et al*., 2001; McDermid *et al.,* 2002; the Functional Annotation of Mouse (FANTOM) consortium, 2005).

As shown above in table 4.1, the expression patterns observed show the efficacy of using the zebrafish model to study this disease. Since Cat Eye Syndromes results in defects in the eye, ear, nose, head and neck, cardiovascular system, gastrointestinal system, various skeletal defects and urogenital system (McTaggart *et al*., 1998; Schinzel *et al*., 1981; McDermid *et al.*, 2002), the genes that correlate to the different phenotypes with respect to their expression patterns in zebrafish are shown in table 4.2.

Phenotypes affecting different organs in CES syndrome	**Genes expressed in different organs based on Zebrafish whole mount *in situ***
Eyes	CECR5, CECR2, ENSG00000138860
Ear	CECR2, O94909(kiaa0819:zf1 and zf2)
Brain	ENSG00000138860, il17r, CECR5, CECR2, BCL2L13
Kidney	CECR5 , PEX26
Liver	Pex26

Table 4.2 List of genes grouped based on their expression in relation to the phenotypes noticed in CES syndrome.

In the case of the novel gene ENSG00000138860, the presence of the pleckstrin homology domain (Ingley and Hemmings, 1994) that is known to be involved in intracellular signaling, may be sensitive to dosage changes and thus be an interesting candidate gene that alters the phenotype when it is over expressed. The gene kiaa0819 with two copies (zf1 and zf2) in zebrafish also showed interesting expression. The zf1 genes in the early embryonic and late stages showed consistent expression in the ear while zf2 showed basal expression throughout the body with slightly high expression in the brain in the early stages. However, in later developmental stages it showed expression similar to zf1 in the ear with both specific expression in early stages and sub-functionalization in later stages. It also was noted

that even though the expression in both cases is in the ear, zf1 was expressed specifically in the otolith and maculae but zf2 was expressed in the otic vesicle. The conservation of transcription factors upstream of these genes are unique to each copy but similar to the single human gene transcription factors that may indicate differential regulation. The study of zf1 gene with ace mutant defective in the fgf8 gene with no expression in ear, allows speculation of the relatedness between this gene and its dependency on fgf8 growth factor, studies that could be expanded by experiments in which the zf1 genes could be knock down and rescued with fgf8 and vice versa. It also was noticed that genes CECR2, BCL2L13, ENSG00000138860 have reduction in expression at 72hpf compared to early stages signifying the early embryonic importance of these genes. Based on the location of each of the genes and their expression pattern in early embryonic stages suggests the important role played by the genes in contributing to the anomalies seen in CES disorder and over-expression of each of the gene or gene group mainly to the distal region may be responsible for the sum total of the alterations in the CES phenotype.

4.3 Genes Postulated to be Involved in DiGeorge Syndrome

DGCR/VCFS is caused by deletions of human chromosome 22q11.2 (genes contained in 1.5 Mb or 3 Mb of 22q11) and these deletions have been intensely studied to better understand the etiology of this syndrome that is mainly categorized as causing a variety of phenotypes affecting multiple organs, including craniofacial defects, thymus hypoplasia, cardiac problems, abnormalities in neural crest migration, hearing loss, cleft palate, skeletal, limb, kidney, increased incidence of schizophrenia, growth

delays. Throughout the earlier work, DGCR has been thought of as a single gene defect, that has lead to numerous studies and arguments pointing to one gene or another in the major DGCR region as being responsible for the observed phenotype as observed in genes such as HIRA (Pizzuti,1999; Roberts *et al*., 1999;2002), COMT (Baker *et al*., 2005), PRODH (Paylor *et al*.,2001; Gogos et al,1999), CRKL (Guris *et al*., 2001), Es2/dgsi (Lindsay et al., 1998), UFD1L (Pizzuti *et al*., 1997;Yamagishi *et al*., 2003), and DGCR6 (Liu *et al*., 2002). Studies using *in situ* hybridization the mouse model (Funke *et al*., 2001), northern blot analysis studies of 24 genes (DGCR2 to ZNF74), deletion and complementary duplication rescue studies in mice (Puech *et al*., 2000), have attempted to account for the observed DGCR/VCF phenotype anomalies. Results have showed that more than one gene is affected during this disease as was observed in the case of UFD1L or shared expression domains as seen in DGSI and GSCL. Recent deletion studies in mice also have shown that Tbx1 is a haplosufficient gene that may independently cause the multiple DGCR/VCF phenotype (Lindsay *et al*., 2001). It also should be pointed out that there is a strong correlation between the lack of the DGCR postulated genes like ZDHHC8 (Mukai *et al*., 2004), ARVCF (Chen *et al*., 2005; Sanders *et al*, 2005) , COMT (Shifman *et al*., 2004; Handoko *et al*., 2005), UFD1L (De Luca *et al*., 2001), PCQAP (De Luca *et al*., 2003; Sandhu *et al*., 2004), DGCR6 and PRODH (Liu *et al*., 2002) and the higher incidences of schizophrenia .

Further refinement in the annotation of the human chromosome 22 has increased the number of genes spanning the critical DGCR region during the course of this dissertation research. The expression of novel genes in addition to known genes not addressed before this study results in the genes listed in table 4.3 and further

establishes the possibility of this disorder being mutli-gene rather than single gene related. The idea of DGCR being a multigene syndrome previously was investigated by studying brain gene expression (Maynard *et al*.,2003).

Summary of the list of DGCR genes:known (9) and unknown (4) gene expression

Gene	Zebrafish whole mount *in situ* studies
ENSG00000185643	Expression in midbrain, hindbrain, otic placode, pectoral fins, pronepheric duct, branchial arches and myotomes at early stages of development in zebrafish
SLC25A1	Expression of this gene in the midbrain and hindbrain, frontal region, and in the pectoral fins during stages 24, 48 and 72hpf
MRPL40	Expression in brain and branchial arches at 24, 48, and 72hpf stages
TBX1	24hpf, 48hpf, and 72hpf zebrafish embryos shows expression in the brain, heart, otic placode, swim bladder, and notochord
ARVCF	Expression in the brain and frontonasal region of the zebrafish embryos
DGCR8	Expression in hindbrain, branchial arches, heart, and pectoral fins.
HTF9C	Expression mainly in the somites, 7 above the yolk cell, 10 above the yolk extension, and 13 posterior to the anus which will later gives rise to myotomes and dermatome. The expression of this gene could be seen at 20-24hpf, higher expression at 48hpf and no expression at 72hpf
ZDHHC8	Expression in the brain, ear, distal part of tail, branchial arches, and heart
ENSG00000182364	Expression in brain, branchial arches, thymus ,pectoral fin, spinal chord and in additional to these at 24hpf , the expression is also seen in the somites and the myotomes at early stages of development in zebrafish.
ENSG00000183099	Expression all over the body at 24hpf and more defined expression at 48hpf in the pronephric duct, liver, pharyngeal arches, brain, somites, myotomes and gradual decrease at 72hpf in zebrafish
SCARF2	Expression in brain , branchial arches, eyes and all over the body at 24hpf and 48hpf but shows dramatic reduction in 72hpf.
ENSG00000185214	Basal expression all over the embryo but mainly in the epidermis especially in the eye in the early stages of development.
PCQAP	Expression in the notochord, pectoral fins and brain
PIK4CA	Expression in the brain and indicated basal levels of expression in pectoral fins. Since this gene is important in brain development.
CRKL	Expression all over especially in the brain, notochord, muscles.
Phf5a	Expression in the brain.

Table 4.3 List of genes in the DGCR region with their expression pattern in zebrafish.

In the present studies, based on the expression pattern as summarized in table 4.4 above, it is reasonable to conclude that if these genes are deleted the result would be sufficient to produce the broad spectrum of phenotypes observed in DGCR/VCF.

Phenotypes noticed in DGCR with respect to different organs	**Genes studied in this dissertation using zebrafish model whole mount *in situ* hybridization (exon specific ssDNA probe)**
Brain	ENSG00000185643 SLC25A1, MRPL40, TBX1, ARVCF, DGCR8, HTF9C, ZDHHC8, ENSG00000182364, ENSG00000183099, SCARF2, ENSG00000185214, PCQAP, PIK4CA, CRKL, Phf5a
heart	TBX1, DGCR8, ZDHHC8
Skeletal (muscles, notochord)	ENSG00000185643,TBX1,ENSG00000182364, ENSG00000183099 ,PCQAP
Limb(pectoral fins)	ENSG00000185643, SLC25A1, DGCR8, ENSG00000182364, PCQAP, PIK4CA,
Thymus	ENSG00000182364
Pharyngeal arches	SCARF2, ENSG00000182364, ENSG00000183099, DGCR8, ZDHHC8
Kidney	ENSG00000183099, CECR5, ENSG00000185643,
Craniofacial (nose, ear, eyes, mouth)	ARVCF, SLC25A1, ENSG00000185214, TBX1, DGCR8, ZDHHC8, SCARF2, ENSG00000185643.

Table 4.4 List of genes grouped based on their expression in relation to the phenotypes noticed in DGCR syndrome.

Information from earlier deletions studies in mice (Df1 (ES2 to UFD1L->1.2Mb), Df2 (Es2 to T10->500Kb) , Df3 (ES2 to CDCREL-1->700kb) , Df4 (T10 to HIRA), and 550kb (DGCR2 to ARVCF) deletion (Lindsay *et al*., 2001;Puech *et al*., 2000;Schinke *et al*., 2001;Prescott *et al*., 2005) and the zebrafish *in situ* expression studies reported in this dissertation, enables further analysis of the effect of deletions in the DGCR/VCF region. The earlier deletions carried out mainly were focused on proving that the TBX1 causes cardiovascular defects (Lindsay *et al*., 2001; Merscher *et al*., 2001) entirely based on non-overlapping gene expression of genes near TBX1, in addition to the loss of function studies that were carried out to show the requirement of TBX1 for inner ear morphogenesis (Vitelli *et al*., 2003; Raft *et al*., 2003). The embryonic stages used during the experiments were limited due to constraints using the mice model. However our studies using zebrafish expression data from very early stages (one cell stage onwards) paves way to further refine these deletion studies in mice, an experiment that is not feasible in zebrafish because the deletions must be

restricted to only a few of the potential disease-related because of the very short stretches of synteny in zebrafish that are spread over different chromosomes in comparison to mice where they are collinear on chromosome 16.

Taken together, the results of our present studies indicate that the majority of the predicted genes in the DGCR critical region indeed are expressed in regions of the developing embryo that are directly related to the phenotype presented by DGCR patients. Thus, one of the major conclusions from this dissertation research is to support and extend the hypothesis that both CES and DGCR/VCFS are multigene based diseases. In the future, it is likely that additional studies aimed at determining the direct involvement of these gene in the various genetic disorders linked to this region of human chromosome 22 will require further analysis of the pathways and other interactions in which the encoded gene products are involved.

References

1. Adams, M.D. et al. (2000) The genome sequence of Drosophila melanogaster. Science, **287**, 2185-2195.
2. Aerts, S., Thijs, G., Coessens, Bert., Staes, M., Moreau, Y and Moor, D,B. (2003). Toucan: deciphering the cis-regulatory logic of coregulated genes. Nucleic Acids Research, **31**, 1753-1764.
3. Aggarwal, S., and Gurney, A.L. (2002). IL-17: prototype member of an emerging cytokine family. Journal of Leukocyte Biology, **71**, 1-8.
4. Alexandersson, M., Cawley, S., and Pachter, L. (2003). Slam: Cross-species finding and alignment with a generalized pair hidden Markov model. Genome Res, **13**, 496-502.
5. Allendorf, F.W., Utter, F.M., and Max, B.P (1975) Gene duplication within the family Salmonidae: II. Detection and determination of the genetic control of duplicate loci through inheritance studies and the examination of population. Isozymes, edited by C.L., Market, Academic Press, NY. pp. 415-432.
6. Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Volff, J-N., and Schartl, M. (2002) Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. Genetics. **161**, 259-267.
7. Altschul, S.F., Gish, W., Miller, W., Myers, W.E and Lipman, J.D. (1990) Basic local alignment search tool. J.Mol. Biol. **215**,403-410.
8. Altschul, S.F., Madden, L.T., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res, **25**, 3389-3402.

9. Amarger, V., Nguyen, M., Laere, A-S, V., Braunschweig, M., Nezer, C., Georges, M., and Andersson, L. (2002) Comparative sequence analysis of the INS-IGF2-H19 gene cluster in pigs. Mammalian Genome, **13**, 388-398.

10. Amemiya, C., Zhong, T., Silverman, G., Fishman, M., and Zon, L.(1998) Zebrafish YAC, BAC, and PAC genomic libraries. Methods Cell Biol. **60**, 235-258.

11. Amores, A., Force, A., Yan, Yi-Lin., Joly, L., Amemiya, C, Fritz, A., Ho, K.R., Langeland, J., Prince, V., Wang, Y-L., Westerfield, M., Ekker, M and Postlethwait, J.H. (1999) Zebrafish hox clusters and vertebrate genome evolution. Science. **282**, 1711-1714.

12. Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M., and Hopkins, N. (1999). A large scale insertional mutagenesis screen in zebrafish. Genes & Dev. **13**, 2713-2724.

13. Ansorge, W., Sproat, B., Steigmann, J., Schwager, C., and Zenke, M. (1987) Automated DNA sequencing: ultra sensitive detection of fluorescent bands during electrophoresis. Nucl. Acid. Res. **15**, 4593-4602.

14. Anzai, T., Shilna, T., Kimura, N., Yangiya, K., Kohara, S., Shigenari, A., Yamagata, T., Kulski, K, J., Naruse, K.T., Fujimori, Y., Fukuzumi, Y, Yamazaki, M., Tashiro, H., Iwamoto, C., Umehara, Y., Imanish, T., Meyer, A., Ikeo, K., Gojobori, T., Baharam, S., and Inoko, H. (2003). Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. PNAS, **100**, 7708-7713.

15. Aparicio, S. et al. (2002). Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science, **297**, 1301-1310.

16. Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhari, et al. (1995). Detecting conserved regulatory elements with the model genome of Japanese puffer fish, Fugu rubripes. Proc. Natl. Acad. Sci. USA, **92**, 1684-1688.

17. Apoil, P-A., and Blancher, A. (2000) Rh Gene Evolution in Primates: Study of Intron sequences. Mol. Biol. Evol. **17**, 127-136.

18. Armengol, L., Pujana, M.A., Cheung, J., Schere, S.W., and Estivill,X. (2003) Enrichment of segmental duplications in regions of breaks of synteny between human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum. Mol. Genet. **12**, 2201-2208.

19. Avery, O.T., Macleod, C.M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation in pneumococcal types. J. Exp. Med, **79**, 137-158.

20. Babcock, M., Pavlicek, A., Spiteri, E., Kashork, C.D., Ioshikhes, I., Shaffer, L.G., Jurka, J., and Morrow, B.E. (2003) Shuffling of genes within low-copy repeats on 22q11 (LCR22) by Alu-mediated rtecombination events during evolution. Genome Res., **13,** 2519-2532.

21. Babenko, V.N., Rogozin, I.B., Mekhedov, S.L and Koonin, E.V. (2004) Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic. Acids. Res. **32**(12), 3724-3733.

22. Bailey, J.A., Yavor, A.M., Viggiano, L., Misceo, D., Horvath, J.E et al. (2002). Human–specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. Am. J. Hum. Genet. **70**, 83-100.

23. Bailey, J.A., Liu, G., and Eichler, E.E. (2003) An Alu transposition model for the origin and expansion of human segmental duplications. Am. J. Hum. Genet, **73**, 823-834.

24. Bailey, A.J., Yavor, M.A., Massa, F.H., Trask, J.B., and Eichler, E.E. (2001) Segmental duplications: organization and impact within the current human genome project assembly. Genome Res, **11**, 1005-1017.

25. Bailey, J.A., Church, D.M., Ventura, M., Rocchi, M., N and Eichler, E. (2004) Analysis of segmental duplications and genome assembly in mouse. Genome Res, **14**, 789-801.

26. Baker, K., Baldeweg, T., Sivagnanasundaram, S., Scambler, P., and Skuse, D.(2005) COMT Val108/158 Met modifies mismatch negatively and cognitive function in 22q11 deletion syndrome. Biol Psychiatry, **58**, 23-31.

27. Band, R. M., Larson, H.J., Rebeiz, M., Green, A.G., Heyen, W.D., Donovan, J., Windish, R., Steining, C., Mahyuddin, P., Womack, E.J., and Lewin, A.H. (2000) An ordered comparative map of the cattle and human genomes. Genome Res,**10**, 1359-1368.

28. Bankier, A.T and Barrell, B.G. (1989). Sequencing single stranded DNA using the chain termination method in nucleic acid sequencing : A practical approach. JRL press Oxford, pp 37-78.

29. Bannert, N., and Kurth, R. (2004) Retroelements and the human genome: new perspectives on an old relation. PNAS, **101**, 14572-14579.

30. Banting, G.S., Barak, O., Ames, T.M., Burnham, A.C., Kardel, M.D., Cooch, N,S., Davidson, C.E., Godbout, R., McDermid, H.E., and Shiekhattar,R. (2005) CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. Hum Mol Genet. **14**(4), 513-524.

31. Barbazuk, W.B., Korf, L., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bede, J.A., McPherson, J.D., Johnson, S.L. (2000) The synyenic relationship of the zebrafish and human genome. Genome Res. **10**, 1351-1358.

32. Barbulescu, M., Turner, G., Su, M., Kim, R., Jensen-Seaman. I.M., Deinard, S, A., Kidd, K.K and Lenz, J. (2001) A HERV-K provirus in chimpanzees, bonobos and gorillas, but not humans. Current Biology, **11**, 779-783.

33. Bateman, A., Birney, E., Ceruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L ., Marshall. M., Sonnhammer, E.L. (2002). The Pfam protein families database. Nucleic Acids Research, **30**,276-280.

34. Batzer, M.A., and Deininger, P.L. (1991). A human-specific subfamily of Alu sequences. Genomics, **9**, 481-487.

35. Batzer, M.A., and Deininger, P.L. (2002) Alu repeats and human genomic diversity. Nature, **3**, 370-380.

36. Batzer, M.A., Gudi, V.A., Mena, J.C., Foltz, D.W., and Herrera, R.J., et al., (1991). Amplification dynamics of human-specific (HS) Alu family members. Nucleic Acids Res. **19**, 3619-3623.

37. Batzer, M.A., Kilroy, G.E., Richard, P.E., Shaikh,T.H., Desselle, T.D., Hoppens,C.L and Deininger, P.L. (1990) Structure and variability of recently inserted Alu family members. Nucleic Acids Res. **18**, 6793-6798**.**

38. Batzer, M.A., Rubin, C.M., Hellmann-Blumberg, U., Algeria-Hartman and Leeflang, E.P. et al., (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. J.Mol. Evol. **247**, 418-427.

39. Batzoglou, S., Pachter, L., Mesirov, J.P., Berger, B., and Lander, E.S. (2000) Human nd mouse gene structure: Comparative analysis and application to exon prediction. Genome Res, **10**,950-958.

40. Beadle, G.W., and Tatum, E.L. Genetic control of biochemical ractions in Neurospora. Proc. Natl. Acad. Sci. USA, **27**, 499-506.

41. Bedell, M.A., Jenkins, N.A., and Copeland, N.G. (1997) Mouse models of human disease: Part I: Techniques and resources for genetic analysis in mice. Genes&Dev, **11**, 1-10.

42. Bedell, M.A., Largaespada, N.A., Jenkins, N.A.,and Copeland,N.G.(1997) Mouse models of human disease: Part II :recent progress and future directions. Genes & Dev. **11**, 11-43.

43. Bennett, C.M., Kanki, J.P., Rhodes, J., Liu, T.X., Paw, B.H., Kieran, M.W., andd Langenau, D.M. (2001) Myelopoiesis in the zebrafish, Danio rerio. American Society of Hematology. Blood, **98**, 643-651.

44. Berti, L., Mittler, G., Przemeck,G.K., Stelzer, G., Gunzler,B., Amati,F., Conti, E., Dallapiccola,B., Hrabe de Angelis,M., Novelli,G., and Meisterernst,M. (2001)

Isolation and characterization of a novel gene from the DiGeorge chromosomal region that encodes for a mediator subunit. Genomics, **74**(3), 320-332.

45. Best, R.G., Diamond, D., Crawford, E., Grass, F.S., Janish, C., Lear,T.L., Soenksen, D., Szalay, A.A., and Moore, C.M. (1998). Baboon/human homologies examined by special karyotyping (SKY): A visual comparison. Cytogenet. Cell. Genet. **82**, 83-87.
46. Betran, E., Wang, W., Jin, L., and Long, L. (2002). Evolution of the Phosphoglycerate mutase Processed Gene in human and chimpanzee revealing the Origin of a New Primate Gene. Mol. Biol. Evol, **19**, 654-663.
47. Birnboim, H.C and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res.**7**, 1513-1523.
48. Blangero, J., MacCluer, J.W., Kammerer, C.M., Mott, G.E., Dyer, T.D., McGill, H.C Jr (1990) Genetic analysis of apolipoprotein A-1 in two dietary environments. Am J Hum Genet. **47**, 414-428.
49. Blanjero, J. (1993) Statistical genetic approaches to human adaptability. Hum Biol. **63**, 941-966.
50. Blechschmidt, K., Schweiger, M., Wertz, K., Poulson, R., Christensen, H-M., Rosenthal, A., Lehrach, H., and Yaspo, M-L. (1999) The mouse Aire gene: comparative genomic sequencing, gene organization, and expression. Genome Res, **9**, 158-166.
51. Bodenteich, A., Chissoe, S., Wang, Y.F., Roe, B.A. (1994). Shotgun cloning as the strategy of choice to generate templates for high throughput

dideoxynucleotide sequencing and analysis techniques. C. Venter, Ed. Academic Press London .

52. Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, D.K., Ovcharenko, I., Pachter, L., Rubin, M.E. (2003). Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science, **299**, 1391-1394.

53. Boffelli, D., Nobrega, M.A and Rubin, E.M. (2004) Comparative genomics at the vertebrate extremes. Nat. Rev. Genet.,**5**, 456-465.

54. Brand, M., Heisenberg, C,-P., Jiang, Y.-J., Beuchle, D., Lun, K., vanEden, F.J. M., Furutani-Seiki, M., Granato, M., Haffter, P., Hammerschmidt, M., Kane, D.A., Kelsh, R. N., Mullins, M. C., Odenthal, J and Nusslein-Volhard, C. (1996) Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain. Development, **123**,179-190.

55. Britten, R.J. (2002). Divergence between samples of chimpanzee and human DNA sequences is 5% counting indels. Proc. Natl. Acad. Sci, **99**, 13633-13625.

56. Britten, R.J., and Kohne, D.E. (1968) Repeated sequences in DNA. Science, **161**, 529-540.

57. Britten, R.J., Baron, W.F., Stout, D.B., Davidson, E.H. (1998) Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci USA, **85**, 4770-4774.

58. Britten, Roy J. (2002), "Divergence between Samples of Chimpanzee and Human DNA Sequences is 5%, Counting Intels," *Proceedings of the National Academy of Sciences*, **99**:13633-13635.

59. Brown, N.L (1984). DNA sequencing methodology in Micro., Academic Press, NY, **17**, 259-312.

60. Caccone, A., Powell, J.F. (1989) DNA divergence among hominoids. Evolution. **43**,925-942.

61. Cagan, R.L., and Zipursky, S.L. (1992) Cell choice and patterning in the Drosophila retina. In M. Shankland and E.R.Macagno (eds): Determinants of Neuronal Identity. San Diego: Academic Press, pp.189-224.

62. Carey, K.D., and Rice, K.S (1996) The aged female baboon as a model of menopause. Paper presented at the annual meeting of the American Society of Primatologists, Madison, August, 11-16.

63. Carlson, C., Sirotkin, H., Pandita, R., Goldberg, R., McKie, J., Wadey, R., Patanjali, S.R., Weissman, S.M., Anyane-Yeboa, K., Warburton, D et al. (1997). Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients. Am. J. Hum. Genet. **61**, 620-629.

64. Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.,Vogel, E., et al., (2001) Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J.Mol. Biol. **311**, 17-40.

65. Cavalier-Smith, T. (1978) Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth and the solution to the DNA C-value paradox. J. Cell. Sci, **34**, 247-278.

66. Chapman, M.A, Charchar, F.J, Kinston, S., Bird, C.P., Grafham, D et al. (2003) Comparative and functional analyses of LYL1 loci establish marsupial sequences as a model for phylogenetic footprinting. Genomics, **81**, 249-259.

67. Chargaff, E. (1951) Some recent studies on the composition and structure of nucleic acids. J.Cell. Comp. Physiol. **38**(Suppl.1) 41-59.

68. Chargaff, E., Lipschitz, R., Green, C., and Hodes, E.M. (1951) The composition of the deoxyribonucleic acid. Nature, **171**, 737-738.

69. Chen, C., Gentles, J.A., Jurka, J., and Karlin, S (2002) Genes, pseudogenes and Alu sequence organization across human chromosomes 21 and 22. PNAS, **99**, 2930-2935.

70. Chen, F.-C and Li, W.-H. (2001) Genomic Divergences between human and other hominoids and the effective population size of the common ancestor of humans and chimpanzee. Am. J. Hum. Genet. **68**, 444-456.

71. Chen, H.Y., Yeh, J.I., Hong, C.J., and Chen, C.H. (2005) Mutation analysis of ARVCF gene on chromosome 22q11 as a candidate for s schizophrenia gene. Schizophr Res. **72**(2-3), 275-277.

72. Chen, K.S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, C.A, Lee, C.C., and Lupski, J.R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. Nat. Genet. **17**, 154-163.

73. Chiang, E.F-L., Yan, Y-L., Guiguen, Y., Postlethwait, J and Chung, B-c. (2001) Two Cyp19 (P450 Aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain. Mol. Biol. Evol. **18**, 542-550.

74. Chieffo, C.; Garvey, N.; Gong, W.; Roe, B.; Zhang, G.; Silver, L.; Emanuel, B. S.; Budarf, M. L. **(**1997) Isolation and characterization of a gene from the

DiGeorge chromosomal region homologous to the mouse Tbx1 gene. Genomics **43**, 267-277.

75. Chiquet, M., Mumenthaler, U., Wittwer, M., Jin, W., and Koch, M. (1998) The chick and human collagen α1(XII) gene promoter. Activity of highly conserved regions around the first exon and in the first intron. Eur. J. Biochem. **257**, 362-371.

76. Chissoe, S.L., Wang, Y.F., Clifton, S.W., Ma, N., Sun, H.J., Lobsinger, J.S., Kenton, S.M., White, J.D., and Roe, B.A. (1991) Strategies for rapid and accurate DNA sequencing methods: A Comparison to Methods in Enzymology. **3**, 55-65.

77. Cho, G. and Doolittle, R.F.(1997) Intron distribution in ancient paralogs supports random insertion and not random loss. J.Mol.Evol., **44**, 573-584.

78. Chomczynski, P and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical Biochemistry, **162**,156-159

79. Chuzhanova, N.A., Krawczak, M., Nemytikova, L.A., Gusev, V.D., & Cooper, D.N. (2000) Promoter shuffling has occurred during the evolution of the vertebrate growth hormone gene. Gene. **254**, 9-18.

80. Clark, G, A., Glanowski, S, Nielsen, R., Thomas, D.P., Kejariwal, A., Todd, A.M., Tanenbaum, M,D., Civello, D., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X., White, J.T., Sninsky, J.J., Adams, D,M ,Cargill, M. (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science, **302**, 1960-1963.

81. Coghlan, A., and Wolfe, K.H (2004) Origins of recently gained introns in Caenorhabditis. PNAS, **101**(31), 11362-11367.

82. Collins, F.S. (1995) Positional cloning mones from preditional to traditional. Nature Genet. **9**, 347-350.

83. Collip, J.B. (1925) The extraction of parathyroid hormone which will prevent or control parathyroid tetany and which regulates the level of blood calcium. J. Bio. Chem. **63**, 395-438.

84. Comeron, J.M., Aguade, M. (1998) An evaluation of measures of synonymous codon usage bias. J.Mol. Evol., **47**, 268-274.

85. Comeron, J.M., and Kreitman, M. (1998) The correlation between synonymous and nonsynonymous substitutions in Drosophila: Mutation, selection or relaxed constraints? Genetics, **150**, 767-775.

86. Conant, G.C., and Wagner, A (2003) Asymmetric sequence divergence of duplicate genes. Genome Res., **13**, 2052-2058.

87. Cooper, G.N. (1994). The human genome project: Deciphering the blueprint of heredity. University Science Books, CA.

88. Courseaux, A and Nahon, J. L. (2001) Birth of two chimeric genes in the Hominidae lineage. Science. **291**, 1293-1297.

89. Cox, L.A. et al. (2002) Identification of candidate genes regulating HDL cholesterol using a chromosomal region expression array. Genome Res. **12**, 1693-1702.

90. Craxton, M. (1991) Linear amplification sequencing: a powerful method of sequencing DNA. Methods: Companion Methods Enzymol. **3**,20-26.

91. Crick, F. H. C., Burnett, L., Brenner, S and Watts-Tobin, R.J. (1961) General nature of the genetic code for proteins. Nature, **192**, 1227-1232.

92. Dagan, T., Sorek, R., Sharon, E., Ast, G., and Graur, D. (2004). AluGene: a database of Alu elements incorporated within protein coding genes. Nucleic Acids Res.**32**, D489-D492.

93. Dalbey, R.E., and Von Heijne,G.(1992) Signal peptidases in prokaryotes and eukaryotes—a new protease family. Trends Biochem Sci. **17**(11),474-478.

94. Danielle and Jean Thierry-Mieg, Michel Potdevin, Mark Sienkiewicz. Identification and functional annotation of cDNA-supported genes in higher organisms using AceView, unpublished. http://www.aceview.org.

95. De Jong, R., Haataja, L., Voncken, J.W., Heisterkamp, N., and Groffen, J. (1995) Tyrosine phosphorylation of murine Crkl. Oncogene, **11**(8), 1469-1474.

96. De Luca A, Pasini A, Amati F, Botta A, Spalletta G, Alimenti S, Caccamo F, Conti E, Trakalo J, Macciardi F, Dallapiccola B, Novelli G. (2001) Association study of a promoter polymorphism of UFD1L gene with schizophrenia. Am J Med. Genet., **105**(6), 529-533.

97. De Luca A, Conti E, Grifone N, Amati F, Spalletta G, Caltagirone C, Bonaviri G, Pasini A, Gennarelli M, Stefano B, Berti L, Mittler G, Meisterernst M, Dallaniccola,B., Novelli,G. (2003) Association study between CAG trinucleotide repeats in the PCQAP gene (PC2 glutamine/Q-rich-associated protein) and schizophrenia. Am J Med Genet B Neuropsychiatr Genet., **116**(1), 32-35.

98. Dehal, P et al. (2002) The draft genome of Ciona intestinalis: insight into chordates and vertebrates origins. Science, **298**, 2157-2167.

99. Deininger, L. P., and Batzer, A.M. (1999) Alu repeats and human disease. Molecular Genetics and Metabolism, **67**, 183-193.

100. Deininger, P.L (1983). Random subcloning of sonicated DNA. Application to shotgun sequence analysis. Anal. Biochem. **129**, 216-233.

101. Deininger, P.L, Moran, J.V., Batzer, M.A., Kazazian, H.H.Jr. (2003) Mobile elements and mammalian genome evolution. Curr. Opin. Genet. Dev. **13**, 651-658.

102. Dermitzakis, T. E and Clark, G.A. (2002) Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. Mol. Biol. Evol, **19**, 1114-1121.

103. Dierick, H., Stul, M., Kelver, W,De., Marynen, P and Cassiman, J-J. (1993) In corporation of dITP or 7-deaza dGTP during PCR improves sequencing of the product. Nucleic Acids Res. **21**, 4427-4428.

104. Dodd, A., Curtis, P.M., Williams, L.C., Love, D.R. (2000) Zebrafish: Bridging the gap between development and disease. Hum. Mol. Genet. **9**, 2443-2449.

105. Don, R.H., Cox, P.T., Wainwright, B.J., Baker, and Mattick.(1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. **19**:4008.

106. Dooley, K and Zon, L.I (2000) Zebrafish: A model system for the study of human disease. Curr. Opin. Genet. Dev. **10**, 252-256.

107. Donovan, A., Brownlie, A., Zhou, Y., Shepard, J., Pratt, S.J., Moynihan, J., Paw, B.H., Drejer, A., Barut, B., Zapata, A., et al. (2000) Positional cloning of

zebrafish ferroportin 1 identifies a conserved vertebrate iron exporter. Nature. **403**, 776-781.

108. Dubchak, I. et al. (2000). Active conservation of noncoding sequences revealed by three-way species comparisons. Genome Res, **10**, 1304-1306.

109. Duncan, C., Biro, P.A., Choudary, P.V., Elder, J.T., Wang, R.R., Forget, B.G., de Riel, J.K., and Weissman, S.M. (1979). RNA polymerase III transcriptional units are interspersed among human non-alpha-globin genes. Proc. Natl. Acad. Sci., **76**, 5095-5099.

110. Duret, L., and Mouchiroud, D. (1999) Expression pattern and, surprisingly gene length shape codon usage in *Caenorhabditis, Drosophila,* and *Arabidopsis*. Proc Natl. Acad. Sci. USA **96**, 4482-4487.

111. Ebersberger, I., Metzler, D., Schwarz,C., and Paabo, S. (2002) Genome wide comparison of DNA sequence between humans and chimpanzees. Am.J. Hum. Genet, **70**, 1490-1497.

112. Edelmann, L., Pandita, R.K., and Morrow, B.E. (1999) Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. Am. J. Hum. Genet., **64**, 1076-1086.

113. Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R.S.K., Magenis, E., Shprintzen, R.J and Morrow, B.E. (1999) A common molecular basis for rearrangement disorders on chromosome 22q11. Hum. Mol. Genet, 8, 1157-1167.

114. Edenberg, H.J. (2000) Regulation of the mammalian alcohol dehydrogenase genes Prog. Nucleic Acid Res. Mol. Biol. **64**, 295-341.

115. Eichler, E.E and DeJong. J.P. (2002) Biomedical applications and studies of molecular evolution: A proposal for a primate genomic library resource. Genome Res, **12**, 673-678.

116. Eichler, E.E. (2001). Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet,17(**11**), 661-669.

117. Eichler. E,E et al.(1996) Duplication of gene cluster between 16p11.1 and Xq28: a novel pericentromic directed mechanism for paralogous genome evolution. Hum. Mol. Genet. **5**, 898-912.

118. Eisen, J.S., and Pike, S.H. (1991) The spt-1 mutation alters segmental arrangement and axonal development of identified neurons in the spinal cord of the embryonic zebrafish. Neuron. **6**, 767-776.

119. Eisen,.J.S., and Weston, J.A. (1993) Development of the neural crest in the zebrafish. Dev. Biol. **159**, 50-59.

120. Ekker, M., Akimenko, M., Allende, M., Smith, R., Drouin, G., et al.(1997) Relationships among msx gene structure and function in zebrafish and other vertebrates. Mol. Biol. Evol. **14**, 1008-1022.

121. Ekker, S.C., Ungar, A.R., Greenstein, P., vonKesseler, D.P., Porter, J.A et al.(1995) Patterning activities of vertebrate hedgehog proteins in the developing eye and brain. Curr. Biol. **5**, 944-955.

122. Ellsworth, R.E., Jamison, D.C., Touchman, J.W., Chissoe, S.L., Braden, M.V., Bouffard, G.G., Dietrich, N.L., Beckstorm-Sternberg, S.M., Iyer, L.M., Weintrab, L.A., et al. (2000) Comparative genomic sequence analysis of the human and

mouse cystic fibrosis transmembrane conductance regulator genes. Proc. Natl. Acad. Sci. **97**, 1172-1177.

123. Elnitski, L., Reimer, C., Petrykowska, H., Florea, L., Schwartz, S., Miller, W., Hardisn, R. (2002) PipTools: A Computational Toolkit to Annotate and Analyze Pairwise Comparisons OF Genomic Sequences. Genomics, **80**,681-690.

124. Emanuel, B.S., Budarf, M.L., Shaikh, T. and Driscoll, D.A. (1998) Blocks of duplicated sequence define the endpoints of DGS/VCFS 22q11.2 deletions. Am. J. Hum. Genet. **63**, A11.

125. Evans, H.M., and Long, J.A. (1921) The effect of anterior lobe of the hypophysis administered intraperitoneally on growth, maturity and the estrous cycles of rat. Anat. Rec. 21, 61-63.

126. Everts-van der Wind, A., Kata, R.S., Band, R.M., Rebeiz, M., Larkin, M.D., Everts, E.R., Green, A.C., Liu, L., Natarajan, S., Goldammer, T., Lee, H.J., Mckay, S., Womack, E.J., and Lewin, A.H.(2004) An 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. Genome Res, 14, 1424-1437.

127. Ewing, B., Hillier, L., Wendl, M., Green, P (1998) Basecalling of automated sequencer traces using Phred I. Accuracy assessment.Gen.Res, **8**,175-185.

128. Fan,Y., Linardopoulou, E., Friedman, C., Williams, E., Trask, B.J. (2002) Genomic structure and evolution of the ancestral chromosome fusion site in 2q13-2q14.1 and paralogous regions on other human chromosomes. Genome Res, **12**, 1651-1662.

129. Farzer, K.A., Elinitski, L., Church, D.M., Dubchak, I., Hardison, R.C. (2003). Cross-species sequence comparisons: a review of methods and available resources. Genome Reszearch, **13**, 1-12.

130. Fedorov, A., Roy, S.,Fedorova, L. and Gilbert, W. (2003) Mystery of intron gain. Genome Res., **13**, 2236-2241.

131. Ferris, S.D., and Whitt, G.S (1979). Evolution of the differential regulation of duplicate genes after polyploidization. J.Mol. Evol. **12**, 267-317.

132. Fickett, W.J and Hatzigeorgiou, G.A. (1997) Eukaryotic promoter recognition. Gen Res, **7**, 861-878.

133. Fisher, R.A. (1935) The sheltering of lethals. Am. Nat. **69**, 446-455.

134. Fishman, M.C. (2001) Zebrafish: The canonical vertebrate. Science. **294**, 1290-1291.

135. Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R.J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K.C., McMorrow, T., Frampton J., Alter, B.P., Frischaul, A-M., and Higgs, D.R. (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the α globin cluster. Hum. Mol. Genet. **10**, 371-382.

136. Footz, K.T., Brinkman-Mills, P., Banting, S.G., Maier, A.S, Roe, A.B, McDermid, E.H et al. (2001) Analysis of the Cat Eye Syndrome Critical region in humans and the region of conserved synteny in mice: A search for candidate genes at or near the human chromosome 22 pericentromere. Genome. Res. **11**, 1053-1070.

137. Force, A et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics. **151**,1531-1545.

138. Fougerousse, F., Bullen, P., Herasse, M., Lindsay, S., Richard, I., Wilson, D., Suel, L., Durand, M., Robson, S., Abitbol, M., Beckmann, J.S., and Strachan, T. (2000) Human-mouse differences in expression patterns of developmental control genes and disease genes. Hum. Mol. Genet. **9**, 165-173.

139. Fraccaro, M., Lindsten, J., Ford, C.E., and Iselius, L. (1980) The 11q;22q translocation: a European collaborative analysis of 43 cases. Hum Genet, **56**, 21-51.

140. Francke, U. (1999) Williams-Beuren syndrome: genes and mechanisms. Hum. Mol. Genet, **8**, 1947-1954.

141. Frazer, A.K., Chen, X., Hinds, A, D., Pant, K.V.P., Patil, N., and Cox, R.D. (2003) Genomic DNA Insertions and Deletions Occur Frequently Between Humans and NonHuman Primates. Genome Res, **13**, 341-346.

142. Fries, R., and Ruvinsky, A. (1999) The genetics of cattle. CABI publishing, Oxon, UK. pp.710.

143. Fuhrmann G, Fayein N-A (1994) In situ hybridization with strand-specific DNA probes generated by a two-step polymerase chain reaction procedure. Anal Biochem **217**,161-163.

144. Fujiyama, A., Watanabe, H., Toyodo, A., Taylor, T.D., Itoh, T., Tsai, S.F., Park, H.S., Yaspo, M.L., Lehrach, H., Chen, Z., Fu, G., Saitou, N., Osoegawa, K., de Jong, P.J., Suto, Y., Hattori, M., and Sakaki, Y. (2002) Construction and analysis of a Human –Chimpanzee Comparative Clone Map. Science, **295**,131-134.

145. Funke B., Puech,A., Saint-Jore, B., Pandita, R., Skoultchi,A., and Morrow, B. (1998) "Isolation and characterization of a human gene containing a nuclear localization signal from the critical region for velo-cardio-facial syndrome on 22q11.'' Genomics, **53**(2), 146-154.

146. Funke, B., Epstein, J.A., Kochilas, L.K., Lu, M.M., Pandiat, R.K., Liao, J., Bauerndistel, R., Schuler, T., Schorle, H., Brown, M.C., Adams, J., Morrow, B.E. (2001) Mice overexpressing genes from the 22q11 region deleted in velo-cardio-facial syndrome/DiGeorge syndrome have middle and inner ear defects. Hum. Mol. Genet. **10**, 2549-2556.

147. Gagneux, P. and Varki, A. (2001) Genetic differences between humans and great apes. Mol Phylogenet Evol, **18**, 2-13.

148. Galas, D.J and McCormack, S.J. (2002) Genomic Technologies: Present and Future.Functional Genomics Series Volume1. Caister Academic Press, Wymondham, UK.345-374.

149. Galili, N., Baldwin, H.S., Lund,J., Reeves,R., Gong,W., Wang,Z., Roe, B.A., Emanuel, B.S., Nayak, S., Mickanin, C., Budarf, M.I., and Buck,C.A. (1997) A region of mouse chromosome 16 is syntenic to the DiGeorge velocardiofacial syndrome minimal critical region. Genome Res. **7**(1), 17-26.

150. Gall, J. G (1981) Chromosome structure and the C-value paradox. J. Cell. Biol, **91**, 3s-14s.

151. Ganguly, A., Dunbar, T., Chen, P., Godmilow, L., Ganguly, T. (2003). Exon skipping caused by an intronic insertion of a young Alu Yb9 element leads to severe hemophilia A. Hum. Genet. **113**, 348-352.

152. Garrity, D.M., Childs, S., and Fishman, M.C. (2002) The heartstrings mutation in zebrafish causes heart/fin TBX5 deficiency syndrome. Development. **129**. 4635-4645.

153. Garrod, A.E. (1908) Inborn errors of metabolism. Lancet, **2**, 1-7, 73-79, 142-148, 214-220.

154. Gawantka, V., Pollet, N., Delius, H., Vingron, M., Pfister, R., Nitsch, R., Blumenstock, C., and Niehrs, C. (1998) Gene expression screening of Xenopus identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. Mech. Dev. **77**, 95-141.

155. Gaxton, M. (1991) Linear amplification sequencing a powerful method of sequencing DNA. Methods: Companion methods in Enzymology. **3**, 20-26.

156. Geisler, R., Rauch, G-J., Baier, H., van Bebber, F., Brob, L., et al. (1999) A radiation hybrid map of the zebrafish genome. Nat. Genet. **23**, 86-89.

157. Gibbons, A. (1998) Which of our genes make us human? Science, **281**, 1432-1434.

158. Gibbons, A. (2000) Genomics. Building a case for sequencing the chimp. Science, **25**, 1267.

159. Gibbs. R and Weinstock, G. (2003). Bovine Genomic sequencing consortium. White paper.

160. Gilad, Y., Oshlack, A., Smyth, G.K., Speed, T.P., and White, K.P. (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. Nature, **440**, 242-245.

161. Glazko, G.V and Nei, M. (2003) Estimation of divergence time for major lineages of primate species. Mol. Biol. Evol. **20**, 424-434.

162. Godsave, S., Dekker, E.J Holling, T., Pannese, M., Boncinelli, E and Durston, A. (1994) Expression patterns of Hoxb genes in the Xenopus embryo suggest roles in anteroposterior specification of the hindbrain and in dorsoventral patterning of the mesoderm. Dev. Biol. **166**, 465-476.

163. Gogos, J.A., Santha, M., Takacs, Z., Beck, K.D., Luine, V., Nadler,J.V., and Karayiorgou, M. (1999) The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. Nat. Genet., **21**, 434-439.

164. Goidts, V., Szamalek, M.J., Hameister, H., and Kehrer-Sawatzki, H. (2004) Segmental duplication associated with the human specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates. Hum. Genet, **115**, 116-122.

165. Goldammer, T., Kata, R.S., Brunner, M.R, Dorroch, Ute, Sanftleben, H, Schwerin, M., and Womack, E.J. (2002) A comparative radiation hybrid map of bovine chromosome 18 and homologous chromosomes in human and mice. PNAS, **99**, 2106-2111.

166. Goldmuntz,E., Wang,Z., Roe, B.A., Budarf, M.L.(1996) Cloning, genomic, organization, and chromosomal localiztaion of human citrate transport protein to the DiGeorge/velocardiofacial syndrome minimal critical region. Genomics, **33**, 271-276.

167. Golling, G., Amsterdam, A., Sun, Z., Antonelli, M., Maldonado,E., Chen,W., Burgress, S., Haldi, M., Artzt, K., Farrington, S., Lin, S,Y., Nissen, R.M., and

Hopkins,N. (2002) Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nat.Genet, **31**,135-140.

168. Gomes, I., Sharma, T.T., Mahmud, N., Kapp, D.J., Edassery, S., Fulton, N., Liang, J., Hoffman, R., and Westbrook, A.C (2001). Highly abundant genes in the transcriptosome of human and baboon CD34 antigen-positive bone marrow cells. Blood, **98**, 93-99.
169. Goodman, M. (1999) The genomic record of Humankind's evolutionary roots. Am. J. Hum. Genet. **64**, 31-39.
170. Goodman, M., Tagle, D.A., Fitch, D.H., Bailey, W., Czelusniak, J., Koop, B.F., Benson, P., Slighton, J.L. (1990) Primate evolution at the DNA level and a classification of hominoids. J.Mol. Evol. **30**, 260-266.
171. Gordon, D., Abajian, C., and Green, Phil. (1998) Consed: A Graphical Tool for Sequence Finishing. Genome Res. **8**, 195-202.
172. Gottgens, B., Barton, L.M., Chapman, M.A., Sinclair, A.M., Knudsen, B., Grafham, D., Gilbert, J.G.R., Rogers, J., Bentley, D.R., and Green, A.R. (2002) Transcriptional regulation of the stem cell leukemia gene (SCL)-Comparative analysis of five vertebrate SCL loci. Genome Res, **12**, 749-759.
173. Gottgens, B., Barton, L.M., Gilbert, J.G.R., Bench, A.J., Sanchez, M-J., Bahn, S., Mistry, S., et al. (2000) Analysis of vertebrate SCL loci identifies conserved enhancers. Nature, **18**, 181-186.
174. Gottgens, B., Gilbert, J.G.R., Barton, L.M., Grafham, D., Rogers, J., Bentley, D.R., and Green, A.R. (2001). Long-range comparison of human and mouse SCL

loci: localized regions of sensitivity to restriction endonucleases correspond precisely with peaks of conserved noncoding sequences. Genome Res, **11**, 87-97.

175. Grantham, R. (1980). Workings of the genetic code. Trends Biochem. Sci. **5**, 327-331.

176. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. (1981) Codon catalog usuage is a genome strategy modulated for gene expressivity. Nucleic Acids Res., **9**, r43-74.

177. Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pave, A. (1980) Codon catalog usage and the genome hypothesis. Nucleic Acids Res. , **8**, r49-62.

178. Graves, J.A.M., Wakefield, M.J., Peters, J., Searle, A.J., Archibald, A., O'Brien, S.J., Womack, J.E. (1995) Report of the Committee on Comparative Gene Mapping. In: Cuticchia, A.J., Chipperfield, M.A., Foster, P.A (comps) Human Gene Mapping. Johns Hopkins University Press Baltimore, pp 1351-1408.

179. Greenberg, F., Guzzetta, V., Montes de Oca-Luna, R., Magenis, R.E., Smith, A.C., Richter, S.F., Kondo, I., Dobyns, W.B., Patel, P.I., and Lupski, J.R. (1991) Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del(17)(p11.2). Am. J. Hum. Genet, **64**, 747-758.

180. Greer, J.M., Puetz, J., Thomas, K.R., and Capecchi, M.R. (2000) Maintenance of functional equivalence during paralogous Hox gene evolution. Nature. 370, 304-307.

181. Gregory, T.R. (2001) Concidence, coevolution or causation? DNA content, cell size and the C-value enigma. Biol. Rev. Camb. Philos. Soc, **76**, 65-101.

182. Gridley, T., P. Soriano, and R. Jaenisch. 1987. Insertional mutagenesis in mice. Trends Genet, **3:** 162-166.

183. Griffith, F. (1928) The significance of Pneumoccocal types. J. Hyg, **27**, 113-159.

184. Grover, D., Majumder, P. P., Rao, B.C., Brahmachari, K, S., and Mukerji, M. (2003) Non random distribution of Alu elements in genes of various functional categories: insight from analysis of human chromosomes 21 and 22. Mol .Biol. Evol. **20**, 1420-1424.

185. Grunwald, D.J., and Eisen, J.S. (2002) Headwaters of the zebrafish- emergence of a new model vertebrate. Nat. Rev. Genet. **3**, 564-572.

186. Guris, D.L., Fantes, J., Tara, D., Druker, B.J and Imamoto, A. (2001) Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome. Nat. Genet., **27**, 293-298.

187. Hacia, G.J (2001) Genome of the apes. Trends in Genetics, **17**,637-645.

188. Haffter, P., and Nusslein-Volhard,C. (1996) Large scale genetics in a small vertebrate, the zebrafish. Int J Dev Biol. **40**(1), 221-227.

189. Haffter, P., Granato, M., Brand, M., Mullins, M.C., Hammerschmidt, M., Kane, D.A., Odenthal, J., van Eeden, F.J.M., Jiang, Y.L., Heisenberg, C.P., et al. (1996) The identification of genes with unique and essential functions in development of the zebrafish, Danio rerio. Development. **123**, 1-36.

190. Haldane, J.B.S. (1933) The part played by recurrent mutation in evolution. Am. Nat. **67**, 5-9.

191. Hall, B.D and Spiegelman, S. (1964) Sequence complementarity of T2-DNA and T2- specific RNA. Proc. Nat. Acad. Sci, **47**,137-146.

192. Hamdi, H., Nishio, H., Zelinski, R., and Dugaiczyk, A. (1999) Origin and phylogenetic distribution of Alu DNA repeats: Irreversible events in the evolution of primates. J. Mol. Biol, **289**, 861-871.

193. Hampsey, M and Reinberg, D. (1999) RNA polymerase II as a control panel for multiple coactivator complexes. Curr. Opin. Genet. Dev, **9**(2),132-139.

194. Handoko, H.Y., Nyholt, D.R., Hayward, N.K., Nertney, D.A., Hannah, D.E., Windus, L.C., McCormack, C.M., Smith, H.J., Filippich, C., James, M.R., and Mowry, B.J. (2005) Mol. Psychiatry, **10**(6), 589-587.

195. Hansson, C.M., Ali, H., Bruder, C.E.G., Fransson, I., Kluge, S., Andersson, B., Roe, B.A., Menzel, U., and Dumanski, J.P. (2003). Strong conservation of the human NF2 locus based on sequence comparison in five species. Mammalian Genome, **14**, 526-536.

196. Hardison, R.C., Oeltjen, J., and Miller, W. (1997) Long human-mouse sequence alignment reveal novel regulatory elements: A reason to sequence the mouse genome. Genome Res, **7**, 959-966.

197. Hare, M., and Palumbi, S.R. (2003) High Intron sequence conservation across three mammalian orders suggest functional constraints. Mol. Biol. Evol. **20** (6), 969-978.

198. Harrub, B., and Thompson, B. (2002) Do Human and Chimpanzee DNA indicate an evolutionary relationship? Apologetics Press, Montogomery, Alabama. Doc's Dissections. 1-9.

199. Hayakawa, T., Satta, Y., Gagneux, P., Varki, A, Takahat, N. (2001) Alu mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. Proc. Natl. Acad. Sci. USA. **98**, 1139-11404.

200. Hayata, I., Seki,M., Yoshida,K., Hirashima, K., Sado,T., Yamagiwa, J., and Ishihara, T. Chromosomal aberratons observed in 52 mouse myeloid leukemias. Cancer Res. **43**,367-373.

201. Hedges, J.D., Callinan, A.P., Cordaux, R., Xing, J., Barnes, E., and Batzer, A, M. (2004). Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res, **14**, 1068-1075.

202. Heisterkamp, N.; Mulder, M. P.; Langeveld, A.; ten Hoeve, J.; Wang, Z.; Roe, B. A.; Groffen, J. (1995) Localization of the human mitochondrial citrate transporter protein gene to chromosome 22q11 in the DiGeorge syndrome critical region. Genomics, **29**, 451-456.

203. Hellmann, I., Ebersberger, I., Ptak, S.E., Paabo, S., Przeworski, M. (2003a) A neutral explanation for the correlation of diversity with recombination rates in Humans. Am. J. Hum. Genet. **72**, 1527-1535.

204. Hellmann, I., Zollner, S., Enard, W., Ebersberger, I., Nickel, B., and Paabo, S. (2003) Selection on Human Genes as Revealed by Comparisons to Chimpanzee cDNA. Genome Res, **13**, 831-837.

205. Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S.A. (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. **25,** 3957-3958**.**

206. Hershey, A.D., and Chase,M. (1952) Independent functions of viral proteins and nucleic acid in growth of bacteriophage. J. Gen. Physiol, **36**, 39-56.

207. Hide, A.W., Babenko, V.N, van Heusden, P.A., Seoighe, G., and Kelso. J.F. (2001) The contribution of exon-skippig events on chromosome 22 to protein diversity. Genome Res. **11**, 1848-1853.

208. Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, **22**, 4673-4680.

209. Higley, J.D., Thompson, W.W., Champoux, M., Goldman, D., Hasert, M.F., Kraemer, G.W., Scalon, J.M. et al. (1993) Paternal and maternal genetic and environmental contributions to cerebrospinal fluid monoamine metabolites in rhesus monkeys. Arch Gen Psychiatry. **50**, 615-623.

210. Hinegardener, R., and Rosen, D. (1972). Cellular DNA content and the evolution of teleostean fishes. Am. Nat. **166**, 621-644.

211. Hisoaka, K.K and Battle, H.I. (1958) The normal development stages of the zebrafish, Brachydanio rerio. J.Morph. **102**, 311-323.

212. Hixson, E.J., Krammerer, M.C., Mott, E.G., Britten, L.M., Birnbaum, S., Poers, K.P and VandeBerg, L.J. (1993) Baboon Apolipoprotein A-IV. The Journal of Biological Chemistry, **268**, 15667-15673.

213. Hixson, J.E., Driscoll, D.M., Birnbaum, S., Britten, M.L. (1993a). Baboon lecithin cholesterol acyltransferase (LCAT): cDNA sequences of two alleles, evolution and gene expression. Gene. **128**, 295-299.

214. Holley, R. W., Appar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merell, S.H., Penswick, J.R and Zamir, A. (1965) Structure of ribonucleic acid. Science, **147**, 1462-1465.

215. Holt, R.A et al. (2002) The genome sequence of the malaria mosquito Anopheles gambiae. Science, **298**, 129-149.

216. Huang, X., and Miller, W. (1991) A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math. **12**, 337-357.

217. Hughes, A.L., and Nei, M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature, **335**, 167-170.

218. Hughes, F.J., and Cofffin, M.J. (2001) Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat Genet, **29,** 487-489.

219. Hughes, M.K., and Hughes, A.L. (1993) Evolution of duplicate genes in tetraploid animal, Xenopus laevis. Mol. Biol. Evol. **10**, 1360-1369.

220. Hukriede, N., Fisher, D., Epstein, J., Joly, L., Tellis, P., et al. (2001). The LN54 radiation hybrid map of zebrafish-expressed sequences. Genome Res. **11**, 2127-2132.

221. Hukriede, N.A., Joly, L., Tsang, M., Miles, J., Tellis, P et al. (1999) Radiation hybrid mapping of the zebrafish genome. Proc. Natl. Acad. Sci. USA. **96**, 9745-9750.

222. Hyatt,D., and Uberbacher, C.E (2002) Computational DNA sequence analysis and annotation. Genomic Technologies: Present and Future. Caister Academic Press, Wymondham, UK.345-374.

223. Ihara, N., Takasuga, A., Mizoshita, K., Takeda, H., Sugimoto, M., Mizoguchi, Y., Hirano, T., Itoh, T., Watanabe, T., Reed, M.K., Snelling, M.W., Kappes, M.S., Beattie, W.C., Bennett, L.G., and Sugimoto, Y. (2004) A comprehensive genetic map of cattle genome based on 3802 microsatellites. Genome Res, **14**, 1987-1998.

224. Ijdo, J.W., Baldini, A., Ward, D.C., Reeders, S.T., and Wells, R.A. (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. Proc. Natl. Acad. Sci. USA, **88**, 9051-9055.

225. Ingley, E., and Hemmings, B.A. (1994) Pleckstrin homology (PH) domains in signal transduction. J. Cell. Biochem. **56**(4), 436-443.

226. Ishii, J., Adachi, H., Aoki, J., Koizumi, H., Tomita, S., Suzuki, T., Tsujimoto,M., Inoue, K., and Arai, H. (2002) J. biol Chem , **277**(42), 39696-39702.

227. Ishikawa, Y. (2000) Medakafish as a model system for vertebrate developmental genetics. Bioessays. **22**, 487-495.

228. Jacob, F and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol, **3**, 318-356.

229. Jaillon,O et al.(2004) Genome duplication in the teleost fish Tetraodon nigroviridis reveals early vertebrate proto-karyotype. Nature, **431**, 946-957.

230. Jayashankar, L., Brasky, M.K., Ward, A.J., and Attanasio, R. (2003) Lymphocyte modulation in a baboon model of immunosenescence. Clinical and Diagnostic Laboratory Immunology, **10**, 870-875.

231. Jeanmougin, F., Thompson J.D., Gouy, M, Higgins, D.G, Gibson, T.J.(1998) Multiple sequence alignment with Clustal X. Trends Biochem. Sci, **23**,403-405.

232. Jerome, C.P, Kimmel, D.B., McAlister, J.A., Weaver, D.S. (1986) Effects of ovariectomy on iliac trabecular bone in baboons. Cacif Tissue Int, **39**, 206-208.

233. Ji, Y., Eichler, E.E, Schwartz, S., and Nicholls, D.R (2000) Structure of chromosomal duplicons and their role mediating human genomic disorders. Genome Res, **10**, 597-610.

234. Johnson, S.D., Africa, D., Horne, S., and Postlethwait. (1995) Half-tetrad analysis of zebrafish: Mapping the ros-mutation and centromere of linkage I group. Genetics. **139**, 1725-1735.

235. Jorgensen, L.A., Laursen, B.H., Jones, C and Bak, L.A. (1992). Evolutionary different alphoid repeat DNA on homologous chromosomes in human and chimpanzee. Proc. Natl, Acad. Sci, **89**, 3310-3314.

236. Jun An, H., Na, D., Lee, Doheon., Lee, H, K., and Bhak, J. (2004). High correlation between Alu elements and the conversion of 3'UTR of mRNAs processed pseudogenes. Genomics & Informatics, **2**, 86-91.

237. Jurka J, Zuckerkandl E. (1991) Free left arms as precursor molecules in the evolution of Alu sequences. J Mol Evol.**33**(1),49–56.

238. Jurka, J. (1993) A new family of recently retroposedAlu repeats. Nucleic Acids Res, **21**, 2252.

239. Jurka, J. (2004) Evolutionary impact of human Alu repetitive elements. Curr. Opin. Genet. Dev. **14**, 603-608.

240. Juyal, R.C., Figuera, L.E., Hauge, X., Elsea, S.H., Lupski, J.R., Greenberg, F., Baldini, A., and Patel, P.I. (1996) Molecular analyses of 17p11.2 deletions in 62-Smith-Magenis syndrome patients. Am. J. Hum. Genet, **58**, 998-1007.

241. Kaessmann, H., Wiebe, V., Weiss, G and Paabo, S. (2001) Great ape DNA sequence reveal a reduced diversity and an expansion in humans. Nat Genet, 27, 155-156.

242. Kammerer, C.M., Sparks, M.L., and Rogers, J. (1995) Effects of age, sex and heredity on measures of bone mass in baboons. J Med Primatol, **24**, 236-242.

243. Kammerer, M.C, Rainwater, L.D., Cox, A.L., Schneider, L.J., Mahaney, C.M., Rogers, J and VandeBerg, L.J. (2002) Locus controlling LDL cholesterol reponse to dietary cholesterol is on a baboon homologue of human chromosome 6. Arterioscler Thromb Vasc Biol, **22**, 1720-1725.

244. Kaplan, J.R, Fontenot, M.B., Berard. J., Manuck, S.B., Mann, J.J. (1995) Delayed dispersal and elevated monoaminergic activity in free-ranging rhesus monkeys. Am J Primatol. **35**, 229-234.

245. Karaman, M.W., Houck, M.L., Chemnick,L.G., Nagpal,S., Chawannakul,D., Sudano,D., Pike,B.L., Ho, V.V., Ryder,O.A., and Hacia, J.G. (2003). Comparative analysis of gene-expression patterns in human and African ape cultured fibroblasts.Genome Res, **13**, 1619-1630.

246. Karlovich, C.A., John, R.M., Ramirez, L., Stainier, D.Y. and Myers, R.M. (1998) Characterization of the Huntington's disease (HD) gene homologue in the zebrafish Danio rerio. Gene, **217**, 117–125.

247. Kataoka, T., Holler, N., Micheau, O., Martinon, F., Tinel, A.,Hofmann, K., Tschopp, J. (2001) Bcl-rambo, a novel Bcl-2 homologue that induces apoptosis via its unique C-terminal extension. J Biol Chem. **276**(22),19548-19554.

248. Kawai J et al., (2001) "Functional annotation of a full-length mouse cDNA collection." Nature, **409**(6821),685-90.

249. Kehrer-Sawatzki, H., Schreiner, B., Taenzer, S., Platzer, M., Muller, S., Hameister, H. (2002) Molecular characterization of the pericentric inversion that causes differences between chimpanzee chromosome 19 and human chromosome 17. Am J Hum Genet, **71**, 375-388.

250. Keller, M.P., Seifried, B.A., and Chance, P.F. (1999) Molecular evolution of CMT1A-REP region: a human- and chimpanzee- specific repeat. Mol. Biol. Evol. **16**, 1019-1026.

251. Kelly, P.D., Chu, F., Woods, I.G., Ngo-Hazelett, P., Cardozo, T et al. (2000) Genetic linkage mapping of zebrafish genes and ESTs. Genome Res. **10**, 558-567.

252. Kent, W.J and Zahler, A.M. (2000) Conservation, regulation, synteny, and introns in a large-scale C. briggsae-C. elegans genomic alignment. Genome Res, **10**, 1115-1125.

253. Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H-.H., Weiss, G., Enard, W., Heissig, F., Arendt, T., Nieselt-Struwe, Kay., Eichler, E.E., and Paabo, S. (2004). Regional patterns of gene expression in human and chimpanzee brains. Genome Res, **14**, 1462-1473.

254. Khaitovich,P., Muetzel, B., She, X., Lachmann, M., Eichler,E., Paabo, S et al. (2004) Regional patterns of gene expression in human and chimpanzee brains. Genome Res. **14**,1462-1473.

255. Kidwell, M.G. 1986. P-M mutagenesis. In Drosophila: A practical approach (ed. D.B. Roberts), pp. 59-81. IRL Press, Oxford,UK.

256. Kim, H.S., Wadekar, R.V., Takenaka, O., Hyun, B.H., and Crow, T.J. (1999) Phylogenetic analysis of a retroposon family in African great apes. J. Mol. Evol. **49**, 699-702.

257. Kimmel, C. B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995) Stages of embryonic development of zebrafish. Dev. Dyn. **203**,253-310.

258. Kimmel, C.B. (1989) Genetics and early development of zebrafish. Trends Genet. **5**:283-288.

259. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995) Stages of embryonic development of the zebrafish. Dev. Dyn. **203**, 253-310.

260. Kimmel, C.B., Kane, D.A., and Ho, R.K. (1991a) Lineage specification during early development of the zebrafish. In Cell-Cell interactions in early development. New York: Wiley-Liss, Inc., pp.203-225.

261. Kimmel, C.B., Kane, D.A., Walker, C., Warga, R.M and Rothman, M.B. (1989) A mutation that changes cell movement and cell fate in the zebrafish embryo. Nature. 337, 358-362.

262. Kimmel, C.B., Schilling, T.F and Hatta. K.(1991b) Patterning of body segments of the zebrafish embryo. Curr. Topics. Dev. Biol. 25, 77-110.

263. Kimura, M.(1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature,267(5608),275-276.

264. King, M.-C and Wilson, A.C. (1975) “Evolution at two levels in humans and chimpanzees”. Science. 188, 107-116.

265. Kitazawa, S., Kitazawa, R., and Maeda, S. (1999) In situ hybridization with polymerase chain reaction-derived single stranded DNA probe and S1 nuclease. Histochem. Cell. Biol, 111,7-12.

266. Klinger, M., Taylor, J.S., Oertle, T., Schwab, M.E., Stuermer, C.A.O., and Diekmann, H. (2004) Identification of Nogo-66 Receptor and homologous genes in fish. Mol. Biol. Evol. 21, 76-85.

267. Knapik, E.A., Goodman, O., Atkinson, C., Roberts, M., Shiozawa, C., Sim, S., Weksler-Zangen, Trolliet, M., Futrell, C et al. (1996) A reference cross DNA panel for zebrafish (Danio rerio) anchored with simple sequence length polymorphism. Development. 123, 451-460.

268. Knebelmann, B., Forestier,L., Drouot, L., Quinones,S., Chuet,C., Benessy, F., Saus, J., and Antignac, C. (1995) Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive A1port syndrome. Hum. Mol.Genet. 4(4),675-679.

269. Knuchel, M.C., Graf, B., Schlaepfer, E., Kuster, H., Fischer, M., Weber, R., and Cone, R. W. (2000) PCR-derived ssDNA probes for fluorescent in situ hybridization to HIV-1 RNA. J.Histochem. Cytochem, 48,285-292.

270. Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., et al. (2000) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development. 127, 1737-1749.

271. Kochilas, L.K., Potluri, V., Gitler, A., Balasubramanian,K., and Chin, A.J. (2003) Cloning and characterization of zebrafish tbx1. Gene Expr Patterns, 3(5), 645-651.

272. Korf, I., Flicek, P., Duan, D., and Brent, M.R. (2001) Integrating genomic homology into gene structure prediction . Bioinformatics, 17, S140-S148.

273. Kozak, M. (1989). The scanning model for translation: an update. J. Cell. Biol. 108, 229-241.

274. Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. Gene 234(2), 187-208.

275. Kreahling, J., and Graveley, B.R. (2004) The origins and implications of Aluternative splicing. Trends Genet. 20, 1-4.

276. Kudoh, T., Tsang, M., Hukriede, N.A., Chen,X., Dedekian, M., Clarke, C.J., Kiang, A., Schultz, S., Epstein, A.J., Toyama, R and Dawid, I.B. (2001). A gene expression in zebrafish embryogenesis. Genome Res.11, 1979-1987.

277. Kumar, S., and Hedges, S.B. (1998) A molecular timescale for vertebrate evolution. Nature, 392, 917-920.

278. Kurtz, S., Choudhari, V.J., Ohlebusch, E., Schleiermacher, C., Stoye, J and Giegerich, R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale Nucleic Acids Research, 29, 4633-4642.

279. Lander, E.S. and Waterman, M.S., 1988, "Genomic mapping by fingerprinting random clones:a mathematical analysis," Genomics, 2, 231-239.

280. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001) Initial sequencing of the human genome. Nature, 409, 860-921.

281. Larkin, M.D., Everts-van der Wind., Rebeiz, M., Schweitzer, A.P., Bachman, S., Green, C., Wright, L.C., Campos, J.E., Benson, D.L., Edwards, J., Liu, L., Osoegawa, Kazutoyo, Womack, E.J., de Jong, J.P ., and Lewin, A.H. (2003) A cattle-human comparative map built cattle BAC-Ends and human genome sequence. Genome Res, 13, 1966-1972.

282. Lash, A.E., Tolstoshev, C.M., Wagner,L., Schuler, G.D., Strausberg, R.L., Riggins, G.J., and Altschul, S.F. (2000) SAGEmap:A Public Gene Expression Resource. Genome Res. 10(7),1051-1060.

283. Lee, K.H., Xu, Q.H and Breitbart, R.E. (1996) A new tinman related gene, nk2.7, anticipates the expression of nkx2.5 and nkx2.3 in zebrafish heart and Pharyngeal endoderm. Dev. Biol. 180, 191-201.

284. Leimer, U., Lun, K., Romig,H., Walter,J., Grunberg,J., Brand, M., and Haass, C. (1999) Zebrafish (Danio rerio) presenilin promotes aberrant amyloid beta-peptide production and requires a critical aspartate residue for its function in amyloidogenesis. Biochemistry. 38(41), 13602-13609.

285. Lewin, B. 2000. Gene VII Oxford University Press, NY.

286. Li, W.-H and Graur, D. (1991) Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA.

287. Li, W.H., Gu, Z., Wang,H., and Nekrutenko, A.(2001) Evolutionary analyses of the human genome. Nature, 409, 847-849.

288. Li, W-H. (1997) Molecular evolution. Sinauer Associates, Inc. Sunderland, MA.

289. Lindsay, E.A., Harvey, E.L., Scambler, P.J., and Baldini, A. (1998) ES2, a gene deleted in DiGeorge syndrome, encodes a nuclear protein and is expresssed during early mouse development, where it shares an expression domain with a Goosecoid-like gene. Hum. Mol. Genet., 7, 629-635.

290. Lindsay, E.A., Vitelli, F., Su,H., Morishima,M., Huynh,T., Pramparo,T., Jurecic, V., Ogunrinu, G., Sutherland, H.F., Scambler,P.J., Bradley,A., Baldini,A. (2001) Tbx1 haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. Nature, 410, 97-101.

291. Lipovich, L., and King, C-.M. (2003). Novel transcriptional units and unconventional gene pairs in the human genome: Toward a sequence-level basis for primate-specific phenotypes. Cold Spring Harb Symp Quant Biol, 68,461-470.

292. Liu, G., NISC Comparative Sequencing Program, Zhao, S., Bailey, A.J., Sahinalp, C.S., Alkan, C., Tuzun, E., Green, D.E., and Eichler, E.E. (2003). Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res, 13, 358-368.

293. Lui, H., Heath, S.C., Sobin, C., Roos, J.L., Galke, B.L., Blundell, M.L., Lenane, M., Robertson, B., Wijsman, E.M., Rapoport, J.L., Gogos, J.A., and Karayiorgou, M. (2002) Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. PNAS, 99(6), 3717-3722.

294. Liu, L., and McKeehan, W.L (2002) Sequence analysis of LRPPRC and its SEC1 domain interaction partners suggest roles in cytoskeletal organization, vesicular trafficking, nucleocytosolic shuttling, and chromosome activity. Genomics, 79, 124-136.

295. Lo, J., Sorcheng, L., Xu, M., Liu, F., Ruan, H., Eun, A., He, Y., Ma, W., Wang, W., Wen, Z., and Peng, J. (2003) 15, 000 unique zebrafish EST clusters and their future use in microarray for profiling gene expression patterns during embryogenesis. Genome Res. 13, 455-466.

296. Locke, D.P., Archidiacono, N., Misceo, D., Cardone, M.F, Deschamps, S., Roe, B., Rocchi, M., Eichler, E.E. (2003) Refinement of a chimpanzee pericentric inversion breakpoint to a segmental duplication cluster. Genome Biol, 4, R50.

297. Long, M., Betran, E., Thornton, K and Wang, W. (2003) The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. 4,865-875.

298. Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., and Frazer, K.A. (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science, 288, 136-140.

299. Loots, G.G., Ovacharenko, I., Pachter, L., Dubchak, I., and Rubin, E.M. (2002). rVista for comparing sequence-based discovery of functional transcription factor binding sites. Genome Res, 12, 832-839.

300. Lund, J., Chen,F., Hua, A., Roe, B., Budarf, M., Emanuel, B.S., and Reeves, R.H (2000) Comparative sequence analysis of 634kb of the mouse chromosome 16

region of conserved synteny with the human velocardiofacial syndrome region on chromosome 22q11.2. Genomics, 63(3), 378-383.

301. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA arrangements and human disease traits. Trends Genet, 14, 417-422.

302. Lynch, M & Force, A. (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics. 154, 459-473.

303. Lynch, M., O'Hely, M., Walsh, B., and Force, A. (2001) The probability of preservation of a newly arisen gene duplicate. Genetics. 159, 1789-1804.

304. MacCluer, J.W., Kammerer, C.M., Blangero, J., Dyke, B., Mott, G.E., VanderBerg J.L., McGill, H.C Jr. (1988) Pedigree analysis of HDL cholesterol concentration in baboons on two diets. Am. J Hum Genet. 43,401-413.

305. Magness, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy,M.B., Proll, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G., and Iadonato, S.P. (2005) Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human. Genome Biology, 6, R60.

306. Mahaney, M.C., William-Blangero, S., Blangero, J., Leland, M.M. (1993) Quantitative genetics of relative organ weight variation in captive baboons. Hum Biol. 65, 991-1003.

307. Mardis, E.R and Roe, B.A (1989) Automated Mehods for single stranded DNA isolation and dideoxynucleotide DNA sequencing reactions on robotic workstation. BioTechniques. 5, 840-849.

308. Martin, C.C., Laforest, L., Akimenko, M.A., and Ekker, M. (1999) A role for DNA methylation in gastrulation and somite patterning. Dev Biol, 206 (2), 189-205.

309. Martin, J.L., Mahaney, C.M., Bronikowski, M.A., Carey, D.K., Dyke, B., Comuzzie, G.A. (2002) Lifespan in captive baboons is heritable. Mechanisms of Ageing and Development, 123, 1461-1467.

310. Martin-Galiano,A.J., Wells, J.M., and de la Campa, A.G. (2004) Relationship between codon biased genes, microarray expression values and physiological characteristics of Streptococcus pneumoniae. Microbiology, 150,2313-2325.

311. Matera, A.G., Hellmann, U., Hintz, M.F., and Schmid, C.W. (1990) Recently transposed Alu repeats result from multiple source genes. Nucleic. Acids. Research, 18(20), 6019-6023.

312. Matsumoto, N.; Tamura, S.; Furuki, S.; Miyata, N.; Moser, A.; Shimozawa, N.; Moser, H. W.; Suzuki, Y.; Kondo, N.; Fujiki, Y. (2003) Mutations in novel peroxin gene PEX26 that cause peroxisome-biogenesis disorders of complementation group 8 provide a genotype-phenotype correlation. Am. J. Hum. Genet. 73, 233-246.

313. Maxam, A.M., and Gilbert, W. (1977) A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74(2), 560-564.

314. Maynard, T.M., Haskell, G.T., Peters, A.Z., Sikich, L., Lieberman, J.A., and LaMantia, A.-S.(2003) A comprehensive analysis of 22q11 gene expression in the developing and adult brain. PNAS, 100 (24), 14433-14438.

315. Mayor, C., Brudno, M., Schwartz, J.R., Polialov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I. (2000). VISTA:visualizing global DNA sequence alignments of arbitrary length. Bioinformatics, 16, 1046-1047.

316. McBride, L.J., Koepf, S.M., Gibbs, R.A., Salser, W., Mayrand, P.E., Hunkapiller, N.W., Kronick, M.N. (1989). Automated DNA sequencing methods involving polymerase reaction. Clin. Chem. 35, 2196-2201.

317. McClintock, J.M., Kheirbek, M.A., and Prince, V.E. (2002) Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. Dev. 129, 2339-2354.

318. McConkey, E.H., and Varki, A. (2000) A primate genome project deserves high priority. Science, 289, 1295-1296.

319. McDermid, H.E., and Morrow, B.E. (2002). Genomic Disorders. Am. J. Hum. Genet, 70, 1077-1088.

320. McDermid, H.E., Duncan, A.M.V., Brasch, K.R., Holden, J. J.A, Magenis, E, Sheehy, R., Burn, J., Kardon, N., Noel, B., Schinzel, A.(1986) Characterization of the supernumerary chromosome in cat eye syndrome. Science, 232, 646-648.

321. McGinnis, W and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell. 68, 283-302.

322. McTaggart, K.E., Budarf, M.L., Driscoll, D.A., Emanel, B.S., Ferreira, P and McDermid, H.E. (1998) Cat eye syndrome chromosome breakpoint clustering: identification of two intervals also associated with 22q11 deletion syndrome breakpoints. Cytogenet. Cell Genet., 81, 222-228.

323. Mead, D.A., McClary, J.A., Luckey, J.A., Witney, F.R and Smith, L.M. (1991). Bst DNA polymerase permits rapid sequence analysis from nanogram amount of template. BioTechniques. 5, 840-849.

324. Merscher, S., Funke, B., Epstein, J.A., Heyer, J., Puech, A., Lu, M.M., Xavier, R.J., Demay, M.B., Russell, R.G., Factor., S., Tokooya, K., St. Jore, B., et al. (2001) TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. Cell, 104, 619-629.

325. Micklos, A.D and Freyer, A.G. (1990) DNA science: A first course in recombinant DNA technology. Colds Spring Harbor Laboratory Press, Colds Spring Harbor, NY.

326. Miki, Y.(1998) Retrotransposal integration of mobile genetic elements in human diseases. J. Hum.Genet. 43(2),77-84.

327. Minghetti, P.P and Dugaiczyk, A. (1993) The emergence of new DNA repeats and the divergence of primates. Proc. Natl. Acad. Sci, 90, 1872-1876.

328. Mitchell, G.A., Labuda, D., Fontaine, G., Saudubray, J.M., Bonnefont, J.P., Lyonnet,S., Brody, L.C., Steel,G., Obie, C., Valle, D.(1991)Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation. Proc. Natl. Acad. Sci.USA.88(3), 815-819.

329. Mizusawa, S., Nishimura, S., and Seela,F.(1986)Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res.14,1319-1324.

330. Moran, J.V, DeBerardinis, R.J, Kazazian, H.H.Jr. (1999) Exon shuffling by L1 retrotransposition. Science. 283, 1530-1534.

331. Morgan, T. H. (1910) Sex-linked inheritance in Drosophila. Science, 32, 120-121.

332. Moriyama, E.N., and Powell, J.R. (1997) Codon usage bias and tRNA abundance in drosophila. J.Mol.Evol., 45, 514-523.

333. Muchmore, E.A., Diaz, S., and Varki, A.(1998) A structural difference between cell surfaces of humans and the great apes. Am. J. Phys. Anthropol. 107,187-198.

334. Mukai, J., Liu,H., Burt, R.A., Swor, D.E., Lai, W.S., Karayiorgou, M., Gogos,J.A. (2004). Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia. Nat. Genet. 36(7), 725-731.

335. Mullins, M.C., M. Hammerschmidt, P. Haffter, and C. Nusslein-Volhard (1994). Large-scale mutagenesis in the zebrafish: In search of genes controlling development in a vertebrate. Curr. Biol., 4, 189-202.

336. Mural, R.J., Adams, M.D., Myers, E.W., Smith, H.O., Miklos, G.L., Wides, R., Halpern, A., Li, P.W., Sutton, G.G., Nadeau, J., et al. (2002). A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. Science 296: 1661-1671.

337. Myers, E.W and Miller, W. (1989) Approximate matching of regular expressions. Bull. Math. Biol, 51, 5-37.

338. Nagase, T., Ishikawa, K., Kikuno, R., Hirosawa, M., Nomura, N., and Ohara,O.(1999) Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 newcDNA clones from brain which code for large proteins in vitro. DNA Res.,6,337-345.

339. Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K., Ohara, O. (2000) Prediction of the coding sequences of unidentified human genes. XIX. The

complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res. 7,347-355.

340. Nakagawa, T., Goto, K., and Kondo,H. (1996) Cloning, expression, and localization of 230-kDa phosphatidylinositol 4-kinase. J.Biol. Chem. 271, 12088-12094.

341. Nasevicius, A., and Ekker, S.C. (2000) Effective targeted gene 'knockdown' in zebrafish. Nat. Genet. 26, 216-220.

342. Nei, M., Gu, X., and Sitnikova, T. (1997) Evolution by the birth-and-death process in multigene families of vertebrate immunesystem. Proc. Natl. Acad. Sci. USA. 94, 7799-7806.

343. Neidhardt L, Gasca S, Wertz K, Obermayr F, Worpenberg S, et al. (2000) Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. Mech Dev, 98 (1-2), 77–94.

344. Nergadze, G.S., Rocchi, M., Azzalin, M. C., Mondello, C., and Giulotto, E. (2004) Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. Genome Res, 14, 1704-1710.

345. Neuhauss, C.F., Solnica-Krezel, L., Schier, A.F., Zwartkruis, F., Stemple, D.L., Malicki, J., Abdelilah, S., Stainier, D.Y.R., and Driever, W. (1996) Mutation affecting craniofacial development in zebrafish. Development. 123,357-367.

346. Nobrega, M.A and Pennachio, L.A. (2003) Comparative genome analysis as a tool for biological discovery. J Physiol, 554.1, 31-39.

347. Nobukuni, T., Kobayashi, M., Omori, A., Ichinose, S., Iwanaga, T., Takahashi, L., Hashimoto, K., Hattori, S., Kaibuchi, K., Miyata, Y., Masui, T., Iwashita, S. (1997) An Alu-linked repetitive sequence corresponding to 280 amino acids is expressed in a novel bovine protein, but not in its human homologue. J. Biol. Chem. 272, 2801-2807.

348. Notredame, C., Higgins, D., Hesinga, J (2000). T-coffee:A novel method for multiple sequence alignments. Journal of Molecular biology. 302:205-217.

349. O'Brien, S.J., Menotti-Raymond, M., Murphy, W.J., Nash, W.G., Weinberg, J., Stanyon, R., Copeland, N.G., Jenkins, N.A., Womack, J.E., Graves, J.A.M. (1999). The promise of comparative genomics in mammals. Science, 286, 458-481.

350. Oates, A.C., Brownie, A., Pratt, S.J., Irvine, D.V., Liao, E.C., Paw, B.H., Dorian, K.J., Johnson, S.L., Postlethwait, J.H., Zon, L.I and Wilks, A.F. (1999) Gene duplication of zebrafish JAK2 homologs is accompanied by divergent expression patterns: Only jak2a is expressed during erythropoiesis. Blood. 94, 2622-2636.

351. Oates, A.C., Wollberg, P., Postlethwait, J., Zon, L.I et al. (1999) Zebrafish stat3 is expressed in restricted tissues during embryogenesis and stat1 rescues cytokine signaling in a STAT-1 deficient human cell line. Dev. Dynamics. 215, 352-370.

352. Ocharenko, I., Boffelli, D and Loots, G.G (2004) eShadow: A Tool for Comparing Closely Related Sequences. Genome Research, 14, 1191-1198.

353. Oeltjen, J.C., Malley, T.M., Muzny, D.M., Miller, W., Gibbs, R.A., and Belmont, J.W. (1997) Large-scale comparative sequence analysis of the human and murine

Bruton's tyrosine kinase loci reveals conserved regulatory domains. Genome Res. 7, 315-329.

354. Oh, Y.S., Lee,S., Won,C., Warnock, D.G. (2001) An Alu cassette in the human epithelial sodium channel. Biochim BioPhys Acta. 1520(1), 94-98.

355. Ohno, S. (1970) Evolution by Gene duplication. Springer-Verlag, Berlin-Heidelberg-NewYork.

356. Ohtsuka, M., Kikuchi, N., Ozato, K., Inoko, H., and Kimura, M. (2004) Comparative analysis of a 229-kb medaka genomic region containing the zic1 and zic4 genes with Fugu, human and mouse. Genomics, 83, 1063-1071.

357. Olson, M.V and Varki, A. (2003) Sequencing the chimpanzee genome: Insights into human evolution and disease. Nature. 4, 20-28.

358. Onyango, P., Miller, W., Lehoczky, J., Leung, C.T., Birren, B et al. (2000). Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. Genome Res, 10, 1697-1710.

359. Ovcharenko, I., Boffelli, D., and Loots, G. (2004) eshadow: A tool for comparing closely related sequences. Genome. Res., 14(6), 1191-1198.

360. Ovcharenko, I., Loots, G.G., Hardison, C.R., Miller, W and Stubbs,L (2004) zPicture: Dynamic Alignment and Visualization Tool for Analyzing Conservation Profiles. Genome Research, 14, 472-477.

361. Pack, M., Solnica-Krezel, L., Malicki, J., Nuehauss, S.C.F., Schier, A.F, et al. (1996) Mutations affecting development of zebrafish digestive organs. Development. 123, 321-328.

362. Pagani, F., Raponi, M., and Baralle, F.E. (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. PNAS, 102(18), 6368-6372.

363. Patel, N.H and Goodman, C.S. (1992) Preparation of digoxigenin labeled single-stranded DNA probes. In Non-radioactive labeling and detection of biomolecules (ed. C. Kessler). Springer-Verlag, Berlin, Germany.p.377.

364. Paule, M.R., and White, R.J (2000). Survey and summary: Transcription by RNA polymerase I and III. Nuclei Acids Res, 28(6),1283-1256.

365. Paylor, R., McIlwain, K.L., McAninch, R., Nellis, A., Yuva-Paylor, L.A., Baldini, A., and Lindsay, E.A. (2001) Mice deleted for the DiGeorge/velocardiofacial syndrome region show abnormal sensorimotor gating and learning and memory impairments. Hum. Mol. Genet., 10, 2645-2650.

366. Pennacchio, L.A and Rubin, E.M. (2001) Genomic strategies to identify mammalian regulatory sequences. Nat. Rev. Genet, 2, 100-109.

367. Pijnacker, L.P., and Ferwerda M.A. (1995) Zebrafish chromosome banding. Genome. 38, 1052-1055.

368. Piotrowski, T., Ahn, D-g., Schilling, T.F., Nair, S., Zon, L.I et al. (2003) The zebrafish van gogh mutation disrupts tbx1, which is involved in the DiGeorge deletion syndrome in humans. Development. 130, 5043-5052.

369. Pizzuti, A., Novelli, G., Ratti, A., Amati, F., Bordoni, R., Mandich, P., Bellone, E., Conti, E., Bengala, M., Mari, A., Silani, V., and Dallapiccola, B. (1999) Isolation and characterization of a novel transcript embedded within HIRA, a gene deleted in DiGeorge syndrome. Mol. Genet. Metab., 67(3), 227-235.

370. Pizzuti, A., Novelli, G., Ratti, A., Amati, F., Mari, A., Calabrese, G., Nicolis, S., Silani, V., Marino, B., Scarlato, G., Ottolenghi, S and Dallapiccola, B.(1997) UFD1L, a developmentally expressed ubiquitination gene, is deleted in CATCH 22 syndrome. Hum. Mol. Genet., 6, 259-265.

371. Plageman, T.F and Yutzey, K..E. (2005) T-Box genes and heart development :putting the "T" in HearT. Developmental Dynamics, 232, 11-20.

372. Pletcher, M.T., Roe, B.A , Chen, F., Do, T., Do, A., Malaj, E., and Reeves, R.H. (2000) Chromosome evolution: The junction of mammalian chromosomes in the formation of mouse chromosome 10. Genome Res. 10, 1463-1467.

373. Pontius J.U., Wagner, L., Schuler G.D. (2003) UniGene: a unified view of the transcriptome. In: The NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information.

374. Postlethwait, J and Talbot, W. (1997) Zebrafish genomics: From mutants to genes. Trends Genet. 13, 183-190.

375. Postlethwait, J., Johnson, S., Midson, C.N., Talbot, W.S., Gates, M., Ballenger, E.W., Africa, D., Andrewd, R., Carl, T., Eisen, J.S., et al. (1994) A genetic linkage map for the zebrafish. Science. 264, 699-703.

376. Postlethwait, J., Yan, Y., Gates, M., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E., Force, A., Gong, Z., et al. (1998) Vertebrate genome evolution and the zebrafish gene map. Nat. Genet. 18, 345-349.

377. Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y-L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., and Talbot, W. (2000) Zebrafish comparative

genomics and the origins of vertebrate chromosomes. Genome Res. 10-1890-1902.

378. Prescott, K., Ivins, S., Hubank,M., Lindsay, E., Baldini, A., and Scambler, P. (2005) "Microarray analysis of the Df1 mouse model of the 22q11 deletion syndrome." Hum. Genet, 116, 486-496.

379. Prince,V.E., and Picket, F.B. (2002) Splitting pairs:Diverging fates of duplicated genes. Nature Reviews Genet. 3, 827-837.

380. Proudfoot, N.(2000) Connecting transcription to messenger RNA processing. Trends. Biochem. Sci. 25(6), 290-293.

381. Puech,A., Saint-Jore, B., Merscher, S., Russell, R.g., Cherif,D., Sirotkin, H., Xu, H., Factor, S., Kucherlapati, R., and Skoultchi, A.I. (2000) Normal cardiovascular development in mice deficient for 16 genes in 550kb of the velocardiofacial/DiGeorge syndrome region. Proc. Natl Acad. Sci. USA, 97, 10090-10095.

382. Quentin, Y. 1992. Origin of the Alu family: A family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. Nucleic Acids Res. 20: 3397-3401.

383. Raft, S., Nowotschin, S., Liao, J., Morrow, B.E. (2003) Supression of neural fate and control of inner ear morphogenesis by Tbx1. Development, 131, 1801-1802.

384. Raft,S., Nowotschin,S., Liao,J., and Morrow,B.E. (2004) Suppression of neural fate and control of inner ear morphogenesis by TBX1. Development, 131(8), 1801-1812.

385. Ransom, D.G., Haffter, P., et al. (1996) Chracterization of zebrafish mutants with defects in embryonic hematopoiesis. Development. 123, 311-319.

386. Raymond, P.A., Barthel, L.K., and Curran, G.A.(1995) Developmental patterning of rod and cone photoreceptors in embryonic zebrafish. J.Comp. Neurol., 359(4), 537-550.

387. Reeves, R., Nissen, M.S. (1990) The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. J.Biol.Chem. 265, 8573-8582.

388. Regnier, V., Meddeb, M., Lecointre, G., Richard, F., Duverger, A., Nguyen, V.C., Dutrillaux, B., Berheim, A., Danglot, G.(1997) Emergence and scattering of multiple neurofibromatosis (NF1)- related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. Hum. Mol. Genet. 6,9-16.

389. Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C and Rozen, S. (2002). Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am. J. Hum. Genet., 71, 906-922.

390. Ricci, V., Regis, S., Di Duca,M., Filocamo, M. (2003) An Alu-mediated rearrangement as cause of exon skipping in Hunter disease. Hum. Genet. 112(4), 419-425.

391. Rijnkels, M., Elnitski, L., Miller, W and Rosen, J.M. (2003) Multispecies comparative analysis of a mammalian- specific genomic domain encoding secretory proteins. Genomics, 83, 417-432.

392. Rithidech, K. N., J. J. Dunn, and C. R. Gordon. (1997) Combining multiplex and touchdown PCR to screen murine microsatellite polymorphisms. BioTechniques 23:36-45.

393. Rithidech,K., Dunn, J.J., Roe,B.A., Gordon,C.R., and Cronkite, E.P. (2002) Evidence for two commonly deleted regions on mouse chromosome 2 in γ ray-induced acute myeloid leukemic cells. Experimental Hematology, 30 ,564-570.

394. Roberts, C., Daw, S.C., Halford, S., and Scambler, P.J. (1997) Cloning and developmental expression analysis of chick Hira (Chira), a candidate gene for DiGeorge syndrome. Hum. Mol. Genet., 6, 237-245.

395. Roberts, C., Sutherland, H.F., Farmer, H., Kimber,W., Halford, S., Carey, A., Brickman, J.M., Wynshaw-Boris, A., and Scambler, P.J. (2002) Targeted mutagenesis of the Hira gene results in gastrulation defects and patterning abnormalities of mesoendodermal derivatives prior to early embryonic lethality. Molecular and Cellular Biol., 22 (7), 2318-2328.

396. Robinson, W.P., Waslynka, J., Bernasconi, F., Wang, M., Clark, D., Kotzot, D., and Schinzel, A. (1996) Delineation of 7q11.2 deletions associated with Williams-Beuren syndrome and mapping of a repetitive sequence within and to either side of the common deletion. Genomics, 34, 17-23.

397. Roe, B.A, Crabtree, J., and Khan, A. (1996) DNA Isolation and Sequencing. John Wiley and Sons Ltd, New York, NY.

398. Roe, B.A. (1997) Protocol for recombinant DNA isolation, cloning and sequencing.

399. Rogers, J and VandeBerg, L, J (2001) Gene Maps of Nonhuman Primates. ILAR Journal, 39, 1-12.

400. Rogers, J., and Hixson, E.J. (1997) Insights from model systems: Baboons as an animal model for genetic studies of common human disease. Am. J. Hum. Genet, 61, 489-493.

401. Rogers, J., Mahaney, C.M., Witte, M.S., Nair, S., Newman, D., Wedel, S., Rodriguez, A.L., Rice, S.K., Slifer, H.S., Perelygin, A., Slifer, M, Palladino-Negro, P., Newman, T., Chambers, K., Joslyn, G., Parry, P., and Morin, A.P.(2000) A genetic linkage map of the baboon (Papio hamadryas) genome based on human microsatellite polymorphisms. Genomics, 67, 237-247.

402. Rowold, J. D., and Herrera, J. R. (2000) Alu elements and the human genome. Genetica, 108, 57-72.

403. Roy, A.M., Carroll, M.L, Nguyen, S.V et al. (2000) Potential Gene conversion and source genes for recently integrated Alu elements. Genome Res. 10, 1485-1495.

404. Roy, A.M., Carroll, M.L., Kass, D.H., Nguyen, S.V., and Salem, A.-H. (1999) Recently integrated human Alu repeats: finding needeles in the haystack. Genetica, 107, 149-161.

405. Roy-Engel, M.A., Carroll, L.M., El-Sway, M., Salem, A-H., Garber, K.R., Nguygen, V.S., Deininger, L.P., and Batzer, A.M. (2002) Non-traditional Alu evolution and primate genomic diversity. J. Mol. Biol. 316, 1033-1040.

406. Rozas, J and Rozas, R. (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics,15,174-175.

407. Ruvolo, M. (1997) Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. Mol. Biol. Evol. 14, 248-265.

408. Saccone, Cecilia., and Pesole, G. (2003) Handbook of comparative genomics: Principles and methodology. John Wiley& Sons Inc, Hoboken, NJ.

409. Saito, T., Stopkova, P., Diaz,L., Papolas, D.F., Boussemart, L., and Lachman,H.M. (2003) Polymorphism screening of PIK4CA: possible candidate gene for chromosome 22q11-linked psychiatric disorders. Am.J.Med.Genet.B. Neuropsychiatr Genet. 16, 77-83.

410. Sakaki, Y., Watanabe, H., Taylor, T., Hattori, M., Fujiyama, A., Toyoda, A., Kuroki, Y., Itoh, T., Saitou, N., Oota, S., Kim, G-.C., Kitano, T., Lehrach, H., Yaspo, L-.M., Sudbrak, R., Kahla, A., Reinhardt, R., Kube, M., Platzer, M., Taenzer, S., Galgoczy, P., Kel, A., Bloecker, H., Scharfe, M., Nordsiek, G., Hellmann, I., Khaitovich, P., Paabo, S., Chen, Z., Wang, Y-.S, Ren , X.S., Zhang, L-. X., Zheng, J-.H., Zhu, F-.G., Wang, F-.B., Zhao, P-.G., Tsai, F.S ., Wu,K., Liu,T-.T., Hsiao, J-.K., Park, S-.H., Lee, S-.Y., Cheong, E-.J., and Choi, H-,S. (2003) Human versus chimpanzee chromosome-wide sequence comparison and its evolutionary implication. Cold Spring Harb Symp Quant Biol, 68, 455-460.

411. Sakate, R., Osada, N., Hida, M., Sugano, S., Hayasaka, I., Shimohira, N., Yanagi, S., Suto, Y., Hashimoto, K., and Hirai, M. (2003). Analysis of 5'-End Sequences of Chimpanzee cDNAs. Genome Res, 13, 1022-1026.

412. Salem, A-H., Kilroy, E.G., Watkins, S.W., Jorde, B.L., and Batzer, A.M. (2003) Recently integrated Alu elements and human genomic diversity. Mol Biol. Evol., 20(8), 1349-1361.

413. Salem, H-A., Ray, A.D., Xing, J., Callinan, A.P., Myers, S.J., Hedges, J.D., Garber, K.R., Witherspoon, J.D., Jorde, B.L., and Batzer, A.M. (2003). Alu elements and hominid phylogenetics. PNAS, 100(22), 12787-12791.

414. Sambrook,J., Fritsch, E. M., and Maniatis, T. (1989) Molecular Cloning:A Laboratory Manual, 2nd Edition., ColdSpring Harbor Laboratory, Cold Spring Harbor, NY. pp.8.11-8.13.

415. Samonte, V.R., and Eichler, E.E. (2002) Segmental duplications and the evolution of the primate genome. Nat Rev Genet, 3, 65-72.

416. Sampath, K et al. (1998) Induction of the zebrafish ventral brain and floorplate requires Cyclops/nodal signals. Nature. 395, 185-189.

417. Sanders, A.R., Rusu,I., Duan, J., Vander Molen, J.E., Hou, C., Schwab, S.G., Wildenauer, D.B., Martinez, M., and Gejman, P.V. (2005) Haplotype association spanning the 22q11.21 genes COMT and ARVCF with schizophrenia. Mol. Psychiatry, 10(4), 353-365.

418. Sandhu,H.K., Hollenbeck,N., Wassink, T.H., Philibert, R.A. (2004) An association study of PCQAP polymorphisms and schizophrenia. Psychiatr Genet., 14(3), 169-172.

419. Sanger, F and Coulson, A.R. (1978). The use of thin acrylamide gel for DNA sequencing. FEBS letters. 87, 107-110.

420. Sanger, F. (1959) Chemistry of Insulin. Science. 129, 1340-1344.

421. Sanger, F., Nicklen, S., A and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74(12), 5463-5467.

422. Sau, K., Sau,S., Mandal, S.C., and Ghosh, T.C. (2005) Factors influencing the synonymous codon and amino acid usage bias in AT-rich Pseudomonas aeruginosa Phage PhiKZ. Acta Biochimica et Biophysica Sinica, 37(9), 625-633.

423. Schibler, L, Vaiman, D., Oustry, A., Giraud-Delville, C., and Cribiu, P.E. (1998) Comparative gene mapping: a fine-scale survey of chromosome rearrangements between ruminants and humans. Genome Res, 8, 901-915.

424. Schier, A.F., Joyner, A.L., Lehmann, R., and Talbot, W.S. (1996) From screens to genes: prospects for insertional mutagenesis in zebrafish. Genes Dev. 10(24),3077-3080.

425. Schinke, M., and Izumo, S. (2001) Deconstructing DiGeorge syndrome. Nature Genet., 27, 238-240.

426. Schinzel, A., Schmid, W., Fraccaro, M., Tiepolo, L., Zuffardi, O., Opitz, J.M., Lindsten, J., Zetterqvist, P., Enell, H., Baccichetti et al.(1981) The 'cat eye syndrome': Dicentric small marker chromosome probably derived from a no.22 (tetrasomy 22pter to q11) associated with a characteristic phenotype. Report of 11 patients and delineation of the clinical picture. Hum. Genet. 57, 148-158.

427. Schuler, G.D.(1997) Pieces of the puzzle: expressed sequence tags and the catalog of human genes. J.Mol. Med, 75(10), 694-698.

428. Schwartz, S., Elnitski, L., Li, M., Weirauch, M., Reimer, C., Smit, A., NISC Comparative Sequencing Program, Green, E., Hardison RC., Miller, W. (2003)

MultiPipmaker and Supporting Tools: Alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Research, 31, 3518-3524.

429. Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Reimer, C., Bouck, J, Gibbs, R., Hardison, R., and Miller, W. (2000). PipMaker-a web server aligning tow genomic DNA sequences. Genome Res, 10, 577-586.

430. Serluca, F.C., Sidow, A., Mably, J.D., and Fishman, M.C. (2001) Partitioning of tissue expression accompanies multiple duplications of the Na+/K+ ATPase α subunit gene. Genome Res. 11, 1625-1631.

431. Seydoux, G and Fire, A. (1995) Whole-mount in situ hybridization for the detection of RNA in Caenorhabditis elegans embryos. Methods In Cell Biol, 48,323-337.

432. Shaikh, T.H., Kurahashi, H., and Emanuel, B.S. (2001) Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: An update and literature review. Genetics In Medicine, 3, 6-13.

433. Shaikh, T.H., Kurahashi, H., Saitta, S.C., Saitta, A., Mizrahy O'Hare, A., Hu, P., Roe, B., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budraf, M.L., and Emanuel, B.S. (2000) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: Genomic organization and deletion endpoint analysis. Hum. Mol. Genet, 9,489-501.

434. Sharp, P.M., Li, W-H.(1986) Codon usage in regulatory genes in Escherichia coli does not reflect selection for "rare" codons. Nucleic Acids Res. 14, 7734-7749.

435. Sharp, P.M., Tuohy, T.M.F., Mosurski, K.R. (1986) Codon usage in yeast: Cluster analysis clearly differentiates between highly and lowly expressed genes. Nucleic Acids Res. 14, 5125-5143.

436. Shaw, C.J and Lupski, J.R. (2004) Implications of human genome architecture rearrangement-based disorders: the genomic basis of disease. Hum. Mol. Genet. 13, R57-R64.

437. Shen, M. R., Batzer, A.M., and Deininger, L. P.,(1991) Evolution of the master Alu gene(s). J. Mol. Evol. 33, 311-320.

438. Shifman, S., Bronstein, M., Sternfeld, M., Pisante, A., Weizman, A., Reznik, I.,Spivak, B., Grisaru, N., Karp, L., Schiffer, R., Kotler, M., Strous, R.D., Swartz-Vanetik, M., Knobler, H.Y., Shinar, E., Yakir, B., Zak,N.B., Darvasi,A.(2004) COMT: a common susceptibility gene in bipolar disorder and schizophrenia. Am J Med Genet B Neuropsychiatr Genet. 128(1), 61-64.

439. Shiohama, A., Sasaki,T., Noda,S., Minoshima, S., and Shimizu, N. (2003) Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. Biochem Biophys Res Commun. 304(1),184-190.

440. Shumeli, O., Horn-Saban, S., Chalifa-Capsi, Vered, Shmoish, M., Ophir, R., Benjamin-Rodrig, H., Safran, M., Domany,Eytan., and Lancet, D. (2003) GeneNote: whole genome expression profiles in normal human tissues. C.R. Biologies, 326, 1067-1072.

441. Sirotkin,H., O'Donnell,H., DasGupta, R., Halford, S., St.Jore, B., Puech,A., Parimoo, S., Morrow, B., Skoultchi, A., Weissman, S.M., Scambler, P., and

Kucherlapati, R. (1997) Identification of a new human catenin gene family member (ARVCF) from the region deleted in velo-cardio-facial syndrome. Genomics 41,75-83.

442. Smit, A.F. (1993) Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res. 21(8), 1863-1872.

443. Smit, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. 9, 657-663.

444. Smit, A.F., Toth, G., Riggs, A.D., and Jurka, J. (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J.Mol.Biol. 246(3), 401-417.

445. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B and Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. Nature, 321, 674-679.

446. Smith, L.P.T., Grosse, M.W., Freking, A.B., Roberts, J.A., Stone, T.R., Casa, E., Wray, E.J., White, J., Cho, J., Fahrenkrug, C.S., Bennet, L.G., Heaton, P.M, Laegreid, W.W., Rohrer, A.G., Chitko-Mckown, G.C., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, W.J. (2001) Sequence evaluation of four pooled tissue normalized bovine cDNA libraries and construction of a gene index for cattle. Genome Res, 11, 626-630.

447. Smith, T.F and Waterman, S.M (1981) Identification of common molecular subsequences. J.Mol. Biol. 147, 195-197.

448. Solnica-Krezel, L., Schier, A.F., and Driever, W.(1994) Efficient recovery of ENU-induced mutations from the zebrafish germline. Genetics. 136, 1401-1420.

449. Solomon, K.S., Logsdon, J.M. Jr., and Fritz, A. (2003) Expression and Phylogenetic analyses of three zebrafish Fox1 class genes. Developmental Dynamics. 228, 301-307.

450. Sorek, R., Ast, G., and Gaur, D. (2002) Alu-containing exons are alternatively spliced. Genome Res. 12, 1060-1067.

451. Spiteri, E., Babcock, M., Kashork, C. D., Wakui, K., Gogineni, S., Lewis, D.A., Williams, K. M., Minoshima, S., Sasaki, T., Shimizu, N., Potocki, L., Pulijaal, V., Shanske, A., Shaffer, L.G., and Morrow, B.E. (2003) Frequent translocations occur between low copy repeats on chromosome 22q11.2 (LCR22s) and telomeric bands of partner chromosomes. Hum. Mol. Genet., 12, 1823-1837.

452. Stahmann, M.A., Huebner, C.F., and Link, K.P. (1941) Studies of the hemorrhagic sweet clover disease. V. Identification and synthesis of the hemorrhagic agent. J.Biol. Chem. 138, 513-527.

453. Stainier D.Y.R, Fouquet, B., Chen, J, Warren, K.S., Weinstein, B.M, et al. (1996) Mutaions affecting the formation and function of the cardiovascular system in the zebrafish embryo. Development. 123, 285-292.

454. Stainier, D.Y.R and Fishman, M.C. (1994) The zebrafish as a model system to study cardiovascular development. Trends Cardiovascular Med. Trend Cardiovasc. Med. 4, 207-212.

455. Stankiewicz, P., Cheung, S.W., Shaw, C.J., Saleki, R., Szigeti, R., and Lupski, J.R. (2003) The donor chromosome breakpoint for a jumping translocation is associated with large low-copy repeats in 21q21.3. Cytogenet Genome Res., 101, 118-123.

456. Stankiewicz, P., Park, S.S., Inoue, Kand Lupski, J.R. (2001) The evolutionary chromosome translocation 4;19 in Gorilla gorilla is associated with microduplication of the chromosome fragment syntenic to sequences surrounding the human proximal CMT1A-REP. Genome Res., 11, 1205-1210.

457. Stenesh, J and Roe, B.A. (1972) DNA polymerase from mesophilic and thermophilic bacteria. I. Purification and properties of DNA polymerase from Bacillus licheniformis and Bacillus stearothermophilus. Biochem. Biophys. Acta. 272, 156-166.

458. Stewart, C.-B and Disotell (1998) "Primate evolution- in and out of Africa". Current Biology, 8, R582-R588.

459. Stickney, H.L., Barresi,M.J.F., and Devoto, S.H. (2000) Somite development in zebrafish. Dev. Dynamics, 219, 287-303.

460. Stojanovic, N., Florea, L., Riemer, C., Gumucio, D., Slightom, J., Goodman, M., Miller, W., and Hardison R. (1999). Comparison of five methods for finding conserved sequences in multiple alignment s of gene regulatory regions. Nucleic Acids Res. 27, 3899-3910.

461. Stoll, M., Kwitek-Black, A.E., Cowley Jr, A.W., Harris, E.L., Harrap, S.B., Krieger, J.E., Printz, M.P., Provoost, A.P., Sassard, J and Jacob, H.J. (2000) New target region for human hypertension via comparative genomics. Genome Res, 10, 473-482.

462. Strausberg RL, Buetow KH, Emmert-Buck M, Klausner R. (2000) The Cancer Genome Anatomy Project: building an annotated gene index. Trends in Genetics, 16,103-106.

463. Strausberg, R.L., Feingold, E.A., Mammalian Gene Collection Program Team.(2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci USA, 99(26), 16899-16903.

464. Streisinger, G., Singer, F., Walker, D., Knauber, D., and Dower, N. (1986) Segregation analyses and gene-centromere distances in zebrafish. Genetics. 112, 311-319.

465. Streisinger, G., Walker, N., Dower, D., Knauber, D., and Singer, F. (1981) Production of clones of homozygous diploid zebrafish (Brahydanio rerio). Nature (Lond). 291, 293-296.

466. Subramanian, S., and Kumar, S. (2003). Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. Genome Res, 13, 838-844.

467. Sultmann, H., Sato, A., Murray, B.W., Takezaki, N., Geisler, R, Rauch, G-J and Klein, J. (2000) Conservation of MHC Class III region synteny between zebrafish and human as determined by radiation hybrid mapping. The Journal of Immunology. 165, 6984-6993.

468. Sutton, W.S. (1903) The chromosomes in heredity. Biol, Bulletin, 4, 231-251.

469. Sverdlov, D.E. (2000) Retroviruses and primate evolution. Bioessays, 22, 161-171.

470. Tabor, S and Richardson, C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Nat. Acad. Set. USA. 84, 4767-4771.

471. Taylor, J.S., Van de Peer, Y., Braach, I., Meyer, A. (2001) Comparative genomics provide evidence for an ancient genome duplication event in fish. Phl. Trans. Roy. Soc. Lond. B 356, 1661-1679.

472. The C.elegans Genome Sequencing Consortium. (1998) Genome sequence of the nematode C.elegans: a platform for investigating biology. Science, 282, 2012-2018.

473. The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group), "The Transcriptional Landscape of the Mammalian Genome". (2005) Science, 309(5740),1559-1563.

474. The Rat Genome Sequencing Project Consortium. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature, 428, 493-521.

475. Thomas, J.W and Touchman, J.W. (2002) Vertebrate genome sequencing: Building a backbone for comparative genomics. Trends. Genet. 18,104-108.

476. Thomas, J.W., Touchman, J.W., Blakesley, R.W., Bauffard, G.G., Beckstrom-Sternberg, S.M., Margulies, E.H., Blanchette, M et al. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature, 424, 788-793.

477. Thomas, W.J and Touchman, J.W. (2002). Vertebrate genome sequencing :building a backbone for comparative genomics. 18, 104-108.

478. Toder, R., Xia, Y., and Bausch, E. (1998) Interspecies comparative genome hybridization and interspecies representational difference analysis reveal gross DNA differences between humans and great apes. Chromosome Res, 6, 487-494.

479. Touchman, J.W., Dehejia, A., Chiba-Falek, O., Cabin, D.E., Schwartz, J.R., Orrison, B. N., Polymeropoulos, M.H., and Nussbaum, R.L. (2001) Human and mouse α-synuclein genes:Comparative genomic sequence analysis and identification of a novel gene regulatory element. Genome Res, 11, 78-86.

480. Trappe, R., Ahmed, M., Glaser,B., Vogel,C., Tascou, S., Burfeind, P., and Engel, W. (2002) Identification and characterization of a novel murine multigene family containing a PHD-finger-like motif. Biochem Biophys Res Commun. 293(2), 816-826.

481. Trappe, R., Schulze, E., Rzymski, T., Frode, S and Engel, W. (2002) The Caenorhabditis elegans ortholog of human PHF5a shows a muscle-specific expression domain and is essential for C.elegans morphogenetic development. Biochem Biophys Res Commun. 297(4), 1049-1057.

482. Uddin, M., Wildman, E.D., Liu, G., Xu, W., Johnson, M.R., Hof, R. Patrick., Kapatos, G., Grossman, I. L., and Goodman, M. (2003) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. PNAS, 101(9), 2957-2962.

483. Ullu E, Tschudi C. (1984) Alu sequences are processed 7SL RNA genes. Nature. 312(5990):171–172.

484. Ureta-Vidal, A., Ettwiller, L., and Birney, E. (2003) Comparative Genomics: Genome-wide analysis in metazoan eukaryotes. Nat. Rev. Genet., 4, 251-262.

485. Van de Peer, Y., Taylor, J.S., Braasch, I., and Meyer, A. (2001) The ghost of selection past: rates of evolution and functional divergence in anciently duplicated genes. J. Mol. Evol. 53, 434-444.

486. van Eeden, F.J., Granato, M., Schach,U., Brand, M., Furutani-Seiki, M., Haffter,P., Hammerschmidt,M., Heisenberg,C.P., Jiang,Y.J., Kane,D.A., Kelsh,R.N., Mullins,M.C., Odenthal,J., Warga,R.M., Allende,M.L., Weinberg, E.S., Nusslein-Volhard,C. (1996) Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. Development, 123,152-164.

487. VandeBerg JL., Williams-Blangero, S. (1996) Strategies for using nonhuman primates in genetic research on multifactorial diseases. Lab Anim Sci 46, 146-151.

488. Vanholme, B., Meutter, D.J., Tytgat, T., Gheysen, D. C.G., Vanhoutte, I and Gheysen, D.R.G. (2002) An improved method for whole mount in situ hybridization of Heterodera schachtii juveniles. Parasitology Research, 88,nr.9,731-733.

489. VanOorschot, R.A.H., and VandeBerg, J.L. (1991) Tight linkage between MPI and NP in baboons. Genomics. 9, 783-785.

490. Varki, A. (2000). A chimpanzee genome project is a biomedical imperative. Genome Res, 10, 1065-1070.

491. Venter, J.C et al. (2001). The sequence of human genome. Science, 291, 1304-1351.

492. Vitelli, F., Viola, A., Morishima, M., Pramparo,T., Baldini, A., Lindsay,E. (2003) TBX1 is required for inner ear morphogenesis. Hum. Molec. Genet. 12, 2041-2048.

493. Vogel, G. (2000) Genomics. Sanger will sequence zebrafish genome. Science. 290, 1671.

494. Wagner, A. (2000) Inferring lifestyle from gene expression patterns. Mol.Biol. Evol., 17(12),1985-1987.

495. Wallace, M.R., Andersen, L.B., Saulino, A.M., Gregory, P.E., Glover, T.W., Collins, F.S. (1991) A de novo Alu insertion results in neurofibromatosis type 1. Nature, 353, 864-866.

496. Wasserman, W.W., Palumbo, M., Thompson, W., Fickett, J.W., and Lawrence, C.E. (2000). Human-mouse genome comparisons to locate regulatory sites. Nat. Genet. 26, 225-228.

497. Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., Kube, M., Taenzer, S., Galgoczy, P., Platzer, M., Scharfe, M., Nordsiek, G., Bloecker, H., Hellmann, I., Khaitovich, P., Paabo, S., Reinhardt, R., Zheng, J-.H., Zhang, L-.X., Zhu, F-.G., Wang, F-.B., Fu, G., , Ren , X-.S., Zhao, P-.G., Chen, Lee, S-.Y., Cheong, E-.J., Choi, H-,S., Wu, M-.K., Liu,T-.T., Hsiao, J-.K., Tsai, F.S ., Kim, G-.C., Oota, S., Kitano, T., Kohara, Y., Saitou, N., Park, S-, H., Wang, Y-.S., Yaspo, L-.M., and Sakaki, Y. (2004) DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature, 429, 382-388.

498. Waterston, R.H et al. (2002) Initial sequencing and comparative analysis of mouse genome. Nature, 420, 520-562.

499. Watson, J.D., and Crick, F. H. C. (1953) Genetical implications of the structure of deoxyribonucleic acid. Nature, 171, 964-967.

500. Watson, J.D., and Crick, F. H. C. (1953) Molecular structure of nucleic acids. Nature, 171, 737-738.

501. Webster, T.M., Smith, G.C.N and Ellegren, H. (2003) Compositional evolution of noncoding DNA in the human and chimpanzee genomes. Mol. Biol. Evol, 20, 278-286.

502. Webster, T.M., Smith, G.C.N., and Ellegren, H. (2002) Microsatellite evolution inferred from human chimpanzee genomic sequence alignments. PNAS, 99, 8748-8753.

503. Weinholds, E, van Eeden, F., Kosters, M., Mudde, J., Plasterk, R.H.A and Cuppen, E. (2003). Efficient target-selected mutagenesis in zebrafish. Genome Res, 13:2700-2707.

504. Weinstein, B.M., Schier, A.F., Abdelilah, S., Malicki, J., Solnica-Krezel, L., et al. (1996) Hematopoietic mutations in the zebrafish. Development. 123, 303-309.

505. Werner, T. (1999) Models of prediction and recognition of eukaryotic promoters. Mammalian Genome, 10, 168-177.

506. Westerfield, M. (1994) The Zebrafish Book. A Guide for the laboratory use of zebrafish (Brachydanio rerio).Edition 2.1,Oregon, University of Oregon Press.

507. Westerfield, M., Liu, D.W., Kimmel, C.B., and Walker, C. (1990). Pathfinding and synapse formation in a zebrafish mutant lacking functional acetylcholine receptors. Neuron. 4, 867-874.

508. Westerfield, M., Wegner, F., Jegalian, B.G., DeRobertis, E.M., and Puschel, A.W. (1992) Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. Genes & Dev. 6, 591-598.

509. Whitfield, T.T., Mullins, M.C., Nusslein-Volhard, C., et al. (1996) Mutations affecting development of the zebrafish inner ear and the lateral line. Development. 123,241-254.

510. Whitfield,T,T. (2002) Zebrafish as a model for hearing and deafness. Journal of Neurobiology, 53(2),157-171.

511. Whitfield,T.T., Riley, B.B., Chiang, M.Y., Phillips,B. (2002) Development of the zebrafish inner ear. Dev.Dyn. 223,1670-1673.

512. Wildman, E, D. (2002) A map of the common chimpanzee genome. BioEssays, 24, 490-493.

513. Wildman, E, D., Uddin, M., Liu, G., Grossman, I, L., and Goodman, M. (2003) Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. PNAS, 100, 7181-7188.

514. Wilkins, M. H. F., Gosling, R.G., and Seeds, W.E. (1951) Physical studies of nucleic acids. Nature, 167, 759-760.

515. Williams, S.H., Mouchel, N., and Harris, A. (2003) A comparative genomic analysis of the cow, pig and human CFTR genes identifies potential intronic regulatory elements. Genomics, 81 628-639.

516. Williams-Blangero, S., Blangero, J. (1995) Heritability of age at first birth in captive olive baboons. Am J Primatol, 37, 233-239.

517. Wilson, D.I., Burn, J., Scamber, P., and Goodship, J. (1993) DiGeorge syndrome: Part of CATCH 22. J. Med. Genet, 30, 852-856.

518. Wilson, M.D., Reimer, C., Martindale, D.W., Schnupf, P., Boright, A.P., Cheung, T.L., Hardy, D.M., Schwartz, S., Scherer, S.W., Tsui, L-C., Miller, W., and Koop, B.F. (2001) Comparative analysis of the gene-dense ACHE/TFR2 region on human chromosome 7q22 with the orthologous region on mouse chromosome 5. Nucleic Acids Research. 29, 1352-1365.

519. Wiltbank, J.N., Rothisberger, J.A., and Zimmerman, D.R. Effect of human chorionic gonadotropin on maintenance of the corpus luteum and embryo survival in the cow. J. Anim. Sci. 64, 642-647.

520. Winship, P.R. (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucleic Acids Res. 17, 1266.

521. Wise, C.A., Garcia, C.A., Davis, S.N., Heju, Z., Pentao, L., Patel, P.I and Lupski, J.R. (1993) Molecular analyses of unrelated Charcot-Marie-Tooth (CMT) disease patients suggest a high frequency of the CMTIA duplication. Am. J. Hum. Genet. 53, 853-863.

522. Wong, K and Cantley, L.C. (1994) Cloning and Characterization of a human phosphatidylinositol 4-kinase. J. Biol. Chem. 269, 28878-28884.

523. Woods, I.G., Kelly, P.D., Chu, F., Ngo-Hazelett, P., Yan, Y.L., et al. (2000) A comparative map of zebrafish genome. Genome Res. 10, 1903-1914.

524. Wray, A. G., Hahn, W. M., Abouheif, E., Balhoff, P.J., Pizer, M., Rockman, V.M and Romano, A.L. (2003) The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. 20(9), 1377-1419.

525. Wu, Y.L., and Gong,Z. (2004) A novel zebrafish kelchlike gene klhl and its human ortholog KLHL display conserved expression patterns in skeletal and cardiac muscles. Gene, 338(1),75-83.

526. Yamagishi, C., Hierck, B.P., Gittenberger-De Groot, A.C., Yamagishi, H., Srivastava, D. (2003) Functional attenuation of UFD1l, a 22q11.2 deletion syndrome candidate gene, leads to cardiac outflow spetation defects in chicken embryos. (2003) Pediatr Res. 53(4), 546-553.

527. Yan, Y.-L., Miller, C.T., Nissen, R., Singer, A., Liu, D., Kirn, A., Draper, B., Willoughby, J., Morcos, P.A., Amsterdam, A., Chung, B.-c., Westerfield, M., Haffter, P., Hopkins, N., Kimmel, C., and Postlethwait, J.H. (2002) A zebrafish sox9 gene required for cartilage morphogenesis. Development. 129, 5065-5079.

528. Yao,Z., Spriggs, M. K.; Derry, J. M. J.; Strockbine, L.; Park, L. S.; VandenBos, T.; Zappone, J.; Painter, S. L.; Armitage, R. J. (1997) Molecular characterization of the human interleukin (IL)-17 receptor. Cytokine 9,794-800.

529. Yap, V.B., and Speed, T. (2005) Rooting a phylogenetic tree with nonreversible substitution models. BMC Evolutionary Biology.,5:2.

530. Zackai, E.H., and Emanuel, B.S. (1980) Site-specific reciprocal translocation, t(11;22)(q23;q11) in several unrelated families with 3:1 meiotic disjunction. Am J Med Genet, 7, 507-521.

531. Zhang, J., Wang, X., and Podlaha, O. (2004) Testing the chromosomal Speciation hypothesis for humans and chimpanzees. Genome Res, 14, 845-851.

532. Zhang, L., Gaut, B.S., and Vision, T.J. (2001) Gene duplication and evolution. Science. 293, 1551.

533. Zhang, W., Hu, G.Y., Deisseroth, A. (1991) Improvement of PCR sequencing by formamide. Nucleic Acids Res. 19,6649.

534. Zhong, T., Kaphingst, K., Akella, U., Haldi, M., Lander, E and Fishman, M. (1997). Zebrafish genomic library in yeast artificial chromosome. Genomics. 48, 136-138.