1-1-2004

Molecular genetic investigation of autosomal dominant hemifacial microsomia

Rebecca Watt
Edith Cowan University

Edith Cowan University

Copyright Warning

MOLECULAR GENETIC INVESTIGATION OF AUTOSOMAL DOMINANT HEMIFACIAL MICROSOMIA

Rebecca Watt BSc

[redacted]

Principal Supervisor - Dr Mel Ziman

Associate Supervisor - Dr Richard Brightwell

This thesis is presented for the degree of Masters of Science by Research

at Edith Cowan University,

2004

Copyright Declaration

Abstract

The overall purpose of this project was to explore the underlying pathogenesis of Hemifacial microsomia (HFM), investigate treatment and management options, identify likely candidate genes and screen candidate genes for mutation(s) causing the form of HFM segregating on chromosome 11 in a West Australian family.

Hemifacial microsomia is a congenital malformation arising from the derivatives of the first and second branchial arches. It is both clinically and genetically heterogeneous, and can occur sporadically or segregate within families in an autosomal fashion. HFM is characterised by significant undergrowth to one side of the face and is a common birth defect with an estimated incidence of 1 in 1,000 to 1 in 5,600 births

Most HFM cases are sporadic, but there are rare familial cases that exhibit autosomal dominant inheritance. These families present the best opportunity for locating and identifying HFM mutations. Two autosomal dominant forms of this disease have been located; one to a 10.7 million DNA base region on chromosome 14 and the other to an 18.8 million base region on chromosome 11. The identification and cooperation of a large West Australian family with HFM has allowed the phenotypic and genotypic study of the disorder presented in this thesis.

Candidate genes were identified from the Draft Human Genome through genome mining and bioinformatics tools. This essentially involved cross-tabulating the genes and predicted genes in the two known hemifacial

microsomia candidate regions and by comparison of gene: names, aliases, acronyms, functions, disease associations (particularly with axial asymmetry), timing of expression and trinucleotide repeat expansions (chromosome 11 hemifacial microsomia appears to show anticipation). The chromosome 11 linkage region was also investigated for homologies to the Completed Mouse Genome.

This project, based on the anticipation described in the family under investigation, and the autosomal dominant nature of this disorder identified four candidate genes at the chromosome 11 linkage locus, three of which were successfully sequenced and eliminated as candidates in this linkage region. Furthermore the sequencing results obtained from this project, together with marker analysis reduced the overall region to be analysed by 355,995 bases.

This project has greatly contributed to the fundamental understanding of this disease and its phenotype and provides a basis for further studies aimed at better diagnosis, treatment and identification of causative genes and other factors.

Thesis Declaration

I certify that this thesis does not, to the best of my knowledge and belief:

(i) *incorporate without acknowledgement any material previously submitted for a degree or diploma in any institution of higher education;*

(ii) *contain any material previously published or written by another person except where due reference is made in the text; or*

(iii) *contain any defamatory material.*

I also grant permission for the Library at Edith Cowan University to make duplicate copies of my thesis as required.

Rebecca Watt *9th February 2005*

TABLE OF CONTENTS

Units

bp	base pair
cM	centi-Morgan
°C	degrees centigrade
g	gram
L	litre
m	metre
M	Molar (moles/litre)
Mb	mega-base
µg	microgram
min	minute
µL	microliter
mM	milliMolar
ng	nanogram
pg	picograms
rpm	revolutions per minute
s	second
V	volts
W	watts

Abbreviations

A/P	anterior/posterior
AD	autosomal dominant
ANRI	Australian Neuromuscular Research Institute
ARIX	Aristaless homeo-box
BLAST	Basic Local Alignment Search Tool
cDNA	complementary DNA
CLNS1A	chloride channel, nucleotide-sensitive, 1A
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddH_2O	double distilled water
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP(s)	deoxynucleotide triphosphate(s)
dsDNA	double stranded DNA
dTTP	deoxythymidine triphosphate
ECU	Edith Cowan University
EDTA	ethylenediaminetetraacetic acid
EtBr	ethidium bromide

GARP	Garpin
GARP	golgi-associated retrograde protein
GS	Goldenhar syndrome
HFM	Hemifacial Microsomia
HOM-C	homeotic gene complex
HOX	homeobox
KEMH	King Edward Memorial Hospital
L/R	left/right
Mb	mega-base (1 million bases)
$MgCl_2$	magnesium chloride
MIM	Mendelian Inheritane in Man
NaAc	sodium acetate
NCBI	National Centre for Biotechnology Information (USA)
OAV	oculoauriculovertebral
Oligo	oligonucleotide
OMENS	Orbital distortion, Mandibular hypoplasia, Ear anomaly, Nerve involvement & Soft tissue deficiency classification system
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction

RED	repeat expansion detection
REF	restriction endonuclease fingerprinting
RNA	ribonucleic acid
RPH	Royal Perth Hospital
RSSCP	RNA SSCP
SAT	Skeletal, Auricule, soft Tissue classification system
SeqEd	Sequence Editor version 1.0.3
Shh	*Sonic Hedgehog*
SOBBS	School of Biomedical and Sports Science, ECU
SSCP	single-stranded conformation polymorphism
T_A	annealing temperature
TAE	Tris/acetate/EDTA buffer
TGF-β	transforming growth factor-β
T_M	melting temperature
UV	ultraviolet
UVRAG	Ultraviolet radiation resistance associated gene
WNT11	Wingless-related MMTV integration site
ZPA	zone of polarizing activity

TABLES and FIGURES

Tables

Figures

Acknowledgements

Thank you firstly to God who gave me the strength to undertake this project in the first place.

Thank you to my late gran (Doreen Peate) for whom I did this project. You're the reason I kept going and I will miss you forever.

Thank you to my family for putting up with me, my paper all over the floor, for standing beside me, staying up late, running back and forth to the printers, your company in the lab's at night and your never-ending faith that I could do it. I did!

Thank you to my friends, Catherine, Kristin, Nat, Paul, Zoe and Sam, Bentley, Bert, Gavin, Greg, Katie, Barry, Mike and Katie, Suella, Treasure, Ashley and Adam. Your friendship has kept my spirits high and I will forever be grateful for your company and distraction techniques that stopped me from going mad!

Thank you to Lori, Dr Laing and everyone at the Australian Neuromuscular Research Institute for their assistance.

Thank you to Dr David Chandler, who's PhD preceded this project.

Thank you to the family who was researched in this study without whom this project would not have existed.

Thank you to Mark Davis and Rebecca at Neuropathology, Royal Perth Hospital for their time, effort and kind support.

Thank you to Associate Professor Jack Goldblatt and his staff at Genetic Services, Western Australia.

Thank you to everyone at the Centre for Human Genetics. In particularly, Professor Alan Bittles for his advice and encouragement, Doreen Mackie (nothing was ever too much to ask), Andrea (you're so focused – its inspiring), Tina (I really appreciated your support and chats on the overpass when things got tough), Judith (thank you for all your advice and encouragement – you're a true friend), Michael (thank you so much for your help in the lab), Hugh (for always having time to help me find things around the lab), Sheena, Marie, Jenny and Cheryl (you are so brilliant,

thank you for your advice and assistance). Thank you all for your patience in allowing me to fumble around before finding my feet.

Thank you to Dr Angus Stewart and Dr Peter Roberts for their advice and support. I am so grateful for everything I learnt from you both.

Thank you especially to Dr Richard Brightwell for your dedication and belief in my ability to complete this undertaking (even when things were really rough).

A huge thank you to Dr Mel Ziman for your continual support, encouragement and quiet strength. Your patience and kind advice has been the driving force behind my completion of this degree. I am so grateful to you.

Thanks go to everyone who has co-operated and collaborated with me throughout the duration of this project. Finally, its over!

Chapter 1: Introduction

1.0 Introduction

1.1 Thesis Overview

The subject of this thesis is the review of current literature surrounding hemifacial microsomia (HFM) and determination of the genetic cause of one form of autosomal dominant hemifacial microsomia, segregating on chromosome 11 in a West Australian family. Fundamental aspects of this study include that:

a) HFM is a congenital defect;
b) at the commencement of this thesis a large West Australian family was known to segregate this disease on chromosome 11; and
c) at the commencement of this thesis, a gene had not been identified for the form of HFM linked to chromosome 11.

The research in this thesis involved searching for the gene causing HFM segregating in this West Australian family and, in doing so develop opportunities for further research into the developmental pathways involved in its manifestation.

This thesis will:

1) Introduce HFM and its correlation with the Human Genome Project (HGP);
2) Review what was known about HFM preceding the commencement of this thesis;
3) Describe the molecular techniques and bioinformatics used throughout this project;
4) Present results of molecular investigations obtained during this project;
5) Describe what is now known about HFM;

6) Discuss avenues for further study; and
7) Discuss treatment avenues based on an increased understanding of the disease.

1.2 Hemifacial microsomia (HFM)

HFM involves malformations of first and second branchial arch derivatives (Cousley, Naora, Yokoyama, Kimura & Otani, 2002; Sze, Paladin, Lee & Cunningham, 2002; Kelberman *et al.*, 2001; Silvestri, Natali & Iannetti, 1996; Rodgers, Eppley, Nelson & Sadove, 1991; Converse, Coccaro, Becker & Wood-Smith, 1973). During embryonic development of higher vertebrates, six transient (branchial) arches are formed, all of which gives rise to specialized structures in the head and neck. In HFM, any derivative of the first and second branchial arches may be affected. Consequently the clinical phenotype of those affected by HFM is highly variable (Appendix 1).

The incidence of HFM is uncertain; it is estimated to affect 1 in 5600 births (Cousley & Calvert, 1997; Cohen, 1991; Gorlin, Cohin & Levin, 1990; Grabb, 1965). Most cases are sporadic (Kearns, Padwa, Mulliken & Kaban, 2000), however the recurrence rate in first-degree relatives is estimated to be between 2-3% (Rollnick & Kaye, 1983). HFM is considered the second most common birth defect worldwide (Monahan *et al.*, 2001; Gorlin *et al.*, 1963).

HFM is a complex deformation displaying a spectrum of anomalies that revolve around its universal feature of axial asymmetry. Despite often being unapparent in children due to excess tissue, this asymmetry is present throughout embryonic development. In addition to craniofacial

anomalies, there may be cardiac, vertebral (Gorlin *et al.*, 1963) and central nervous system defects (On-line Mendelian Inheritance in Man-164210 (OMIM-164210), 1998). HFM deformity affecting only one side of the body (unilateral deformity) includes: abnormality in external ear size and shape; rudimentary nodules and tags of ear tissue located anterior to the ear often containing a core cartilage (preauricular nodule or tag); and the closure (atresia) or absence (anotia) of the auditory canal. Nodule and tag differentiation is merely one of size, nodules being smaller than tags. Eye defects associated with HFM may involve small cysts containing skin or structures connected with skin, upon the eyeball (epibulbar dermoids). Furthermore, absence of the upper eyelid (coloboma) is frequent. Preauricular tags are also displayed in Goldenhar syndrome (OMIM-164210, 1998). Goldenhar syndrome is classified as the same disease as chromosome 14 HFM, hence both have the same Mendelian Inheritance in Man (MIM) number of 164210.

Accurate assessment of (facial) anomalies for effective diagnosis and treatment of HFM is crucial (Cousley & Calvert, 1997). Numerous classification systems have been devised to facilitate the classification of HFM, none of which have been universally adopted. The two most predominant classification systems include Orbital distortion, Mandibular hypoplasia, Ear anomaly, Nerve Involvement and Soft tissue deficiency (OMENS) and SAT (S = skeletal, A = auricle, T = soft tissue) (cited in Rodgers *et al.*, 1991).

The pathogenesis and genetic predisposition to HFM is currently the subject of extensive investigation. Little success has been achieved in identifying genetic, clinical or teratological information (Cousley &

Wilson, 1992). HFM appears to be caused by a range of factors, including defective genes, teratogenic substances and/or vascular anomalies. Pathogenetic theories are numerous and range from neural crest cell involvement to a deficiency in the development of the branchial arch and head mesoderm. Due to the complexity of its phenotype and inheritance pattern, a variety of genetic models have been described. The reduced penetrance of an autosomal dominant gene may explain some reported pedigrees (Gibson, Sillence & Taylor, 1996; Cousley & Wilson, 1992), although alternatively polygenetic inheritance has also been proposed to explain the genesis of HFM (Cousley & Wilson, 1992). More so, HFM has been associated with trisomy 18, trisomy 7 and 9 mosaicism and *cri du chat* syndrome (Horgan *et al.*, 1995). To date, linkage regions have been identified on chromosomes 11 and 14, by Chandler (2001) and Kelberman *et al.*, (2001) respectively.

Currently, those affected with significant facial asymmetry are confronted with multiple surgical interventions (Kelberman *et al.*, 2001). Effective management of HFM must take into account the variability in phenotype specific to the individual (Cousley & Calvert, 1997). Much debate surrounds the appropriate time at which to initiate surgical intervention. Effective treatment appears more successful through an interdisciplinary approach to affect a long-term coordinated treatment plan.

The overall purpose of this project was to explore the underlying pathogenesis of HFM, investigate treatment and management options, identify likely candidate genes and the mutation(s) causing the form of HFM segregating on chromosome 11 in a West Australian family.

Outcomes of this project may significantly advance investigations into the molecular pathology of hemifacial microsomia providing those affected with chromosome 11 HFM some hope of better diagnosis, prognosis, prevention and treatment in the future.

1.3 The West Australian Chromosome 11 HFM Family

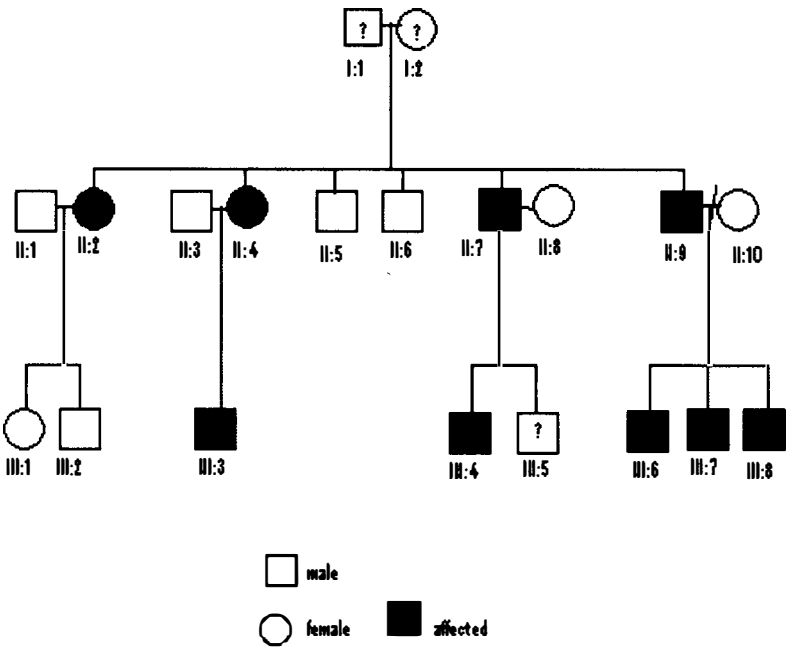


Figure 1: West Australian Pedigree.

Diagram of a pedigree family - affected individuals are depicted by darkened circles and squares. Of a family of 20 people, 9 are affected with the form of HFM that has been linked to chromosome 11.

(Singer *et al.* 1994).

Studies described in this thesis were conducted in a West Australian family afflicted with hemifacial microsomia. Singer, Haan, Slee and Goldblatt (1994) first described this pedigree in 1994 (Figure 1, p6). This family afflicted with HFM, is one of the few families large enough to provide useful genetic linkage analysis to allow identification of a linkage region on chromosome 11, as identified by Chandler (2001).

The symptoms of the nine affected individuals in this family are presented in Table 1.1 (p7).

Table 1.1 Clinical pathologies for West Australian HFM family

(modified from Chandler, 2001)

Pedigree number	Preauricular nodules	Preauricular tags	Epibulbar dermoid	Hemifacial microsomia	Macrostomia (increase in mouth width)	Other ear defects
II:2	Right					
II:4		Left	Left			
II:7		Bilateral				Bilateral
II:9	Right				Mild (right)	
III:3	Right	Right	Right	Right		
III:4		Bilateral		Right		Right
III:6		Unilateral				Right
III:7		Unilateral		Right	Right	
III:8		Bilateral				

There are nine affected individuals in this family, inheriting varying symptoms ranging from pre-auricular nodules and slight asymmetry of the jaw, to pre-auricular tags and hemifacial microsomia. The presence of epibulbar dermoids in two individuals (II:4 and III:3) indicates a possible overlap of chromosome 11 HFM and Goldenhar syndrome, similar to the chromosome 14 linked form of HFM (Singer *et al.*, 1994).

In 1994 individual II:4 had an affected child, III:3 who has an epibulbar dermoid and is more severely affected than his mother. Generally the severity and number of symptoms increase through succeeding generations, thereby indicating some degree of genetic

anticipation. This is most clearly evident in the 'hemifacial microsomia' column of Table 1.1 (p7), where the most pronounced feature of this disease, the under-development of one side of the face, becomes evident in generation III.

Table 1.2 Diagnostic criteria for HFM in WA family (Cousley & Calvert, 1997)	
I	Ipsilateral (one side) mandibular and ear (external/middle) defects
II	Asymmetrical mandibular or ear (external/middle) defects in association with two or more indirectly associated anomalies or a positive family history of hemifacial microsomia

1.4 Diagnostic criteria used to diagnose this family with HFM

The fundamental diagnostic criteria presented in Table 1.2 (p8) were used to diagnose the West Australian family. Presently, there is no agreed minimum diagnostic criterion for HFM however (pre)auricular abnormalities such as nodules or tags are noted as possibly the mildest manifestation of the disease (Rollnick, 1983) and these were found present in the majority of affected family members. Moreover, characteristic unilateral facial anomalies are present in the third generation patients in this family indicating increased severity (Gorlin, 1990). The term 'HFM' is generally used to describe disorders where the most predominant feature is asymmetric hypoplasia, or under-development of one side of the face (Cousley & Calvert, 1997). Mandibular deformations are also present and were a prime criterion used to assess the family under investigation.

1.5 Haplotypes

The West Australian family under investigation comprises 20 members, 8 unaffected and 9 who are diagnosed as HFM positive. Dr David Chandler (2001) showed linkage of this family's hemifacial microsomia disease gene to a region on chromosome 11, between the markers D11S1883 and D11S937. Haplotypes for chromosome 11 are available from D11S1765 to D11S876 for both affected and unaffected individuals within this family (Chandler, 2001) (Figure 2, p10). From the haplotypes it is clear that a proposed linkage region for HFM exists between markers D11S1883 to D11S911, approximately 18.8cM in length (Figure 3, p11).

1.5.1 Refinement of the Chromosome 11 Linkage Region which predisposes the West Australian family to HFM

The region on chromosome 11 which is linked to HFM in the WA family was refined by Chandler (2001) using 152 polymorphic markers, and initially, through exclusion mapping, a Lod score of 2.1 (at theta=0) in an affected individual only analysis was obtained. Genotyping of a further 10 markers between D11S1985 and D11S1396 showed no recombinants between these two markers (Chandler, 2001). A possible recombination was suggested through the negative scores of markers D11S1765 (3'), D11S1883 (3'), D11S937 (5') and D11S1362 (5'). Markers D11S987, D11S4207 and D11S916 derived the highest two-point Lod score of 2.11. This was the highest possible Lod score achievable with this family in an affected-only analysis (Figure 4, p12).

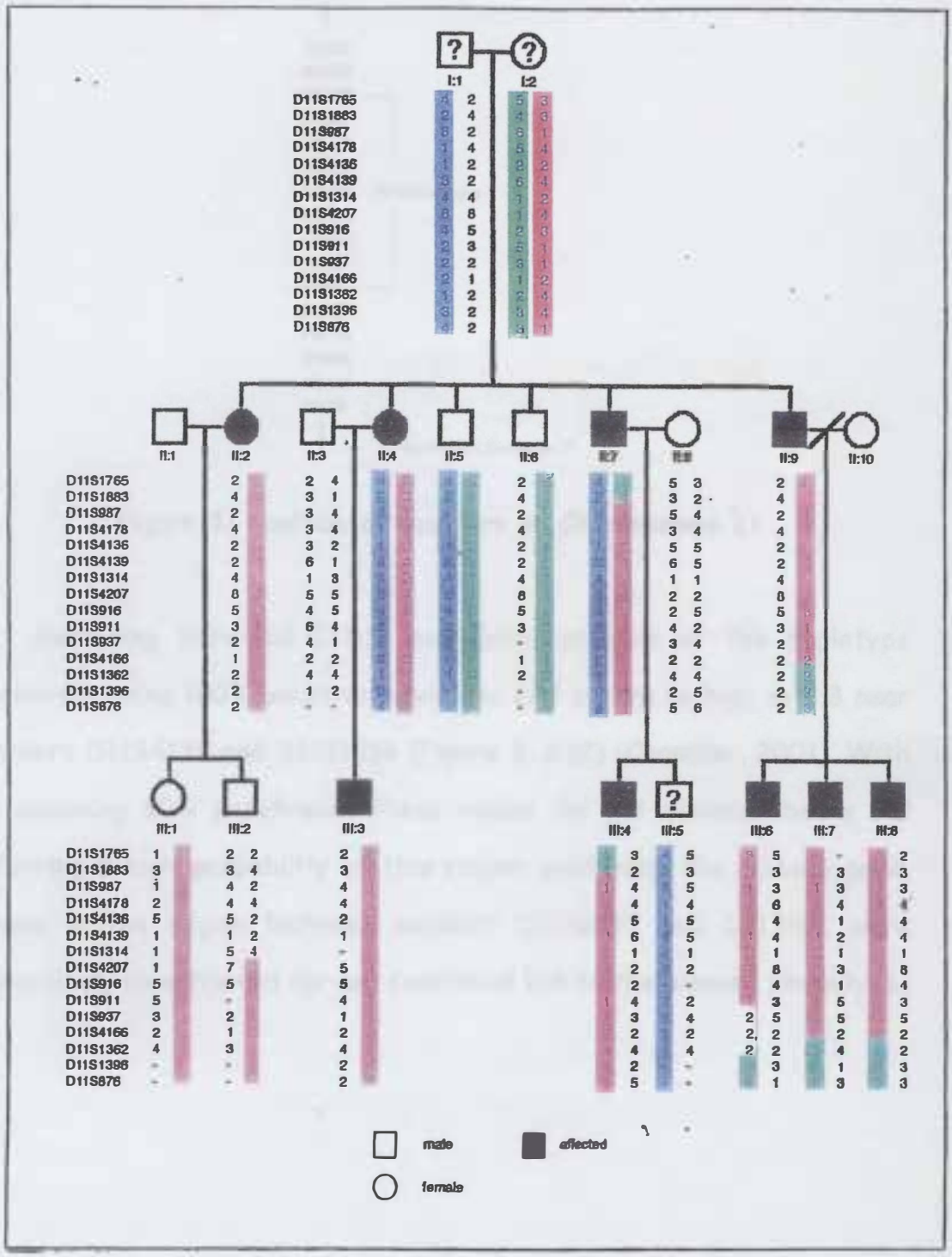


Figure 2: Haplotypes of the West Australian hemifacial microsomia family

Between markers D11S1765 to D11S876. Red bars designate the putative affected haplotype (Adapted from Chandler, 2001)

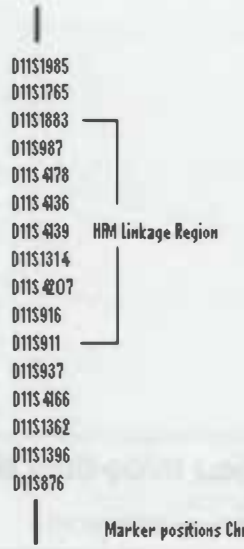


Figure 3: Position of markers on Chromosome 11

Excluding individual III:1, multi-point analysis of the haplotype region assuming 100% penetrance yielded Lod scores as high as 3.3 near markers D11S4139 and D11S1314 (Figure 3, p.12) (Chandler, 2001). With an assuming 80% penetrance these values did not diminish below 3.0 inferring a high probability of this region containing the disease gene. Genes in the region between markers D11S1883 and D11S911 were accordingly investigated for any functional link to the disease phenotype.

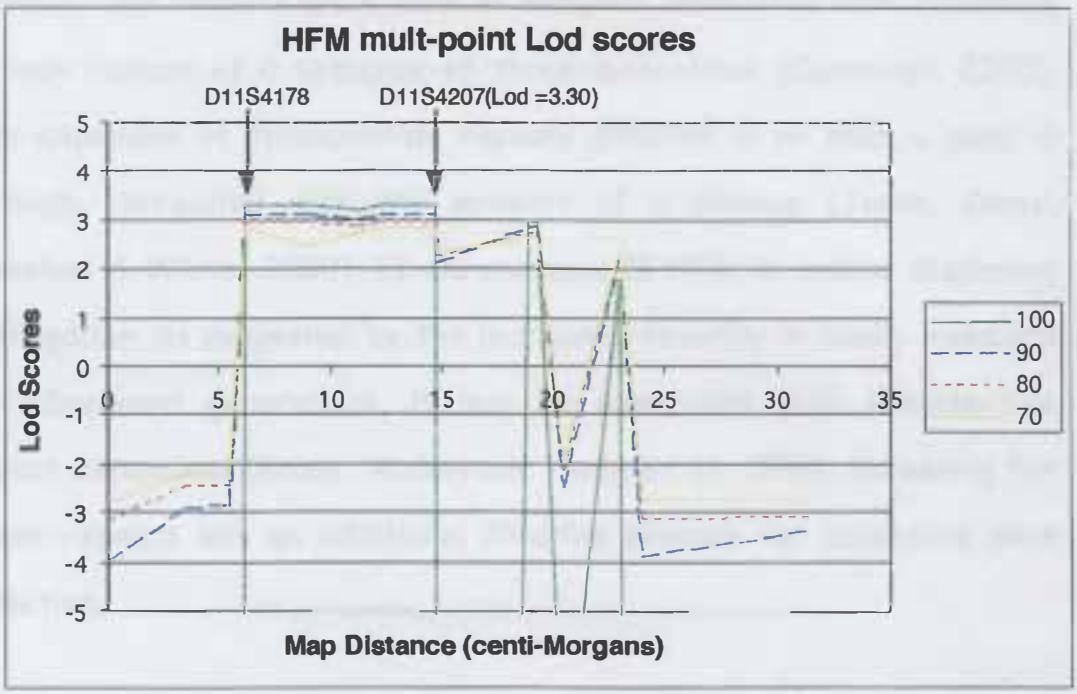


Figure 4: Multi-point Lod score graph

The conserved HFM haplotype region for the West Australian family assuming 70, 80, 90 and 100% penetrance levels. Data from individual III:1 was excluded thereby providing a maximum Lod score of 3.30 at D11S4207 (Adapted from Chandler, 2001).

1.6 Identifying potential chromosome 11 HFM disease genes

Dysmorphology can be caused by mutations in several genes. To identify candidate genes associated with HFM in this family, three major approaches were adopted. Firstly, each gene present in the chromosome 11 HFM linkage area between D11S1883 and D11S911, was investigated for embryonic expression and association with axial (a)symmetry. Moreover, genes were assessed for trinucleotide repeat regions. Trinucleotide repeats are a form of mutation associated with expansion in copy number of a sequence of three nucleotides (Cummings, 2000). The expansion of trinucleotide repeats situated in or near a gene is strongly correlated with the severity of a disease (Jorde, Carey, Bamshad & White, 2000). If chromosome 11 HFM is indeed displaying anticipation, as suggested by the increasing severity in family members of subsequent generations, it may be associated with trinucleotide repeat expansions (Brook, McCurrach, Harle *et al.*, 1992). Screening for these repeats was an additional filtering process for candidate gene selection.

Secondly, regions of the mouse genome homologous to the chromosome 11 HFM linkage area were identified and genes within these regions were investigated for the same criteria. Homologies were determined for three of the five chosen candidate genes; *WNT11, GARP* and *CLNS1A* demonstrated homology to regions on mouse chromosome 7.

Thirdly, as HFM was additionally linked to chromosome 14, a comparison of the genes in the chromosome 11 and 14 HFM linkage regions was undertaken in regard to the same criteria.

In the congenital disorder HFM, as with many malformation syndromes, there are progressive symptoms that may be attributed to normal growth patterns of malformed tissue or continual incorrect expression of a disease gene (characteristics of a likely disease gene were investigated as described above). Advances in gene technology as a result of the Human Genome Project have resulted in availability of a variety of tools which can be implemented to identify candidate genes: these include advanced linkage analysis software, automated genotyping protocols, comprehensive genomic and cDNA databases with full public access (Table 3.1, p82).

As the volume of online data and sophisticated web interfaces increased, the accessibility of powerful linkage and sequence analysis software also increased and hence added to the variety of tools available to be used in this project.

1.7 Candidate Genes (Genome mining)

Identification of candidate genes in the chromosome 11 HFM linkage area was achieved using the aforementioned approaches:

1. Each gene present in the chromosome 11 HFM linkage area was inspected for three criteria, a) embryonic expression, b) association with trinucleotide repeats, and c) association with axial (a)symmetry.

Eleven genes out of a total of 131 genes were identified as potential candidates. In order of likelihood of possessing the mutation responsible for HFM, thereby fulfilling the criteria listed above, they are *UVRAG*, *ARIX*, *WNT11*, *GARP*, *CLNS1A*, *PKRIR*, *CAPN5*, *LGALS12*,

DKFZP564M082 and *OMP*. Of these eleven genes, three were identified as prime candidates, *UVRAG, WNT11* and *ARIX*. These particular three genes were selected as they were located within the linkage region, despite re-mapping of markers on numerous databases and complied with the selection criteria. Furthermore they demonstrated some association with craniofacial development and/or malformation. Chandler (2001) and Kaledijiva (2002) (personal communication) had previously selected the developmental gene *WNT11* as a candidate for HFM and sequenced this gene in both an unaffected and an affected individual from the West Australian chromosome 11 HFM pedigree. The affected and unaffected sequences did not vary significantly from each other or from the published sequence and as no mutations were found, this gene was eliminated as a candidate. *GARP* and *CLNS1A* are genes lying within this HFM candidate region and according to Ensembl may be possible candidate genes and for this reason both will be tested along with *UVRAG* and *ARIX* (Specific information on each of the genes is described in chapter 2.11.2, p 64).

2. Regions of the mouse genome homologous to the human chromosome 11 HFM linkage region were identified through web-searches. Genes lying within homologous regions were inspected for the same three criteria; embryonic expression, association with trinucleotide repeats and association with axial asymmetry. Moreover genes associated with HFM in a number of animal models were investigated for homology with chromosome 11 genes. Significant insight into craniofacial development has been recently unveiled through generation of transgenic mouse mutants. These mice are engineered with mutations in genes responsible for craniofacial development and the resultant craniofacial

syndromes are assessed and compared by classical comparative embryological approaches in a range of species (Jones & Trainor, 2004). Evidence from two mouse models supports genetic involvement in HFM. Hemifacial maxillary malformation in heterozygotes with incomplete penetrance has been observed to result from alteration in the genetic background of the mouse *far (first arch)*, a recessive lethal mouse mutant (Juriloff *et al.*, 1987). The second mouse model is an autosomal dominant insertional mutation on mouse chromosome 10 which produces a *Hfm* phenotype in hemizygous mice which resembles HFM in humans, including microtia, low set ears and an abnormal bite resulting from hypoplasia of the second branchial arch accompanied by local haemorrhage (Naora *et al.*, 1994).

3. A comparison of all the genes in the chromosome 11 and chromosome 14 HFM linkage regions was undertaken and all were interrogated in regard to the same three criteria; embryonic expression, association with trinucleotide repeats and association with axial asymmetry. Genes sharing homology in both HFM linkage regions would become prime HFM candidates, particularly when they fulfilled the three aforementioned criteria. Since the chromosome 14 region has been definitively linked to autosomal dominant HFM (Kelberman *et al.*) this experiment would provide useful information on genes likely to be responsible for chromosome 11 HFM.

1.8 Project Feasibility

This project was an attempt to discover a mutation in a region of DNA approximately 18.8cM in length between markers D11S1883 and D11S911 on chromosome 11. The difficultly of this task is highlighted by

the fact that the mutation could be as small as one altered DNA base. The fact that the causative gene is expressed embryonically further complicates this task as it denies the opportunity to observe the first stages of the disease. Nevertheless this project was thought feasible and involved selection of candidate genes using methods described above and their interrogation by standard laboratory techniques available at Edith Cowan University.

Laboratory methods included the polymerase chain reaction (PCR) of coding regions (exons) of HFM candidate genes, sequencing of these exons (Sanger, 1981), and the comparison of 'affected sequences', 'unaffected sequences' and published sequences. Such methodologies could reveal DNA sequence alterations segregating within affected individuals.

The Draft Human Genome was inspected across several genome databases to monitor the location of the HFM flanking makers, D11S911 and D11S987 on chromosome 11. By doing this on a regular basis, the length of the chromosome 11 HFM candidate region [http://www.ncbi.nlm.nih.gov/genome/guide/human] was reduced, as more detailed analysis of the genome became available. Moreover the candidate genes on chromosome 11 and the chromosome 14 linkage region were compared. This allowed us to confirm candidature of known and predicted genes and to exclude others as HFM candidate genes. This was of assistance as genes identified as potential candidates in preliminary investigations were excluded later if they no longer mapped within the reduced linkage regions.

Secondly, it was thought that the likelihood of this project identifying the causative mutation would be improved through careful selection of HFM candidate genes by genome mining. This method could assist in identification of candidate genes by cross-tabulating relevant details of known and predicted genes in the two HFM candidate regions on chromosome 11 and 14. Information on candidate genes was cross-tabulated with available information in the mouse genome databases. Many mouse genes demonstrate high homology to those of humans and it was thought a feasible approach to assist in determining gene function for the many human chromosome 11 HFM linkage region genes which are of unknown function. Criteria relevant to analysis of chromosome 11, chromosome 14 and mouse genes included: names, aliases, acronyms, functions, disease associations (particularly with axial asymmetry), timing and site of expression, and association with trinucleotide repeat expansions.

1.9 Originality of this research

This project is innovative in that it seeks to identify the mutation causing the form of hemifacial microsomia segregating on chromosome 11. If successful, this would be the first identification of a mutation responsible for this common congenital disorder. Secondly, the adoption of genome mining enhances this project. Genome mining is novel simply because it has only recently become possible to attempt such analyses with sufficient rigour. Prior to the completion of the Human Genome Project in 2004 only moderate-resolution maps of chromosomes 11 were available. The publication of the Draft Human Genome, the rapidly expanding DNA databases for mouse and man, and the availability of

powerful search engines have combined to make such an undertaking possible.

1.10 Overall purpose and significance of this research

The rationale of this project was that identification of a causative mutation responsible for chromosome 11 HFM could improve the diagnosis, prognosis, prevention, treatment and future research into the molecular pathology of autosomal dominant HFM. Although it is impossible to predict what molecular connections future research may find between the differing forms of HFM, it will become successively easier to identify possible causative factors as this and other research discoveries emerge. The enhancement of understanding with regard to craniofacial developmental deformation will greatly improve future prognoses, diagnosis and treatment of such disorders.

Inevitably, as HFM mutations are discovered, functions of the associated gene(s) will become better understood. Studying the effects of disease causing mutations is a powerful way of determining the biological role of gene products and linking genotype to phenotype.

1.11 Conclusion

This project identified four genes, namely *UVRAG, ARIX, CLNS1A* and *GARP* from the chromosome 11 linkage region as being likely candidates for the West Australian family segregating HFM. All four genes were thoroughly investigated and the majority of three genes were sequenced, by implementation of numerous laboratory techniques. In this way this project contributed to enhancing the understanding of HFM through the adoption of genome mining and bioinformatics.

However sequencing of these three genes in an affected and unaffected individual identified no differences from the published sequence. Thus the research presented here excludes these genes as candidates for the disease in this family and identifies a potential candidate gene for future research. As a result, this project of gene mining for hemifacial microsomia, together with advances of the Human Genome Project have refined the candidate gene linkage region associated with HFM on chromosome 11. Ultimately, an enhanced understanding will lead to improved treatment strategies.

Chapter 2: Comprehensive literature overview on HFM

2.0 Comprehensive literature overview on HFM

This chapter reviews the current literature surrounding hemifacial microsomia. It discusses clinical aspects of the disease, assessment and classification, malformation causes, laterality defects, specification of axis formation and the numerous proposed pathologies surrounding HFM. Furthermore, genetics, phenotypic and pathogenetic data will be discussed. Candidate genes selected in this project and animal models for HFM will be described.

2.1 Nomenclature

The term 'hemifacial microsomia', was initially coined by Gorlin *et al.*, 1963 (as cited in Seow *et al.*, 1998) to refer to patients displaying unilateral microtia, macrostomia, and hypoplasia of the mandibular ramus and condyle (Cavaliere & Buchman, 2002). Alternative terminology used to describe HFM includes; first and second branchial arch syndrome, otomandibular dysostosis, lateral facial dysplasia, Goldenhar syndrome, facio-auriculo-vertebral sequence/complex, craniofacial microsomia and oculoauriculovertebral (OAV) dysplasia/spectrum (Wang, Lin & Yi, 2001; Cousley & Calvert, 1997; Jacobsson & Granstrom, 1997; Tiner & Quaroni, 1996; Johnson, Fairhurst & Clarke, 1995; Kaye *et al.*, 1992; Boles, Bodurtha & Nance, 1987; Mansour, Wang, Henkind, Goldberg & Shprintzen, 1985; Yanagihara, Yanagihara & Kabasawa, 1979; Gorlin, Pindborg & Cohen, 1976).

Numerous authors have found increasing evidence, in the form of associated anomalies, supporting the proposition that hemifacial microsomia (HFM), Goldenhar syndrome (GS) and oculoauriculovertebral dysplasia

(OAV) are part of a spectrum within a single entity (Brady *et al.*, 2002; Tiner & Quaroni, 1996; Kaye *et al.*, 1992; Figueroa & Friede, 1985; Yanagihara *et al.*, 1979; Gorlin *et al.*, 1976). OAV *dysplasia* has subsequently been changed to OAV *spectrum*, as dysplasia and syndrome were considered inaccurate descriptions (Cousley & Calvert, 1997). Generally, the term HFM is widely applied, especially in surgical texts, to illustrate the full clinical spectrum of the salient features of facial hypoplasia and axial-asymmetry (Cousley & Calvert, 1997).

2.2 Clinical History

Hemifacial microsomia (HFM), a complex congenital condition (Cousley & Calvert, 1997), is recognized as etiologically and phenotypically heterogenous (Cousley, 1993; Cousley & Wilson, 1992; Bassila & Goldberg, 1989), and has been classified as the second most common birth defect, second to cleft lip and palate (Monahan *et al.*, 2001; Fischer & Prahl-Andersen, 1996; Satoh, Shibata, Tokushige & Onizuka, 1995; Rodgers *et al.*, 1991). Secondary to underdevelopment, HFM is the commonest asymmetrical craniofacial deformity of the craniofacial skeleton and soft tissues (Ji, Li, Shamburger, Jin, Lineaweaver & Zhang, 2002; Whyte, Hourihan, Earley & Sugar, 1990). Furthermore, HFM involves deformities in first and second brachial arch derivatives (Cousley *et al.*, 2002; Sze *et al.*, 2002; Kelberman *et al.*, 2001; Silvestri *et al.*, 1996; Rodgers *et al.*, 1991; Converse, Woodsmith, McCarthy, Cocaro & Becker,1974; Converse *et al.*, 1973). As several structures arise from these arches and not all are necessarily affected, the clinical presentation may vary considerably. During embryonic development of higher vertebrates, six transient arches,

also termed pharyngeal or visceral arches, are formed that give rise to specialised structures in the head and neck [On-line Medical Dictionary: http://cancerweb.ncl.ac.uk/omd/]. The skeletal elements of these arches are derived from neural crest and lateral plate mesoderm (Appendix 2).

The estimated international incidence of HFM ranges from 1 in 1000 births (Horgan *et al.*, 1995), to between 1 in 5600 births (Cousley & Calvert, 1997; Cohen, 1991; Gorlin, Cohin & Levin, 1990; Grabb, 1965;) and 1 in 4000 births from studies conducted in London (Poswillo, 1973). Although most cases are considered sporadic, the estimated recurrence rate in first-degree relatives is approximately 2-3% (Rollnick & Kaye, 1983).

HFM is a complex deformation that displays a spectrum of anomalies involving soft-tissue and skeletal structures (Carvalho, Song, Vargervik & Lalwani, 1999; Cousley, 1993). It is characterized by varying degrees of facial asymmetry (the general causes of which are outlined in Table 2.1, p26) and hypoplasia of the bony and/or muscular and soft tissues of the face (Loescher, 1996), auricular anomalies, skin tags, pits and different gradations of microtia (Gorlin *et al.*, 1976). Extracraniofacial-associated malformations can occur in cardiac, pulmonary, gastrointestinal, renal, and central nervous systems (Horgan *et al.*, 1995; Lauritzen *et al.*, 1985). An example of the extreme variety of anomalies associated with HFM includes the presence of cartilage in cervical appendages (Rollnick & Kaye, 1985). The majority of reported cases of HFM are considered sporadic in origin, however substantial evidence for genetic involvement has come to light in recent years (Cousley *et al.*, 2002; Connor, & Fernandez, 1984), including

rare familial cases that display autosomal dominant inheritance (Kelberman *et al.*, 2001), as is the case in the West Australian family under investigation.

Table 2.1 Causes of asymmetry in humans

Causes	Examples
Gene Mutations	Kartagener syndrome
Embryopathies	
Malformations	Unilateral cleft lip
Disruptions	Amniotic band disruption
Fetopathies	
Deformations	Mandibular asymmetry
Hemi-asymmetries	
Hemihyperplasia	Beckwith-Weidemann syndrome
Hemihypoplasia	Hemifacial microsomia
Hemiatrophy	Romberg syndrome
Craniosynostoses	Unilateral coronal synostosis
Hamartoses	Sturge-Weber angiomatosis
Common and/or well-known	
Infections	Facial cellulitis of dental origin
Cysts	Lymphoepithelial cyst
Tumors	Pleomorphic adenoma of parotid
Trauma	Unilateral temporomandibular joint ankylosis
Others	Fibrous dysplasia

Adapted from Cohen (1985a)

2.3 Clinical Ascertainment

The clinical presentation of HFM generally manifests as one of three principal deformities (Silvestri, Natali & Fadda, 1996; Kay & Kay, 1989; Mulliken & Kaban, 1987): 1) auricular hypoplasia, 2) mandibular hypoplasia, or 3) hypoplasia of the soft tissues. The orofacial manifestations of HFM are characteristically unilateral, although is has been suggested that bilateral

involvement can occur, with more severe expression on one side (Seow *et al.*, 1998; Kallen, Harris & Robert, 1996; Bassila & Goldberg, 1989; Figueroa & Pruzansky, 1982) (Appendix 2).

A patient displaying HFM in its fullest expression would typically exhibit unilateral (rarely bilateral) underdevelopment of the middle and external ear, mandible, zygoma, maxilla, temporal bone, facial muscles, mastication muscles, palatal muscles, the tongue and parotid gland, as well as macrostomia and a first branchial cleft sinus (Grabb, 1965) (Appendix 3). Fortunately however, HFM is frequently incompletely expressed. The two major features of deformation in HFM are auricular and mandibular hypoplasia (Converse, Coccaro, Becker & Wood-Smith, 1973).

Other variations of HFM may demonstrate predominance of either auricular or mandibular dysplasia (Converse, Coccaro, Becker & Wood-Smith, 1973). In cases exhibiting predominant mandibular dysplasia, additional deformities characteristically associated with HFM such as pre-auricular tags are less prevalent.

Anomalies deriving from the variety of malformations of the branchial arches in HFM have been reported to include congenital heart disease, contracture of a finger, acetabular dysplasia, clubfoot, imperforate anus with recto-vaginal-urethral fistula, bilateral failure of rotation of the kidneys and hypospadias (Grabb, 1965). Alternatively, maximal auricular deformity may be evident while the mandibular deformity may be unapparent upon clinical examination. Roentenographic studies (a derivative

of radiology) have demonstrated that every case of external auditory canal and auricular hypoplasia with middle ear deformation encompasses skeletal changes on the affected side, suggesting the severity of ear deformity does not parallel that of the mandibular defect (Murray, Kaban & Mulliken, 1984).

Mandibular deformation characteristic of HFM may present without significant auricular or temporal bone maldevelopment. Differentiation between these cases and postnatal deformities caused by injury is difficult, although the diagnosis is evident if the deformity was present at birth.

In HFM there is a specific, but none the less incorrect, tendency to consider hypoplasia as predominantly bony. However muscular hypoplasia, involving powerful muscles such as the masseter, medial and lateral pterygoid and temporalis can emerge which would undoubtedly influence skeletal development (Converse, Coccaro, Becker & Wood-Smith, 1973). This impairment in muscle function has an apparent impact, not only on the developing musculature but also on the morphologic character of the attached bones and nerves. Hence paralysis or weakness of the facial nerves is an occasional finding.

Other anomalies associated with HFM may include microtia-atresia; mandibular, maxillary and orbital hypoplasia; absence of the parotid gland; fistulae, microphthalmia, coloboma of iris and choroid, epibulbar lipodermoid, strabismus, macrostomia, conductive or sensorineural hearing loss and hypoplastic facial muscles (Converse, Coccaro, Becker & Wood-

Smith, 1973). It is evident that HFM is an encompassing term that covers a variety of congenital defects. Yet the extremely variable phenotype revolves around the universal feature of axial asymmetry (Cousley & Wilson, 1992; Rollnick & Kaye, 1985). In some cases this asymmetry is masked by normal adjacent soft tissues. This especially applies to young children whose faces are generally more buxom, thereby expressing an incomplete phenotype. Facial soft-tissue deformities ranges from clinically normal to severely deficient on an affected side(s) (Murray, Kaban & Mulliken, 1984). Patients who exhibit mild involvement typically display minimal deformity, including ear tags, slight macrostomia and minimal subcutaneous-muscular hypoplasia without auricular or cranial nerve involvement. On the other hand, severe soft-tissue deformities refer to patients with major contour deficiency, neuromuscular weakness and major soft tissue deficiency associated with ear distortion, nerve deficits and/or clefts of the face or lips (Satoh *et al.*, 1995; Mulliken & Kaban, 1987; Murray, Kaban & Mulliken, 1984). Between these two extremes, patients are classified as moderate.

Both HFM and the branchio-oto-renal syndrome (BOR) are associated with malformations of the external ears, preauricular tags, pits, or sinuses and conductive or mixed hearing loss (Abdelhak *et al.*, 1997; Jacobsson & Granstrom, 1996; Sensi, Cocchi, Martini, Garani, Trevisi, & Calzolari, 1996; Rollnick & Kaye, 1985). However, both sensorineural hearing loss and facial nerve dysfunction are common features specific to HFM (Brady *et al.*, 2002; Carvalho *et al.*, 1999; Cousley, 1993).

2.4 Classification

HFM is typically not diagnosed until late infancy or early childhood, as recognizable features such as facial asymmetry are subtle in newborn infants (Carvalho *et al.*, 1999; Marsh, Baca, & Vannier, 1989). Accurate facial anomaly assessment is essential for effective diagnosis, classification and management of HFM (Cousley & Calvert, 1997; Converse, Wood-Smith, McCarthy, Coccaro & Becker, 1974).

The spectrum of anomalies consistent with HFM has made systematic and inclusive classification difficult (Vento, LaBrie & Mulliken, 1991; Bassila & Goldberg, 1989). Subsequently, numerous systems have been devised to facilitate the classification of individualized components of this complex condition (Polley, Figueroa, Liou & Cohen, 1996; Rodgers *et al.*, 1991), although a system suitable for universal adoption is yet to be devised (Polley *et al.*, 1996; Figueroa & Pruzansky, 1982). Classification systems categorizing patients as HFM type I, type II etc. are not recommended since they are solely based on mandibular deformation, thereby neglecting the multitude of other anomalies associated with HFM (Mulliken & Kaban, 1987; Figueroa & Pruzansky, 1982; Swanson & Murray, 1978).

Keusch, Mulliken and Kaplan (1990) recommend the distinction between malformation and deformation when referring to congenital structural anomalies. Malformations specifically incorporate errors of morphogenesis in which intrinsically abnormal structures develop throughout gestation, whereas deformations result from abnormal mechanical forces on presumably intrinsically normal structures. Although effective in

categorising the origin for a predominant congenital anomaly, this classification system oversimplifies a complex disorder (Keusch *et al.*, 1990).

The most favourable systems of classification used in HFM diagnosis are (1) the Skeletal, Auricle and Soft-tissue (SAT) multisystem classification, proposed in 1987 by David and colleagues, and (2) the Orbital distortion, Mandibular hypoplasia, Ear anomaly, Nerve involvement and Soft tissue deficiency (OMENS) classification system, proposed by Vento, LaBrie and Mulliken (1991). The SAT system allows each component addressed (S = skeletal, A = auricle, T = soft tissue) to be graded according to the severity of involvement (Appendix 4). Similarly, the OMENS system permits grading according to the severity of involvement of each component. The differentiation between key phenotypic elements will lead to improvements in diagnosis, prognostic predictions, treatment planning, data evaluation and case correspondence (Cousley & Calvert, 1997). Both these systems are useful in HFM classification as they allow bilaterally affected patients to be classified by analysis of each affected side separately. The SAT system is perhaps less comprehensive than the OMENS system as it does not provide a separate grading category for nerve involvement. Neither system allows for cardiac, renal, nervous system or other anomalies that have been associated with HFM (Rodgers *et al.*, 1991). Another flaw of the OMENS system is that the categorisation of the five major factors (O.M.E.N.S) displays bias as the authors themselves have decided what is to be considered 'major'. Yet there is well established evidence to support the inclusion of cardiovascular anomalies as part of the

oculoauriculovertebral spectrum (Casey, Braddock, Haskins, Carey & Morales, 1996; Cohen, 1991). The diagnostic criteria used to assess the WA family under investigation by Singer *et al.* (1994) for degree of deformation is tabulated in Table 1.2 (p8).

Three other classification systems deserve brief mention for their contribution to the classification of HFM and associated anomalies: the Pruzansky classification system, Tenconi and Hall's phenotypic classification system and Munro and Lauritzen's system (all cited in Rodgers *et al.*, 1991). These are alternative classification systems devised to guide treatment planning and diagnosis of HFM, although they are not as widely accepted as the SAT and OMENS classification systems.

The Pruzansky classification system is particularly effective in describing mandibular deformities characteristic of HFM, however this system has two primary disadvantages, namely failing to include the absence of condylar description and to address the variety of other anomalies frequently observed in HFM (Pruzansky, 1969).

A disadvantage of the Tenconi and Hall phenotypic classification system is that nerve involvement is not acknowledged and ear abnormalities are not addressed, despite being a frequent feature of HFM.

Munro and Lauritzen's system is unique in that it is an "anatomical-surgical" classification system. The various categories are distinct, reproducible and thorough however this system does not account for nerve

involvement, which becomes increasingly evident with deformation severity (Rodgers *et al.*, 1991).

Through reviewing the numerous classification systems in existence to date and considering the inherent variability of deformities of HFM, a multi-system classification seems justified (David, Mahatumarat & Cooter, 1987). In reference to problems associated with HFM classification, it is the proximity of the structures deriving from the branchial arches that complicates classification, as many of the structures affected are comparable to Goldenhar syndrome, rather than specifically to a diagnosis of hemifacial microsomia alone (Converse, Coccaro, Becker & Wood-Smith, 1973). This is evident in the WA family under investigation. In an ideal situation, the anatomical structure, aforementioned deriving from the first two branchial arches, would not overlap and would be clearly defined. This would encourage more accurate identification of abnormal anatomical structures and thereby support a precise method for classification and diagnosis.

2.5 Causes of Malformations

Teratogens and genetic factors are two predominant factors known to cause malformations. Their effects play an important role in development, for example susceptibility or resistance to certain teratogens may be provided by genetic factors, thereby complicating interpretation. The determination of the relative importance of genetic and environmental effects as a cause of malformation is made challenging by the overlap between these factors. Independent consideration can be artificial but

expedient in circumstances where either an environmental factor or specific genetic problem is presumably the sole cause of malformation. Teratogens have been subject to considerable research as they are a potentially preventable cause of malformation.

Alternatively numerous studies have confirmed genetic involvement in malformation, including recurrence rate of common malformations such as cleft lip or neural tube defects in family studies (Clayton-Smith & Donnai, 1997) as well as the existence of rare, Mendelian inherited malformations (Wilkie, 1994). With regards specifically to HFM, Otani *et al.* (1991) demonstrated the involvement of both teratogens and genes by determining that hypoplasia of the branchial arch was not the primary cause but a result, and that vasculature served as a threshold for the defective gene or drug effect.

2.6 Molecular basis of dysmorphology

2.6.1 Embryology

Pre-implantation involves ovulation, fertilisation and cleavage, blastocyst formation and zygote implantation. Genetically normal sperm and ova are evidently crucial for the initiation of this process. Aside from non-fatal genetic defects, errors occurring during pre-implantation usually result in early pregnancy termination rather than malformation (Martini, Ober, Garrison, Welch & Hutchings, 1998).

Proliferation and condensation of mesenchyme to form somites, alongside marked ectodermal activity is the primary embryologic change

inaugurating auricle formation in the fourth and sixth weeks of embryonic development (Stark & Saunders, 1962). Patterns formed during these initial stages of development form the template for the future body plan. Pharyngeal arches are formed during the fourth week of embryonic development and can be likened to the gills of fish. They reflect the fact that that the development of an individual (ontogeny) resembles the species evolution (phylogeny) and are important in appreciating the final structure and innervation of the head and neck (Zillmusom, 2003). Initially the branchial arches consist of a mesodermal core surrounded by epithelium (Shanahan, 2004). The mesenchyme will form muscles, arteries, connective tissue, cartilage and parts of the skeleton. Each of these arches contains a cartilaginous core, an aortic arch and a cranial nerve which eventually supplies the structures that develop from the mesenchyme of the arch (Marino, 2004). Structures derived from the branchial arches are outlined in Table 2.2 (p35).

Table 2.2 Structures derived from branchial arches

Arch/ Nerve	Skeletal	Ligaments	Muscles	Pouch
First (V)	1. Malleus 2. Incus	1. Anterior ligament of malleus 2. Sphenomandibular ligament	1. Muscles of mastication 2. Tensor tympani 3. Tensor palati 4. Anterior belly of Digastric	1. Auditory tube 2. Tympanic cavity
Second (VII)	1. Stapes 2. Styloid process 3. Hyoid bone-lesser horn, upper half of body	Stylohoid ligament	1. Muscles of facial expression 2. Stapedius 3. Stylohoid	Lining (crypts) of palatine tonsils
Third (VIII)	Hyoid bone-greater horn lower half of body		Digastric Stylopharyngeus	1. Inferior parathyroid gland 2. Thymus
Fourth (X)	Cartilages of larynx		1. All muscles of larynx 2. All muscles of pharynx (except stylopharyngeus)	1. Superior parathyroid gland 2. C-cells of thyroid
Sixth (XI)			1. Sternocleidomastoid 2. Trapezius	

Adapted from Zillmusom (2003)

Patterning genes are responsible for governing cell differentiation in regions of overlapping gradients along rostro-caudal, dorso-ventral and medial-lateral axes across the embryo (Gavalas, Davenne, Lumsden, Chambon, & Rijli, 1997). Through the development of the body pattern, cells migrate to their assigned positions and differentiate into specific tissues.

Gene expression coding for transcription factors, adhesion factors, receptors and structural proteins permit the complex signalling systems between cells, the extracellular matrix and soluble factors that brings about co-ordinated development of the embryo.

At the region of the mid- and hind-brain, patterns of gene expression are transferred from the neuroectoderm to the oral ectoderm derived epithelia which in turn form a number of tissues, including cartilage, bone, tooth and tongue. It has been suggested that co-ordinated HOX genes, growth factors and receptors orchestrate the timing and positional information for the distribution, determination and demarcation of these tissues.

2.6.2 Genes involved in Dysmorphology

Birth defects are caused by mutations in a variety of genes including transcription and growth factors, receptors and signal transduction molecules, enzymes, transporters and structural proteins (Table 2.4, p40). DNA disruptions too small to distinguish using cytogenetic techniques are known to cause over 1750 inherited malformations (Wilson, 1992). Of these, disorders, over 1000 are multiple defect syndromes. By 1997, 38 genes had been found and a further 20 mapped by positional cloning (Winter, 1998). The following year, more than 3000 non-chromosomal developmental syndromes were reported, of which 2000 seemed to involve single gene aetiology. Genes associated with 200 of these syndromes have been located and a further 120 syndromes mapped (Winter, 1998). Evidently, the pace of

gene discovery for malformations has increased alongside the pace of discovery of all human genes.

2.7 Developmental genetics

The neural crest is an embryonic cell population that originates at the border of the neural plate and prospective epidermis (Basch, Garcia-Castro & Bronner-Fraser, 2004) and throughout the period of embryonic neural tube closure, neural crest cells emigrate from the neural tube, migrate along defined paths and differentiate into a range of derivatives, including the craniofacial skeleton, pigment cells and the peripheral nervous system. Many of the more common craniofacial birth defects are a corollary of abnormal neural crest development, a proposed pathological cause of HFM. Despite the extensive research on the migration and differentiation of the neural crest, research is reasonably new in investigations of how the tissue originates. Due to the fact that craniofacial abnormalities account for a third of all human congenital defects, it is essential to strive for understanding of the patterning mechanisms responsible for controlling head development, in particular the neural crest cell derivatives and their contribution to connective tissue, bones and nerves (Trainor & Krumlauf, 2000).

The migration of neural crest cells occurs in three stages; initiation, dispersion and cessation of migration. Initiation involves the process by which the neural crest cells undergo an epithelial to mesenchymal transition, thus causing the cells to break free from the neural tube. Results of this process include an increase in intercellular space and increased motility in

the neural crest cells. In chick embryo's, *slug*, a transcription factor is thought to be responsible for initiating these factors (Stark & Saunders, 1962). The second migratory stage, dispersion, involves the migration of cells to their final destinations where they proliferate. Two mechanisms are thought to guide the control by which the cells migrate, namely contact inhibition and contact guidance. And finally, cessation of migration involves a reversal of the mechanisms used to initiate migration (Stark & Saunders, 1962).

2.8 Axis Specification

Axes specification and formation are crucial developmental events as they determine the body plan orientation. Axial patterning is a critical component of ontogeny and the developmental process whereby an embryo establishes the orientation of its body axis (or axes) and subdivides the axis into distinct regions. Fundamentally, the anterior/posterior axis is used to describe the axis between the head and the tail; the dorsal/vental axis is used to describe the axis from the back (dorsum) to the belly (ventrum); and the medial/lateral axis is used to distinguish between the middle and lateral sides of the body. Moreover, the vertebrate body is not symmetrical across the L-R axis, instead there are two predominant sides, left and right. Generally the heart and spleen are located on the left whereas the liver is located on the right.

In regards to axial skeleton development, the ventral portion of the sclerotome surrounds the notochord and forms the rudiment of the vertebral body. The dorsal portion of the sclerotome surrounds the neural

tube and forms the rudimentary vertebral arches. The primary signal or sclerotome induction appears to be a notochord-produced factor, *Shh*. *Pax* genes are additionally involved in mediating interactions between the notochord and the developing sclerotome (Bonaventure, 1998).

2.8.1 Formation of the Anterior/Posterior (A/P) Axis

The primitive streak is responsible for determining the A/P axis of a developing mammalian embryo. Patterning along the A/P axis is controlled by a cluster of genes encoding transcription factors that contain a 60-amino acid DNA-binding domain (homeodomain). These genes compose the homeotic gene complex (*HOM-C*) in *Drosophila*, the organism in which they were first isolated through mutation identification (Shankland, 2004). Four copies of *HOM-C* (*HOX* A through D) are found in the human and the mouse. There are 39 *HOX* genes, each 100kb gene cluster located on different chromosomes (Jorde *et al.*, 2000; Bonaventure, 1998). These HOX genes act to specify specific regional identities along the AP axis of the developing organism (Gilbert, 1998; Finnerty, 1994).

2.8.1.1 Homeobox genes

Homeobox genes regulate the embryonic pattern along the anterior-posterior axis. The name '*Homeobox*' genes was originally derived from a multitude of *Drosophila* phenotypes termed 'homeotic transformations'. Homeotic transformations are mutations that alter the identity of particular segments, transforming them into copies of alternate segments. Mutations in the drosophila homeotic gene complex (HOMC) account for dysmorphic phenotypes (Gellon & McGinnis, 1998). HOMC contains a cluster

Table 2.4 Genes involved in dysmorphology

Class of gene	Gene	Disorder	Reference
Transcription factors			
PAX	PAX3, PAX 6	Waardenburg syndrome (WS1), WS2, WS3 Aniridia	Hoth *et al.*, 1993 Jordan *et al.*, 1992
HOX	HOXD13	Synpolydactyly	Muragaki, Mundlos, Upton & Olsen, 1996
MSX	MSX2	Craniosynstosis	Li *et al.*, 1993
Zinc fingers	GLI3/GCPS	Postaxial polydactyly	Radhakrishna, Wild, Grzeschik & Antonarakis, 1997
	ZIC3	Visceral heterotaxy	Gebbia *et al.*, 1997
HMG domain	SOX9	Campomelic Dysplasia	Foster *et al.*, 1994
Helix-loop helix / Leucine zipper	MITF	WS2A	Tassabehji, Newton & Read, 1994
POU domain	POU3F4	Mixed deafness	de Kok, 1995
Structual proteins			
Collagen	COL1A1	Osteogenesis imperfecta	de Vries & de Wet, 1986
Fibrillin	FBN2	Congenital contractural Camptodactyly	Putnam, Zhang, Ramirez & Milewicz, 1995

Enzymatic deficiencies			
Lysosomal enzymes	alpha-1-iduronidase	Hurler/Scheie syndrome	Moskowitz, Tieu & Neufeld, 1993
Peroxisomal enzymes	Actl-CoA oxidase	Pseudonatal Adrenoleukodystrophy	Fournier *et al.*, 1994
Receptor/Signal transducers			
Growth factor receptors	FGFR1	Pfeiffer syndrome	Muenke *et al.*, 1994
	FGFR2	Crouzon syndrome	Reardon *et al.*, 1994
	FGFR3	Achondroplasia	Shiang *et al.*, 1994
Hormone receptors	PTH-PTHrP Receptor	Metaphyseal chondrodysplasia	Shiapani, Kruse & Juppner, 1995
Other receptors, cell adhesion molecules and gap junctions			
Endothlin receptors	EDNRB	Hirschsprung disease	Puffenberger *et al.*, 1994
Adhesion molecules	KAL	Kallmann syndrome	Bick *et al.*, 1992
	Connexin 43	Cardiac Malformations	Britz-Cunningham, Shah, Zuppan & Fletcher, 1995
GEFs	FGD1	Aarskog-Scott syndrome	Pasteris *et al.*, 1994
G proteins	LIS1/MDCR	Lissencephaly	Chong *et al.*, 1996
	GNAS1	Albright heridatary Osteodystropathy	Patten *et al.*, 1990

Adapted from Chandler, 2001

of eight homeobox genes. Every gene in the HOMC cluster contains a region known as a 'homeobox', which codes for a 60 amino acid part of protein termed the 'homeodomain'. Homeodomains are DNA binding domains of proteins that enhance gene expression of genes to which the protein is bound. HOMC gene expression and mutation effects suggest the involvement of genes of the HOMC, in regulation of segmentation along the developing embryo's anterior/posterior (AP) axis. Interestingly, the homeobox sequence is exceptionally conserved throughout phylogeny (Russell, 2002; Gellon & McGinnis, 1998). Insect HOMC expression variation has been suggested as partially responsible for body form variation. In all vertebrates, there are four groups of *HOX* genes (*HOX*A-D) on four different chromosomes (Krumlauf, 1993; Acampora *et al.*, 1989). It has been postulated that increasing numbers of *HOX* genes would accommodate the increasing complexity of the vertebrate body form (Acampora *et al.*, 1989).

HOX genes are expressed along the dorsal axis from the anterior boundary of the hindbrain to the tail, certain derivatives are expressed in specific body parts such as the head, thorax and abdomen (Gilbert, 1998). 3' *HOX* genes are expressed earlier (temporal collinearity during pattern formation along the primary and secondary axes of vertebrates) than and anterior to the 5' *HOX* genes (Bonaventure, 1998). The overlapping domains of *HOX* gene expression generate combinations of codes that specify the positions of cells and tissues along the AP axis of the trunk and limbs (Jorde *et al.*, 2000).

2.8.2 Specification of the D-V axis in the developing embryo (Embryonic patterning)

D-V axis asymmetry is generated through a cascade of signalling molecules, beginning with asymmetric expression of *Sonic hedgehog* (*Shh*) from the ventral side and BMPs from the dorsal surface. Furthermore, *Shh* together with BMPs are responsible for dorsal/ventral patterning of the central nervous system, induction of the neural plate, patterning of the limbs and establishment of the embryonic L/R axis (axis specification). It is also a key mediator of the polarizing activity that regulates patterning along the anterior-posterior (AP) axis and has been observed to influence the signalling pathways regulating calvarial growth and cranial suture morphogenesis (Ingham & McMahon, 2001; Bonaventure, 1998). Ingham and McMahon (2001) observed expression of genes by *Shh* in three key signalling centres in the vertebrate embryo; the notochord, a mesodermal rod that underlies the ventral neural tube, the floor plate, a specialized population of support cells at the ventral midline of the developing central nervous system, and the zone of polarizing activity (ZPA), a population of apical, posterior mesenchyme cells in the limb bud. The notochord is responsible for expression of *Shh* which is both necessary and sufficient for the introduction of distinct ventral cell identities in the spinal cord (Ingham & McMahon, 2001). Cell-autonomous activation or inhibition of *Shh* signalling within the neural tube indicates that the protein acts both directly and at long range to specify cell fate (Ingham & McMahon, 2001). Both *Pax* and a diverse group of homeodomain-containing transcriptional regulators provide the initial response to *Shh* signalling (Ingham & McMahon, 2001; Lacombe, 1989).

The TGF- β supergene family comprises a large group of structurally related genes that encode proteins that form homodimers or heterodimers. Members of the TGF-β superfamily exert a diverse range of biological effects on cell growth and differentiation, such as governing embryonic axial patterning and restricting *Shh* expression (Dupont & colleagues, 2003). Various members of the TGF-β superfamily, including BMPs, produced by the dorsal ectoderm and roof plate of the dorsal neural tube appear to be the dominant influence in the specification of dorsal cell identities (Ingham & McMahon, 2001).

Wnt is an additional patterning gene family critical in the establishment of the primary D/V and A/P axes in vertebrates (Chen & Johnson, 2002; Saint-Jennet, He, Varmus and David, 1997; Ekker, 1993) and is involved in the induction of head structures (Gilbert, 2000). This gene family is responsible for encoding secreted glycoproteins that participate in a variety of developmental processes, including specification of the dorsal/ventral axis and formation of the brain, muscle, gonads and kidney (Jorde *et al.*, 2000).

2.8.3 Formation of the medial-lateral (M/L) and left-Right (L/R) Axes

The cellular and molecular mechanisms that regulate regional specification of the forebrain are largely unknown. Shimamura and Rubenstein (1997) identified *Shh* as the molecular regulator of the medial neural and thus medial-lateral asymmetry is embryonically established by expression of *Shh* at the mid-line. In contrast, gene expression in the lateral neural plate, (Shimamura & Rubenstein, 1997) is regulated by non-

neural ectoderm and BMPs. This suggests employment of similar patterning mechanisms across both the medial-lateral and dorsal-ventral axes.

The left and right sides of the embryo and organs must also be patterned. *Nodal*, a member of the TGFß family is responsible, in part, for the distinction between the left and the right side and is thus a crucial element of left/right axis formation. The mechanisms by which *nodal* expression is activated however, differs among vertebrate classes. As the primitive streak reaches maximum length, transcription of *Shh* gene ceases on the right side due to expression on this side of *activitin* and its receptor. The expression of *activin* blocks the expression of *Shh* and through a cascade of events prevents transcription of *caronte* genes, the absence of which results in BMPs blocking expression of *nodal* and *lefty-2* (Gilbert, 2000).

Despite the fact that mechanisms initiating left-right patterning may be distinct in different vertebrate classes, the conservation of asymmetric gene expression patterns of *nodal* and *lefty* in the left lateral plate mesoderm has become apparent (Bisgrove & Yost, 2001).

Further evidence for the role of *nodal* in L/R axis specification has come from the observation that L/R expression of *nodal* is randomised in a naturally occurring mouse mutant that exhibits a form of randomised organ asymmetry called *inversus viscerum* (Jorde *et al.*, 2000; Jurliff & Harris, 1983).

2.8.4 Craniofacial development

Hox genes play a critical and definitive role in the patterning of regions within the craniofacial complex. In vertebrates, the *Hox* genes are involved in patterning regions of the hindbrain and second branchial arch. By contrast, the first branchial arch is patterned by groups of homeobox genes distinct from those of the *Hox* gene clusters. Trainor and Krumlauf (2000) argued that *Hox* gene identity of the neural crest is pre-patterned, carrying positional information acquired in the hindbrain to developing branchial arches. More recently, investigations have revealed plasticity of *Hox* gene expression in the hindbrain and cranial neural crest of chick, mouse and zebrafish embryos, thus postulating that craniofacial development is regulated by a complex integration of cell and tissue interaction, rather than neural crest pre-patterning, as once assumed.

Pharyngeal arch development is a complex process involving a number of distinct embryonic populations, the ectoderm, endoderm, neural crest and mesoderm, all of which have to be coordinated in order to generate the necessary components and identity of each branchial arch (Graham & Smith, 2001; Graham, Hixon, Bacino, Daack-Hirsch, Semina, Murray, 1995; Cousley & Wilson, 1992). In mammalian embryos, neural crest cells from the presumptive forebrain and midbrain region give rise to the nasal processes, palate, and mesenchyme of the first pharyngeal pouch. This mesenchyme in turn forms the maxillar, mandible, incus, and malleus. The neural crest cells of the presumptive anterior hindbrain migrate and differentiate to become the mesenchyme of the second pharyngeal pouch and the stapes and facial cartilage. Cervical neural crest cells produce the mesenchyme of the third,

fourth, fifth and sixth pharyngeal arches (in humans, the sixth pharyngeal arch degenerates). This mesenchymal tissue becomes the muscle and bone of the neck. Bones of the skull are thought to develop directly from mesenchyme produced by neural crest cells (Jorde *et al.*, 2000; Bonaventure, 1998).

2.9 Pathogenesis of HFM

Theories explaining the pathogenesis of HFM are many and varied. Previously, the study of the causative mechanism behind HFM has been focused on developmental defects involving the 1st and 2nd branchial arches derivatives. Gorlin *et al.* (1990) emphasized the potential pathogenic significance of abnormalities involving the vasculature that supplies the developing embryos cephalic mesoderm, and thus vascular impairment of vessels supplying the branchial arches of the embryo (Satoh, Shibata, Tokushige & Onizuka, 1995), could cause a deformity (Robinson, Hoyme, Edwards & Jones, 1987).

Hematoma formation and the consequent tissue destruction that follows is an alternative widely accepted mechanism of HFM induction (Brown & Salinas, 1984). Theoretically, this could be induced by a developmental defect of the stapedial artery, a hypothesis encouraged through an appreciation of the early vasculature of the embryonic face (Robinson, Hoyme, Edwards & Jones, 1987). Further findings by Robinson *et al.* (1987) support the hypothesis that the developmental pathogenesis of some unilateral craniofacial defects, including HFM are secondary to interruption in embryonic blood flow (Johnston & Bronsky, 1991; Robinson,

Hoyme & Edward, 1987; Poswillo, 1973). Tissue ischemia and necrosis both result from an interruption of blood flow to developing facial structures. 'Catch-up growth' would consequently vary, depending on the extent and degree of tissue injury, and would present phenotypically as in HFM patients. Louryan, Heymanns and Goffard (1995) opposed this theory by arguing that the stapedial artery did not play a direct role in the genesis of branchial and particularly ossicular, abnormalities. Yet both latter theories were supported by Kearns *et al.* (2000) as they were consistent with the variable and asymmetric nature of HFM. However, neither theory explains the progressive nature of deformation following birth. Furthermore, craniofacial defects can appear on the side opposite to a possible haematoma. Hematoma formation can explain some HFM cases, it can not explain them all.

Evidence for an alternative developmental mechanism is provided through the stimulation of facial and asymmetric brain development following evidence of acute maternal teratogen exposure during gastrulation. Acute ethanol exposure for example can compromise embryonic development (Brown & Salinas, 1984). Poswillo's work (1973) further illustrates this by using triazine in a pregnant rat and thalidomide in a pregnant monkey. Within mere days of administering the treatment, haemorrhage and hematoma formation resulted in the areas of stapedial artery distribution. Interestingly, asymmetric craniofacial defects in the region of the 1st and 2nd branchial arches were observed in the treatment group but not in controls for those animals delivered at term. Furthermore, premature accutane (13-*cis*-retinoic acid) administration can destroy neural

crest cells thereby interfering with their normal migratory patterns, and results in characteristic HFM deformities. In general, there is a lack of understanding surrounding the ramifications of embryonic neural crest cell destruction on structure development and long-term growth potential of the affected derivatives of the first and second branchial arches.

It appears that HFM could result from a variety of causal factors including defective genes, vascular anomalies and/or teratogenic materials, which either singularly or collectively incur disruption of the developing (pre-osteogenesis) facial skeleton (Kearns, Padwa, Mulliken & Kaban, 2000; Cousley & Calvert, 1997; Cousley & Wilson, 1992). Johnston and Bronsky (1991) argued that although HFM is considered to be at least partly caused by mutations in patterning genes (Rollnick & Kaye, 1983), the craniofacial and cardiovascular features of this complex syndrome suggest additionally primary crest cell involvement. A deficiency in the development of the branchial arch and head mesoderm is assumed to result from failure of the neural crest cells to migrate and contribute adequately to the facial primordia. This would cause alteration of standard morphogenetic interactions through matrix deficiency and consequently encourage malformations consistent with HFM (Jacobsson & Granstrom, 1996). Failure of neural crest cells to migrate and/or proliferate could be numerous and involve genetic, vascular and teratogenic changes.

2.10 Genetics

Despite the complex nature of HFM, evidence suggests a major genetic determinant in some cases of this condition. Rollnick and Kaye

(1983) revealed that 45% of 97 participants had a family history of some features of the disorder. Rollnick and Kaye (1983) further identified first-degree relatives as being the most frequently affected, often with mild phenotypic expression. "Formes fruste" (microforms include those displaying low penetrance and mild phenotype) are represented by auricular abnormalities (Converse, Wood-Smith, McCarthy, Coccaro & Becker, 1974).

Numerous theories have been proposed to explain the genesis of first and second branchial arch syndromes. Since few congenital disorders follow strict Mendelian segregation, a selection of explanatory genetic models has been developed. Those commonly applied to HFM include the multifactorial, polygenic and single gene (autosomal dominant and recessive) models (Cousley & Calvert, 1997; Cousley & Wilson, 1992; Connor & Fernandez, 1984).

Most affected individuals are cytogenetically normal, however a number of chromosome abnormalities have been reported including trisomy 18 (Greenberg *et al.*, 1988), deletion chromosome 5p (Neu, Friedman & Howard-Peebles, 1982), short arm chromosome 4 deletion (Zellweger, Bardach, Bordwell & Williams, 1975), terminal deletion (22q) Hathout *et al.*, 1998; Herman *et al.*, 1988), monosomy chromosome 6q (Greenberg *et al.*, 1988), chromosome 8q deletion (Townes & White, 1978), ring chromosome 21 (Greenberg *et al.*, 1988), recombinant chromosome 18 (Sujansky & Smith, 1981), trisomy 22 (Kobrynski *et al.*, 1993), chromosome 22q deletion (Greenberg *et al.*, 1988), *cri du chat* (5p-) syndrome (Neu, Friedman, Howard-Peebles, 1982) and 47 XXY (Klinefelter syndrome) (Garavelli *et al.*,

1999; Poonawalla, Kaye, Resonethal & Pruzansky, 1980; Kushnick & Colondrillo, 1975). Additionally chromosomal mosaicism has been reported including trisomy 7 (Hodes *et al.*, 1981) and 9 (Wilson & Barr, 1983). This offers an explanation for localised features and low recurrence risk. A number of instances of monozygotic twin discordance exist which complicates interpretation of twin studies. However, rare cases of concordance with varying expression have been reported (Burck, 1983, Ryan *et al.,* 1988). These studies all confirm genetic association for this disease and thus support the hypothesis that the West Australian family contain a genetic predisposition for this disease.

A lack of extensive kinships with many affected individuals has hampered previous genetic studies. Preliminary studies by Rollnick and Kaye (1983) involved sporadic cases and families with small numbers of affected individuals. While their results suggested a multifactorial origin for HFM. It was agreed that HFM is genetically transmitted (Kaye *et al.*, 1992).

Subsequent investigations performed by Kaye *et al.* (1992) involving segregation analysis provided evidence for autosomal dominant inheritance with incomplete penetrance. Recently, two HFM families have demonstrated autosomal dominant (AD) segregation, one linked to chromosome 14 (a London family) (Kelberman *et al.*, 2001) and the other to chromosome 11 (a West Australian family) (Chandler, 2001), the latter being the family under investigation. Moreover, the WA family under investigation appears to be indeed demonstrating anticipation with increasing severity of symptoms and affected individuals in subsequent generations which supports the

presumption that this form of HFM is genetically linked, specifically to chromosome 11.

2.11 Haplotypes of the West Australian Family

This West Australian family of 20 members contains 8 unaffected and 9 affected members. This family is one of a very few segregating with autosomal dominant hemifacial microsomia that is large enough to allow genetic linkage analysis to locate the disease gene. Linkage analysis performed by Dr David Chandler (2001) using 152 polymorphic markers suggested linkage of this family's hemifacial microsomia disease gene to a region of chromosome 11, between the markers D11S987 and D11S911. D11S911 was located on chromosome 11, position 75,971,191 - 75,971,382, as located using the Ensembl database. While D11S916 had been reported to be located centromerically on chromosome 11 (Chandler, 2001), this marker could not be located by this author at the time of experimentation. A comprehensive list of relevant, available markers and their respective positions on chromosome 11, deduced from Ensemble can be seen in Appendix 5.

Haplotypes for chromosome 11 are available from D11S1765 to D11S876 for both affected and unaffected individuals within this family (Chandler, 2001) (Figure 2, p11). The two paternal haplotypes were equally shared and similar maternal haplotypes were received by all four affected individuals in the second generation (i.e. minimal recombination). The maternal haplotypes received by affected individuals II:7 (D11S1883) and II:9 (D11S4166) demonstrates single recombination (Chandler, 2001). Aside

from a double recombination that occurred in individual III:6, Chandler (2001) observed no alterations to the grand-maternal haplotypes in third generation affected individuals. This author determined that since the marker D11S911 is non-informative in individual III:6, and marker D11S916 could not be located during experimentation the lower breakpoint is most probably at marker D11S1396 thereby maximising the distance (9.2cM) over which such an infrequent happening could occur (probability of 0.0084) (Chandler, 2001). From the haplotypes it is apparent that the linkage region for HFM stretches from D11S1883 to D11S911, approximately 16cM in length (Figure 2, p 10). It is only in one of the 'unaffected' children (III:1) that this (D11S1883-D11S911) haplotype occurs, although the lower part of the haplotype (D11S4207 to D11S911, 4.6cM) occurs in an unaffected sibling, III:2. This could suggest a single recombination at D11S1314. Of the three unaffected individuals (III:1, III:2 & III:5), none inherited the conserved haplotypes. Given that the unaffected individual III:1 possesses the non-recombinant grand-maternal haplotype, it seems likely that they are a non-penetrant carrier of the disease gene. Thereby, if individual III:2 is indeed unaffected then the critical region for HFM is reduced to between D11S1883 and D11S4207, 11.5cM. Unfortunately, at the time of experimentation marker D11S4207 could not be located on Ensembl and so genes within the region D11S1883 to D11S911 (16cM) were mined and several genes were identified as potential candidates for HFM in individual III:6.

2.11.1 Screening for Detected Polymorphisms / Mutation Detection

The final step of all human disease gene cloning strategies is the identification of the mutations on chromosomal DNA within the identified linkage regions, contributing to each disease. Mutations can be translocations, deletions, duplications, repeat expansions, insertions, splice site, mis-sense, frameshift or stop codon mutations. They can affect transcribed and non-transcribed portions of the genome.

Deletions that include markers adjacent to the disease gene can be recognised from non-Mendelian and/or homozygosity marker segregation in pedigrees for both autosomal dominant and recessive diseases. PCR methods detect smaller mutations in cloned DNA lying inside or outside coding regions. Commonly used mutation detection methods are listed in Table 2.3, p60. Of these mutation detection methods the direct sequencing of transcripts is of primary relevance and is discussed further in Chapter 3.

2.12 Data Mining

This project was innovative in that it adopted a novel strategy, genome mining with which four hemifacial microsomia candidate genes were identified in this project. This process involves cross-tabulating relevant details of the genes and predicted genes in the two known hemifacial microsomia candidate regions, then cross-tabulating all this information with what is available in vertebrate and invertebrate genome databases.

Bergeron (2003) defined data mining as the "process of automatically extracting meaningful patterns from (usually) very large quantities of

seemingly unrelated data" (p263). Data mining is not simply about searching for every possible relationship in a database, rather it has the ability to initiate queries that are not restricted by the user's fluency in authoring effective database inquiries. Data mining allows the researcher to utilize their pattern recognition skills and knowledge of the field of interest, to determine which data-mining results warrant further investigation. For example, with relation to this project, genes identified in the chromosome 14 linkage region would warrant further investigation if they were expressed embryonically, were associated with axial asymmetry or if they contained trinucleotide repeats. Their homology to the chromosome 11 linkage region would then be investigated. If a gene was found in the chromosome 11 linkage region that was homologous with a gene involved in axial asymmetry from the chromosome 14 linkage region, it would be considered a strong candidate for chromosome 11 HFM.

Data mining principally involves the identification of patterns and relationships in data that is often obscure in complex data sets. Pattern recognition is a major part of data mining as is, by extension, pattern discovery. Pertaining to bioinformatics, pattern recognition is primarily concerned with character sequence classification representative of nucleotide bases or molecular structures. For bioinformatics to be a beneficial research tool, the sequence and structure of proteins and other molecules must be linked to functional genomics. The online bibliographic data base PubMed, is a primary resource of this as it stores functional data that links clinical medicine, sequence and structure data. It is expected that by mining such databases, relationships between structure and

function will be revealed. PubMed was used extensively throughout the data mining experimental portion of this project.

Additionally, numerous databases were searched for research articles involving HFM were identified in databases. These articles were consequently researched for any indication of possible candidate genes or their tissue expression, which would highlight genes as potential candidates, particularly when homologous to different organisms. Alternatively, the candidate genes identified by data mining were looked up in PubMed for any indication of their involvement with different diseases.

Table 2.3 Common PCR Mutation Detection Methods

Method	References
*Direct Sequencing of Gene/Chromosome	Sanger 1981.
Denaturing Gradient Gel Electrophoresis	Fischer & Lerman, 1983.
RNase Cleavage	Myers *et al.*, 1985.
Deletion screening ('multiplex PCR')	Chamberlain *et al.*, 1988.
Chemical Mismatch Cleavage	Cotton *et al.*, 1988.
DNA Single-strand Conformation Polymorphism (SSCP)	Orita *et al.*, 1989.
RNA SSCP (rSSCP)	Sarkar *et al.*, 1992a.
Dideoxy Fingerprinting	Sarkar *et al.*, 1992b.
Repeat Expansion Detection (RED)	Schalling *et al.*, 1993.
Restriction Endonuclease Fingerprinting (REF)	Liu & Sommer, 1995.
Heteroduplex Cleavage with Bacterial Resolvases	Youil *et al.*, 1995.

* Relevant to this thesis and reviewed on page 57

2.12.1 Candidate genes: selection strategies

Aforementioned, the neural crest is a principle cell population that participates in craniofacial development (Jones & Trainor, 2004).

Additionally neural crest cells constitute the connective tissue-forming mesenchyme of the facial skeleton. It is for this reason, as mentioned previously, that genes associated with defects in patterning, particularly of neural crest cells, were targeted as candidate genes associated with craniofacial abnormalities (Jones & Trainor, 2004; Graham & Smith, 2001).

Key patterning factors include bone morphogenetic proteins (BMPs), retinoic acid (RA), fibroblast growth factors (FGFs), WNT and Hedgehog proteins (Melton, Iulianella & Trainor, 2004). Of these, *WNT11,* of the *WNT* gene family of patterning genes was identified as a candidate gene for HFM in the WA family by gene mining strategies (Chandler, Kaledijiva, personal communication, 2002). Although sequencing ruled out *WNT11* as a candidate, the identification of *WNT11* as a possible candidate exemplifies genome mining as a technique suitable for candidate gene identification and was considered a feasible method for use in this project.

Transcription factors regulate cell development by governing gene expression, differentiation and growth (Lacombe, 1999). Transcription factors, 'patterning genes', are the most likely to be involved in dysmorphism and are hence the most likely candidate genes (Lacombe, 1999). Of the patterning genes, *HOX* genes are of interest in the molecular dysmorphology field due to their association with dysmorphic *Drosophila* and mouse phenotypes.

Disrupted *HOX* genes in transgenic mice were found to generate dysmorphic phenotypes (Lufkin *et al.*, 1992; Chisaka & Capecchi, 1991;

Balling, Mutter, Gruss & Kessel, 1989). The generated phenotypes of the mice were similar to recognized human HFM malformations. Consequently, *HOX* genes were thought to be disrupted in human malformations (Chisaka & Capecchi, 1991).

There are 4 *Hox* gene clusters: HOXA (formerly HOX1) on chromosome 7, HOXB (formerly HOX2) on chromosome 17, HOXC (formerly HOX3) on chromosome 12, and HOXD (formerly HOX4) on chromosome 2. No *Hox* gene clusters are located on human chromosome 11, the region associated with the WA HFM. Therefore *Hox* genes were ruled out as candidate genes in this family.

Within the linkage region on chromosome 11, five genes were identified as suitable potential candidate genes for the form of HFM segregating in the West Australian family under investigation; *UVRAG, WNT11, GARP, CLNS1A* and *ARIX*. *UVRAG* (UV Radiation resistance Associated Gene - 11q13.5) is expressed embryonically and has been specifically associated with axial asymmetry (LocusLink 7405). Aforementioned, *WNT11* (11q13.5) is a patterning gene from the WNT family critical to the establishment of the D/V and A/P axes (Saint-Jeannet *et al.*, 1997). Saint-Jeannet *et al.* (1997) reasoned that the dorsally expressed *Wnt* molecules could play an important role in the regulation of neural crest cell formation. However, sequencing ruled out *WNT11* as a candidate (Chandler, Kaledijiva, personal communication, 2002). *ARIX* (Aristaless homeobox - 11q13.2) has been associated with left-right asymmetry determination and craniofacial abnormalities (LocusLink 401).

Specifically *ARIX* is expressed in noradrenergic cell types of the sympathetic nervous system, brain and adrenal medulla (Johnson, Smith, Johnson, Rhodes, Rinchik, Thayer & Lewis, 1996). Johnson *et al.* (1996) determined *ARIX* to be homologous to a gene on mouse chromosome 7 (approximately 50cM distal to the centromere of mouse chromosome 7) and defined regions of conserved synteny between mouse and human genomes Rinchik, Magnuson, Holdener-Kenny, Kelsey, Bianchi, Conti, Chartier, Brown, Brown, & Peters, 1992).

When investigating map locations of both mouse and human *ARIX* genes, Johnson *et al.* (1996) described it as highly likely that *ARIX* pertained to "inherited developmental disorders linked to human 11q13" such as HFM (p527).

CLNS1A (Chloride channel, nucleotide-sensitive, 1A - 13q13.5-q14) is involved in auxiliary transport protein activity, chloride transport, circulation, regulation of cell volume and visual perception and has been linked to human diseases such as leukemia (Buyse, De Greef, Raeymaekers, Droogmans, Nilius & Eggermont, 1996) and human breast carcinoma (Bekri and colleagues, 1997) (LocusLink 1207). With regards to auxiliary transport protein activity, *CLNS1A* facilitates the transport across one or more biological membranes without participating directing in transport itself. Bekri et al. (1997) through experimentation identified *CLNS1A*, *GARP* and *UVRAG* as genes present within the region 11q13.5-q14.1, which they linked to estrogen receptor positive breast carcinomas prone to metastasis. Thus

these genes have previously been identified as playing a crucial role in human disease associated with aberrant cell proliferation and migration.

The functions of *GARP* (Garpin complex) include receptor interactions, enzyme inhibition, cell adhesion and cellular trafficking. Pruitt, Katz, Sicotte and Maglott (2000) revealed its involvement in early mammalian development, neural development, cell polarization, regulation of gene expression and apoptosis signalling. Furthermore *GARP* plays a critical role in the morphology and dynamics of the cytoskeletal framework. Pruitt *et al.* (2000) also identified *GARP* as having homology with the *Drosophila Aristaless* gene. Similar to the human gene *ARIX* (identified as a candidate gene in this project) the *Drosophila Aristaless* gene is specifically expressed in noradrenergic cell types of the sympathetic nervous system, of the brain and adrenal medulla of *Drosophila*.

2.13 Animal Models

Significant insight into craniofacial development has been recently unveiled through generation of transgenic mouse mutants. These mice are engineered with mutations in genes responsible for craniofacial development and the resultant craniofacial syndromes are assessed and compared by classical comparative embryological approaches in a range of species (Jones & Trainor, 2004). These studies (Jones & Trainor, 2004), have emphasised the importance of developmental transcription factors (patterning genes) in craniofacial development.

Moreover, evidence from two mouse models supports genetic involvement in HFM. Hemifacial maxillary malformation in heterozygotes with incomplete penetrance has been observed to result from alteration in the genetic background of the mouse *far (first arch)*, a recessive lethal mouse mutant (Juriloff *et al.*, 1987).

This mutation was first discovered as an autosomal recessive disorder that caused bilateral facial deficiencies in homozygotes. Transfer of the mutation to a different strain produced hemifacial deficiencies in heterozygotes. Juriloff, Harris and Froster-Iskenius (1987) proposed that bilateral and unilateral abnormalities of tissue derived from the first branchial arch may be due to a defect in a single gene and suggested that the first branchial arch syndromes in mice, which include tissue derangements and deficiencies in the maxillary prominence-derived tissues may be analogous to those in HFM of the human.

The second mouse model is an autosomal dominant insertional mutation on mouse chromosome 10 which produces a phenotype in hemizygous mice which resembles HFM in humans, including microtia, low set ears and an abnormal bite resulting from hypoplasia of the second branchial arch accompanied by local haemorrhage (Naora *et al.*, 1994). The mutation was caused by the insertion of multiple, tandem copies of a myelin basic protein during the creation of transgenic mice. Of 21 transgenic mice with the same transgene randomly inserted, only one displayed facial deformities suggesting that the site of the transgene insertion was responsible for the phenotype, rather than the transgene itself (Otani, Manzoku, Shibaski &

Nomachi, 1978). *In situ* hybridization, performed on transgenic mice by Naora *et al.* (1995), identified chromosome 10 as the locus of integration and therefore of association with the *Hfm* phenotype in mice. At the time of this study, no naturally occurring mutations were mapped to this locus in mice. Further investigation of this *Hfm* model revealed that integration of the transgene was accompanied by a 23kb deletion of genomic DNA. Evolutionary conserved sequences were identified within and beside the deleted region, suggesting that they may correspond to either a gene(s) or a regulatory structure(s) that is associated with the HFM phenotype. No mouse without the transgene but containing the 23kb deletion was described. The absence of homozygous mice in the offspring of hemizygous males and females suggests that the homozygous is lethal during the prenatal period.

Naora *et al.* (1994) noted that in terms of pathogenesis, the *far* and *Hfm* mutations were at different loci and supported different aetiologies of human HFM, those of vascular abnormality (*Hfm*) (Cousley & Wilson, 1992) and those pertaining to cartilage perturbation (*far*) (Juriloff *et al.*, 1987). The region of mouse chromosome 10 has syntony to a portion of human chromosome 6q24; there are no known genes in this region that are obvious candidates for HFM. Although there is no known region of human synteny for the *far* mouse mutant, this homologous region cannot be excluded as containing candidate genes for HFM in the human (Chandler, 2001).

Further potential candidate loci have been derived through experimentation in the mouse. Inactivation of *FGF8*, located on human chromosome 10q25, in the first branchial arch resulted in a failure to develop cartilaginous and skeletal first branchial arch derived structures (Trumpp, Depew, Rubenstein, Bishop & Martin, 1999). Mice with heterozygous disruption of *FGF8* in the first branchial arch, demonstrated a less severe phenotype observed on only one side of the head (Trumpp *et al.*, 1999).

Another mouse mutation that deserves mention is the *Msx1* knockout mouse, *Msx1* being a homeobox gene expressed in numerous tissues in the developing foetus including the limb buds and branchial arches. Satokata and Mass (1994) identified the expression of genes of the *Msx* class in the epithelial or mesenchymal components of disparate tissues undergoing morphogenesis and observed bilateral facial malformation. *Msx1* appears to be inherited in a recessive fashion. Satokata and Mass (1994) further discovered that the phenotype in *Msx* knockout mice was mainly restricted to the first pharyngeal arch, despite strong evidence for a role for *Msx1* in limb pattern formation. Juriloff, Harris & Froster-Iskenius (1987) proposed that this may be modified by genetic background, although transfer of the mutation to different strains of mice to test this proposition has not yet been done. Despite none of the five candidate genes displaying homology to known mouse mutants this process assisted through eliminating mouse homology as a technique for candidate gene selection in the genome mining process.

Table 2.5 Chromosomal abnormalities responsible for HFM

Chromosome abnormality	Reports	Candidate	Reference
inv1	1		236
del5p	3		237, 238
Monosomy 6q	1		239
del6q	1		239
tri7 mosaicism	1	HOXD	235
Duplication7q	1		240
del8q	1		241
tri9 mosaicism	1		242
tri 18 mosaicism	1		243
Tri 18	2		239, 244
Recombinant 18	1		245
del18q	1		246
del 22q	2		239,247
49 XXXXY	1		248
47 XXY	1		249
49 XXXXX	1		250
47 XXX	1		251

Adapted from Chandler (2001)

2.14 Bioinformatics

As a result of the Human Genome Project, there are a number of freely available tools which can be implemented to identify candidate genes: advanced linkage analysis software, automated genotyping protocols, comprehensive genomic and cDNA libraries in a variety of vectors, and

public access to genetic databases. Combinations of these tools were used to 'data mine' the genes present in the chromosome 11 linkage area.

2.15 Candidate genes; Selection by association studies

Taking into account the complexity of HFM classification, it would be unwise to assume that the syndrome is caused by a defect in a single gene. It is more likely, as pointed out by Mulvihill (1995) that differing mutations in multiple but related genes could produce the range of phenotypes displayed in HFM. As yet, the only region linked by family studies with West Australian families segregating HFM are the chromosome 11 and chromosome 14 studies. Alternate chromosomal abnormalities responsible for HFM are listed in Table 2.5, p70. Although the report by Hodes *et al.* (1981) of a HFM patient with a trisomy seven suggests linkage of the disorder to chromosome 7 is not localised enough to define a candidate gene. The only other indications of candidate genes come from the animal models described above.

Primarily, through bioinformatics this project researched the chromosome 14 linkage region for any indications of homology to the chromosome 11 linkage region. Using data mining techniques (2.14), a gene list from markers D14S1142 to D14S267 (Kelberman *et al.*, 2001) was created for chromosome 14 (Appendix 16). Each gene in this linkage region was researched and documented under headings including alternative symbols, alias names, expression, tissue specificity, family from which the gene was derived, function, locus link reference number, position in the linkage region and homologs (Appendix 6 & 8). This linkage region on

chromosome 14 identified by Kelberman *et al.* (2001) harbours the *Goosecoid* gene, a candidate for the chromosome 14 linked HFM based on mouse expression and phenotypic data.

2.16 Conclusion

Hemifacial microsomia is a common birth defect involving first and second branchial arch derivatives. It is a complex deformity that displays a spectrum of anomalies involving soft-tissue and skeletal structures. Additionally it can involve deformation of the cardiac, pulmonary, gastrointestinal, renal and central nervous systems. The majority of reported cases are considered sporadic in origin, however substantial evidence supporting genetic involvement has evolved.

Substantial debate has surrounded the cause and pathogenesis of HFM. It appears HFM results from a culmination of causal factors, including defective genes, teratogenic substances or vascular anomalies. Theories attempting to explain the pathogenesis of HFM are numerous and range from neural crest cell involvement to a deficiency in the development of the branchial arch and head mesoderm.

As any structure deriving from the first or second branchial arch may be affected, the clinical picture can vary considerably. Of significance is the universal demonstration of axial asymmetry. Predominant features of HFM include progressive underdevelopment of the mandible, zygoma, external and middle ear.

Accurate assessment of facial anomalies is essential for effective diagnosis, classification, treatment and management plans. The most recent systems of classification pertaining to diagnosis include the Skeletal, Auricle and Soft-tissue (SAT) multisystem classification and the Orbital distortion, Mandibular hypoplasia, Ear anomaly, Nerve involvement and Soft tissue deficiency (OMENS) classification system. Alternative systems, including the Pruzansky classification system, Teconi and Hall's phenotypic classification system and Munro and Lauritzen's system were alternative classification systems devised to guide treatment planning and diagnosis of HFM, although not as widely accepted as the SAT and OMENS classification system. Involvement of facial soft tissue can be mild or severe and needs to be thoroughly assessed prior to surgical intervention and treatment strategies.

The term *hemifacial microsomia* was initially coined by Gorlin, Jue, Jacobson and Goldschmidt in 1963 to refer to patients displaying unilateral microtia, macrostomia and hypoplasia of the mandibular ramus and condyle (Politi, Sembronio, Robiony & Costa, 2002). The variable expressivity of an autosomal dominant gene may explain the variability in the phenotype of HFM in two reported HFM pedigrees. No genes had as yet been identified for the WA family, although several likely candidate genes were selected, as a result of careful selection criteria.

Despite analysis of genes in the chromosome 11 linkage region, known mouse mutants, and genes identified in the chromosome 14 linkage region no

homologies were discovered and thus this project implemented PCR and sequencing to assess candidate genes.

2.17 Aims

1. To identify HFM candidate genes in one WA family clearly segregating with HFM

2. To find mutations in these genes on chromosome 11 as would cause HFM

To accomplish these aims, a combination of sophisticated genetic techniques was utilised, including bioinformatics, mutation detection by PCR and sequencing, and data mining.

Chapter 3: Research Methods

3.0 Introduction

This chapter contains in-depth details of experimental protocols. It also includes details of software used to obtain and analyse data for this thesis. Protocols have been adapted from a number of sources, all of which are referenced within the text. It is through methodology detailed in this chapter that candidate genes were selected and sequenced for this project.

3.1 Online Databases

Table 3.1 - Web-based databases used to obtain and analyse data

Database	Web address (URL)
PubMed	www.ncbi.nlm.nih.gov/entrez/query.fcgi
Draft human sequence	www.genome.cse.ucsc.edu
ENSEMBL	www.ensembl.org
Online Mendelian Inheritance in Man (OMIM)	www.ncbi.nlm.gov/OMIM
GenBank	www.ncbi.nlm.nih.gov
The Genome Database (GDB)	www.gdb.org

To identify literature and data relevant to this project, searches were performed under (a combination of) the following headings 'hemifacial microsomia', 'malformation', 'congenital disorders', 'birth defects', 'axial asymmetry', 'trinucleotide repeat', 'Goldenhar syndrome', 'branchial arch derivatives', 'oculo-auriculo-vertebral spectrum', otomandibular dysostosis', 'lateral facial dysplasia', craniofacial microsomia' in a number of databases (Table 3.1, p70).

3.1.1 Draft Human Genome Sequence and ENSEMBL

The Draft Human Genome Sequence and ENSEMBL are two different interfaces that have been developed for accessing the draft human sequence. Both databases were initially released in June 2000 and

have since been regularly upgraded. Each interface is searchable by marker, gene name, clone name and/or base pair number along a chromosome. Moreover, both sites incorporate a variety of features including positioning of linkage markers, gaps, known and predicted genes and the position and types of repeats.

The physical size of the chromosome 11 region in linkage with hemifacial microsomia in the West Australian family was estimated from the sequence of the Human Genome Working Draft available at UCSC genome website (http://www.genome.ucsc.edu/) on April 2001. This was achieved by mapping out the positions of the markers flanking the chromosome 11 HFM candidate region, namely D11S1883 and D11S911 as the gene responsible for chromosome 11 HFM is expected to lie within this region. Markers refer to polymorphisms such as microsatellite repeats that are linked to a disease locus (Jorde *et al.*, 2000). This estimation of the linkage region was updated in September 2001, December 2001, April 2002 and September 2002, during the initial stages of this thesis (See Haplotypes, Figure 2).

The UCSC (University of California, Santa Cruz) database was used to search for these same flanking markers, namely D11S1883 and D11S911, to obtain a complete map of the region and associated genes (http://www.genome.cse.ucsc.edu). By clicking on the linkage region a complete list of genes within this region was displayed.

3.1.2 Online Mendelian inheritance in man (OMIM) (http://www.omim.com/)

OMIM is a database containing a definitive collection of human genes and associated genetic disorders. Cytogenetic maps of diseases (morbid map) and genes (gene map) are available from this site.

Figure 5 Chromosome 11

Linkage region identified in red

Adapted from data presented in USCS (http://www.genome.ucsc.edu)

A list of genes within this region is depicted in Appendix 9

From these candidate genes were selected by data mining as described in Section 3.4

In particular this database was used to identify the markers flanking the region of interest on chromosome 11, locate each of the candidate genes on chromosome 11, and identify their nucleotide position in relation to other genes present within the region. Furthermore OMIM was used to obtain information on each of the genes selected as candidates for chromosome 11 HFM. The OMIM databases provided links to the NCBI interface which within itself offered links to over 11 databases (Table 3.2, p75). Each link unraveled new databases associated with the gene in question and thereby generated a cascade of information on each gene within the chromosome 11 linkage region. Based on information retrieved from these databases, four candidate genes were selected with the aid of strict selection criteria given in Section

3.6. Relevant information obtained through bioinformatics investigations is displayed in Appendix 15.

3.1.3 GenBank (http://www.ncbi.nlm.nih.gov)

GenBank at the National Institutes of Health, USA, contains DNA sequences from all species tested. This database was used by the author of this project to obtain the sequences of known genes located in the linkage regions for hemifacial microsomia. GenBank searches for candidate gene sequences were performed by BLAST using a GenBank accession number or, if this number was not known, a partial DNA sequence of the chromosome 11 HFM linkage region, obtained from the Draft Human Sequence, was used. GenBank was also used to confirm primer design for PCR and sequencing reactions and as the source of 'normal' sequence for comparative analysis with sequence generated from the DNA of the patient affected with HFM.

Aforementioned, the NCBI database comprises a number of databases, the majority of which were used. These are alphabetically listed and briefly described in Table 3.2 (p75). More specifically, the database 'Books' was used for general information on a gene, primarily expression and/or association with a disease. 'Genome' was used extensively to map the genes within the chromosome and obtain genomic sequences. OMIM was specifically used for information on genes and associated pathologies, as was PubMed, which was further used to identify syntony to other genes within the human genome.

3.2 Online Analysis Software

3.2.1 Basic local alignment sequence tool (BLAST)

(http://www.ncbi.nlm.nih.gov/BLAST)

The BLAST program enables a given DNA sequence to be compared with all sequences in GenBank in both forward and reverse orientation. BLAST uses algorithms to identify optimal sequence alignments, and search results are typically displayed as a series of statistical analyses of the comparisons (Fassler, Nadel, Richardson, McEntyre, Schuler, McGinnis, Pongor & Landsman, 2000). The database provides a list of sequences as pair wise comparisons between the test sequence and each related identified sequence (Fassler *et al.*, 2000). BLAST was used to compare sequences generated from PCR products of candidate genes in patient DNA against the wild type sequence of that gene in GenBank.

Mismatches indicate a possible mutation. Should a difference between the known published 'normal' unaffected sequence and the HFM affected sequence be noted, further investigation would ensue to determine the specific mutation from which HFM could derive.

Table 3.2 Databases included in NCBI Entrez

Database	Description
3D domains	Protein domains from NCBI's Conserved Domain Database
Books	A biomedical book collection
Genome	Views of genomes, chromosomes, sequence maps, and integrated physical and genetic maps
Nucleotide	Nucleotide sequence data from GenBank, the European Molecular Biology Laboratory (EMBL) and the DNA database of Japan (DDBJ), the Genome Sequence Database (GSDB) and patent sequences from US Patent and Trademark Office (USPTO) as well as other international patent offices
OMIM	Human genes and genetic disorders
PopSet	Nucleotide and protein sequence data that has been aligned and submitted as a set resulting from a population, phylogenetic or mutation studies
Probeset	Hybridization array and Gene Expression Omnibus (GEO)
Protein	Protein Information Resource (PIR) (vertebrate protein sequences) SWISS-PROT, Protein Research Foundation (PRF) and Protein Databank (PDB), and from the coding sequences that have been translated from GenBank, EMBL and DDBJ DNA sequences
PubMed	Biomedical literature
Structure	Protein Databank (PDB) derived experimental data from crystallographic and NMR structure determinations
Taxonomy	Names of all the organisms represented in NCBI's genetic database

Adapted from Bergeron (2003)

3.3 Sequence Maps

Similar to the use of flow diagrams being used to portray content and describe the organization of various components within a program, gene maps provide a sophisticated view of a gene and the location of its nucleotide sequence within a chromosome (Bergeron, 2003). The Web-based Map Viewer is a fundamental application used to map genes within the genome giving its chromosome location (Figure 5). The Map Viewer is part of NCBI's integrated Entrez system and provides a composite interface with access to various online databases. Through these programs, homologous genes can be located in the genomes of different species. Furthermore it visually demonstrates the distance between

genes in a chromosomal region and information on the sequence of a gene in the critical region of a chromosome. Graphs are typically used to map the collated sequences from the sequential databases of NCBI and relevant links are supplied corresponding to associated databases that relate to specific diseases and sequences of the genes themselves (Fassler *et al.*, 2000). The gene list obtained from the Map Viewer program (Appendix 11) was a starting point for the bioinformatics methods used to identify candidate genes present in the chromosome 11 linkage region as it is from this database that the original list of genes was derived [http://www.ncbi.nlm.nih.gov/genome/guide/human/] (Appendix 9). Each specific gene was then researched for criteria relating to candidate gene selection (Appendix 7 & 12).

The Map Viewer was monitored regularly to check the positions of the flanking markers and the genes within the linkage region. As more of the Draft Human Genome was completed markers within and flanking the linkage region were not always mapped at previous locations. It therefore became essential to continually monitor the marker positions to ensure the identified candidate genes remained within the linkage region - Map Viewer was used for this purpose. This is explained further in 4.1.

3.4 Data Mining

This project adopted a strategy known as 'genome mining' (Houle, Cadigan, Henry, Pinnamaneni & Lundahl, 2001) to identify hemifacial microsomia candidate genes. This strategy involved obtaining all relevant data on the genes from available databases, cross-tabulating relevant details of the genes and predicted genes in the two known hemifacial

microsomia candidate regions on chromosome 11 and 14, and then cross-tabulating relevant information with available information on vertebrate and invertebrate genomes from appropriate databases. Information obtained from databases for each gene included: names, aliases, acronyms, functions, disease associations (particularly with axial asymmetry), and association with trinucleotide repeat expansions. Appropriate analytic methodology incorporating data mining, continual investigation of the linkage region using numerous databases and thorough evaluation of all the genes present in the chromosome 11 linkage region enabled identification of candidate genes.

The size and complexity of current genomic databases assisted with data mining as it was possible to acquire sufficient information on gene sequence, structure and function. This was particularly relevant to this project, as novel strategies such as data mining have not only become possible through the introduction and expansion of various web databases but have become an eminent research tool, as displayed in this project.

3.5 Data Collection / Storage

3.5.1 Pedigree data

Pedigree data includes previously collected clinical and personal data for patients affected with chromosome 11 HFM. Clinical ascertainment was conducted by Professor Jack Goldblatt from King Edward Memorial Hospital (KEMH). Information provided by the family was coded to ensure confidentiality. DNA of one affected patient was provided by Professor Goldblatt, Royal Perth Hospital (RPH) for use in this project. Prior to the collection of DNA, the patient was required to sign a

consent form, a copy of which is given in Appendix 13. This DNA sample was coded with a unique identifier using the DNA Daybook system employed by the Department of Neuropathology, RPH. This code was annoted to the HFM pedigree under investigation. The only information provided for the genetic investigations reported in this thesis was an annoted pedigree with HFM status indicated. Patient (III:6, DOB 05.01.1981) is affected with HFM as previously identified by Singer *et al.* (1994). This patient demonstrated microtic right ear with ipsilateral macrostomia. On a previous occasion, this was surgically repaired postnatally with simultaneous excision of a right pre-auricular skin tag. This patient had a right deviated chin point and upper right occlousal plane. Soft tissue was hypoplastic and overlay both the right infraorbital and ramus region. Singer *et al.* (1994) describes that upon oral examination this patient was found to have a "Class 1 malocclusion in the mixed dentition characterized by the displacement of the upper and lower centre lines to the right" (p287). On opening, the mandible deviated to the right despite no history of temporomandibular joint dysfunction. No teratogenic exposure or parental consanguinity was identified (Singer *et al.,* 1994). This patient is an excellent sample for DNA mutation analysis because the patient is in the III generation and displays a moderately severe phenotype of HFM.

3.6 Identification of Candidate Genes

Mutations in different genes, perhaps numerous genes, are likely to cause dysmorphology. As mentioned in the introduction, prime candidate genes from the chromosome 11 HFM linkage region were identified by adopting specific selection criteria, described below, followed by

comparison with mouse dysmorphology genes and chromosome HFM 14 genes.

Firstly, each and every gene located in the chromosome 11 HFM linkage area was inspected for three selection criteria; (1) candidate genes should be expressed embryonically as HFM is a congenital disorder. (2) Candidate genes would be expected to be associated with axial asymmetry, a definitive and unique feature of hemifacial microsomia and (3) genes should be associated with trinucleotide repeats. Trinucleotide repeat expansions are important to research as they are a form of mutation associated with expansion in copy number of a sequence of three nucleotides, situated in or near a gene (Cummings, 2000). Trinucleotide repeats, alternatively termed microsatellites, are well documented as being unstable and have been strongly correlated with disease severity (Jorde, Carey, Bamshad & White, 2000). These repeats are particularly relevant to this study as the chromosome 11 form of HFM appears to be more severe with succeeding generations. If chromosome 11 HFM is indeed displaying anticipation, the appearance of more severe symptoms at earlier ages in succeeding generations would be expected (Cummings, 2000) and this was indeed observed (Table 1.1, p7); such anticipation could be associated with trinucleotide repeats (Brook, McCurrah, Harle *et al.*, 1992).

Several diseases have been documented in which the number of trinucleotide repeats increases with successive generations, accompanied by increasing severity of the disorder. This phenomenon is particularly evident in cases of myotonic dystrophy and Huntington Disease and Fragile X (Brook, McCurrah, Harle *et al.*, 1992). Typically, in disorders

representative of varying degrees of anticipation, any initial alteration in the number of copies of a repeat sequence, either within or near a gene, would increase the probability that further modification in repeat numbers will occur.

This in turn generates alleles with full mutations. For example, in regions of the genome containing trinucleotide repeats, this event increases the chance of and promotes further expansion and subsequent development of the disease phenotype (Cummings, 2000). Inheritance of such disorders might be explained through forms of genome instability as opposed to simple Mendelian inheritance (single gene mutation inheritance), as once assumed (Cummings, 2000; Jorde, Carey, Bamshad & White, 2000). Genome instability is an abnormal condition in which there is a substantial increase in mutations throughout the genome and includes the widespread mutations, chromosome breaks and aneuploidy (Jorde, Carey, Bamshad & White, 2000). Screening for these repeats would be an additional filtering process for candidate gene selection. To do this, the entire genomic sequence for a particular gene was copied onto a word document and then each possible trinucleotide combination *AAC*, *ACT*, *GCC*, *GCG* etc. within the entire chromosome 11 linkage region was identified using the 'find' function in Microsoft Word (Appendix 10).

Using the methods described above, eleven genes out of a total of 131 genes were identified as potential candidates. In order of likelihood of possessing the mutation responsible for HFM, they are *UVRAG*, *ARIX*, *WNT11*, *GARP*, *CLNS1A*, *PKRIR*, *CAPN5*, *LGALS12*, *DKFZP564M082* and *OMP*. *UVRAG* (UV Radiation resistance Associated Gene - 11q13.5) is expressed embryonically (Perelman *et al.*, 1997) (LocusLink 7405), is

associated with axial (a)symmetry and contains two trinucleotide repeats, *AAC* at 74,105,973-74,105,987 and at 74,272,958-74,272,978 (Ensembl gene ID ENSG00000137493).

ARIX (Aristaless homeobox - 11q13.2) has been associated with left-right asymmetry determination and craniofacial abnormalities (Pattyn, Morin, Cremer, Goridis & Brunet, 1997) (LocusLink 401) and is situated near two GTT trinucleotide repeats (75,904,244-75,904,267 and 75,934,321-75,934,344) (LocusID 116648). Aforementioned, *ARIX* is specifically expressed in noradrenergic cell types of the sympathetic nervous system, brain and adrenal medulla (Johnson, Smith, Johnson, Rhodes, Rinchik, Thayer & Lewis, 1996). Johnson *et al.* (1996) determined *ARIX* to be homologous to the mouse chromosome 7 (*Phox2*). Using backcross analytical techniques, Johnson *et al.* (1996) mapped ARIX to mouse chromosome 7, approximately 50 cM distal to the centromere, in a region showing conservation of synteny with human chromosome 11. This author's research also identified such regions of conserved synteny between mouse and human genomes and confirms the selection of *ARIX* as a candidate gene for "inherited developmental disorders linked to human 11q13" such as HFM (Johnson *et al.*, 1996, p,527).

WNT11 is a patterning gene from the *WNT* family critical to the establishment of the D/V and A/P axes. As previously discussed, this gene family is specifically expressed along the D/V axis of the developing neural tube (Saint-Jeannet *et al.*, 1997). This gene family consists of structurally related genes encoding secreted signalling proteins that have been implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning

during embryogenesis (LocusLink 7481). The dorsally expressed *Wnt* molecules play an important role in the regulation of neural crest cell formation (Saint-Jeannet *et al.*, 1997) and in anterior-posterior patterning of the limb. *WNT* is expressed both embryonically and fetally and is located near a *CAG* trinucleotide repeat (73,925,066-73,925,080) and near two GTT repeats (73,869,466-74,369,532 and 73,881,196 - 73,881,210) (Ensembl gene ID ENSG00000085741).

CLNS1A (Chloride channel, nucleotide-sensitive, 1A - 11q13.5-q14 is involved in auxiliary transport protein activity, chloride transport, circulation, regulation of cell volume and visual perception and has been linked to human diseases such as leukemia (Buyse *et al.*, 1996) and human breast carcinoma (Bekri *et al.*, 1997) (LocusLink 1207). With regards to auxiliary transport protein activity, *CLNS1A* facilitates the transport across one or more biological membranes without participating directing in transport itself. Bekri *et al.* (1997) identified *CLNS1A*, *GARP* and *UVRAG* as genes present within the region 11q13.5-q14.1, which is linked to estrogen receptor positive breast carcinomas prone to metastasis.

The functions of *GARP* (Garpin complex) includes receptor interactions, enzyme inhibition, cell adhesion and cellular trafficking. As previously discussed Pruitt, Katz, Sicotte and Maglott (2000) reveal its involvement in early mammalian development, neural development, cell polarization, regulation of gene expression and apoptosis signalling. Furthermore *GARP* may play a critical role in the morphology and dynamics of the cytoskeletal framework. This author also identifies *GARP* as having homology with the Drosophila *Aristaless* gene and to the human gene *ARIX*. The *ARIX* gene is specifically expressed in

noradrenergic cell types of the sympathetic nervous system, brain and adrenal medulla in mice and Drosophila (Pruitt *et al.*, 2000). With specific regards to the selection criteria primarily used to identify candidate genes in this project both *CLNS1A* and *GARP* are located near two trinucleotide repeats (GTT repeats near *GARP* at 73,512,866-73,512,889 and *CLNS1A* at 76,696,182-76,696,199) (Ensembl gene ID ENSG00000137507 and Ensembl gene ID ENSG00000074201 respectively).

PRKRIR (protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repressor of (P58 repressor) - 11q13.5) and *CAPN5* (Calpain 5 - 11q14) are both expressed embryonically (Gale *et al.*, 1998; Dear, Matena, Vingron & Boehm, 1997) and are located near trinucleotide repeat expansions (PKRIR - (*CGG*) 73,782,820-73,782,834; (*CAG*) 73,925,066-73,925,080; (*GCG*) 73,782,819-73,782,833 ; (*AAC*) 73,741,283-74,047,866 and 73,677,006-73,677,020 CAPN5 - (CTT) 73,125,943-73,125,963) (Ensembl gene ID ENSG00000137492 and Ensembl gene ID ENSG00000149260 respectively).

Similarly *LGALS12* (lectin, galactoside-binding, soluble, 12 (galectin 12) - 11q13) and OMP (olfactory marker protein - 11q13.5) are fetally expressed (Yang, Hsu, Yu, Ni, Liu, 2001; Evans *et al.*, 1993 respectively) (PMID 9302557, Ensembl gene ID ENSF00000009799 respectively). The OMP gene structure and protein sequence are highly conserved between mouse, rat and human (Ricnchik *et al.*, 1992). *DKFZP564M082* (DKFZP564M082 protein - 11q13.4) is homologous to mouse chromosome 7 and is phenotypically known to produce congenital abnormalities (LocusID: 25906).

Of these eleven genes, due to their compliance with the selection criteria and specific association with craniofacial deformation and axial development, three were identified as prime candidates, *UVRAG, WNT11* and *ARIX.* Chandler (2001) and Kaledijiva (2002) (personal communication) had previously selected the developmental gene *WNT11* as a candidate for HFM due to its involvement in embryonic axial patterning, and sequenced this gene in an unaffected and an affected individual from the West Australian chromosome 11 HFM pedigree. The *WNT11* sequences for the HFM affected individual and the unaffected individual did not vary significantly from each other or the published sequence, and as no mutations were found, this gene was eliminated as a candidate.

Due to the properties of *GARP*, namely its homology to Drosophila *ARIX* and its involvement in cytoskeletal framework and neural development, it was chosen as a possible candidate gene and was sequenced for mutations in the HFM patient, along with *UVRAG* and *ARIX*, as will *CLNS1A*.

Regions of the mouse genome homologous to the human chromosome 11 HFM linkage region were identified through web-searches and genes lying within homologous regions were inspected for the same three criteria; embryonic expression, association with trinucleotide repeats and association with axial asymmetry (Appendix 15).

Genes in the chromosome 11 and chromosome 14 HFM linkage regions were compared in regard to the same three criteria; embryonic expression, association with trinucleotide repeats and association with

axial asymmetry. Genes that satisfied these criteria and shared homology with both HFM linkage regions would become prime HFM candidates. This comparison was essential as the chromosome 14 region has been definitively linked to autosomal dominant HFM (Kelberman *et al.*) and could thus provide crucial information on genes likely to be responsible for chromosome 11 HFM.

3.7 Sequencing of Candidate Genes to Identify Mutations

The four candidate genes, *UVRAG, ARIX, CLNS1A* and *GARP* were analyzed for mutations by standard PCR methodology. Exonic sequences of each of three candidate genes in an affected patient were sequenced and compared with normal gene sequence, obtained from the NCBI database. Several attempts to sequence the fourth gene (*ARIX)* were unsuccessful.

3.7.1 *UVRAG* (Candidate gene 1)

The *UVRAG* gene is 328,418 bases in length, including both exonic and intronic sequence and spans base numbers 74,047,885 to 74,376,860 on chromosome 11q13.5 (as at April 2002). *UVRAG* contains 15 exons. The number of bases in each exon and their positions on chromosome 11 are depicted in Table 3.3 (p86), to clarify the nucleotide positions and regions of *UVRAG* to be sequenced.

Table 3.3 *UVRAG* exons and positions on chromosome 11

Exon	Position (bases)	No. bases
1	74,049,885-74,050,179	294
2	74,086,537-74,086,655	118
3	74,096,400-74,096,435	35
4	74,114,532-74,114,694	162
5	74,123,482-74,123,557	75
6	74,146,607-74,146,693	86
7	74,196,098-74,196,204	106
8	74,218,041-74,218,168	127
9	74,238,659-74,238,744	85
10	74,242,188-74,242,276	88
11	74,243,461-74,243,522	61
12	74,251,469-74,251,635	166
13	74,300,364-74,300,486	122
14	74,350,578-74,350,670	92
15 (i)	74,375,366 - 74 375 968	602
15 (ii)	74,375,842-74,376,425	583
15 (iii)	74,376,320-74,376,860	540
15 (iv)	74, 376,238-74,376,502	464

Given the enormous size of exon 15 (Appendix 14), and for purposes of experimentation, the exon was divided into four parts, and PCR primers were chosen to span both halves so that the entire exon could be analyzed by sequencing, exon 15(i) spanned 602 bases, exon 15(ii) spanned 583 bases, exon 15(iii) spanned 540 bases and exon 15(iv) spanned 464 bases.

3.7.2 ARIX (Candidate gene 2)

The ARIX gene is 4,426 bases in length, including both intronic and exonic sequence and spans from base numbers 75,883,451 to 75,887,841 on chromosome 11 (as at April 2002). The ARIX gene contains 3 exons. The number of bases in each exon and their positions on chromosome 11 are as depicted in Table 3.4 (p87), to indicate regions of the gene to be sequenced.

Table 3.4 ARIX exons and positions on chromosome 11

Exon	Span	No. bases
1	75,883,451-75,883,803	352
2	75,886,282-75,886,470	188
3	75,887,370-78,887,820	450

3.7.3 *GARP* (Candidate gene 3)

The *GARP* gene is 12,322 bases in length, including both intronic and exonic sequence and spans from base numbers 74,891,183-74,903,592 on chromosome 11 (as at April 2002). The *GARP* gene contains three exons. The third exon is 5,097 bases in length and was divided into eight parts for choice of PCR primer positions so that the entire exon could be analyzed by sequencing. The number of bases in each exon and their positions on chromosome 11 are depicted in Table 3.5 (p87), to indicate nucleotides that require sequencing for mutations in affected and unaffected samples.

Table 3.5 *GARP* exons and positions on chromosome 11

Exon		Span	No. bases
1		74,891,183- 74,891,371	188
2		74,894,952-74,895,264	88
3	(i)	74,899,273-74,899,971	699
	(ii)	74,899,814-74,900,391	578
	(iii)	74,900,281-74,900,911	630
	(iv)	74,900,591-74,901,191	600
	(v)	74,901,257-74,901,866	609
	(vi)	74,901,773-74,902,343	570
	(vii)	74,902,297-74,902,898	601
	(viii)	74,902,782-74,903,490	708

3.7.4 *CLNS1A* (Candidate gene 4)

The *CLNSIA* gene is 21,648 bases in length, including both intronic and exonic sequence and spans base numbers 75,889,882 to 75,937,186 on chromosome 11. The *CLNS1A* gene contains 7 exons. The number of

bases in each exon and their positions on chromosome 11 are depicted in Table 3.6 (p88), to indicate regions of the gene to be sequenced.

Table 3.6 *CLNS1A* exons and positions on chromosome 11

Exon	Span	No. bases
1	75,889,882-75,890,095	213
2	75,897,788-75,897,925	137
3	75,901,868-75,901,970	102
4	75,902,616-75,902,724	108
5	75,905,005-75,905,179	174
6	75,907,980-75,908,070	90
7(i)	75,910,899-75,911,371	472
7(ii)	75,911,312-75,911,704	392

Given the size of exon 7, and for purposes of experimentation, the exon was divided into two overlapping parts, exon 7(i) spanned 472 bases (92 intronic region bases, 380 exonic bases) (75,910,899-75,911,371), exon 7(ii) spanned 392 bases (75,911,312-75,911,704). For exon 7(i) intronic sequence was used to facilitate primer binding in the PCR reaction. The intronic primer position was considered to be the best choice according to the primer design programs, indicated and explained further in Section 3.9.1.2.

3.8 Molecular Methods

3.8.1 Markers

Individuals of the WA family had been previously genotyped for linkage analysis using a variety of markers identified from a number of genetic maps (Chandler, 2001; Figure 2, p10 and Appendix 5). A total of 152 polymorphic markers with an average intermarker distance of 20cM were analysed for the WA family. These markers were chosen from a variety of sources including Genethon and CHLC and genotyped at ANRI. Chandler (2001) performed a two-point analysis and obtained a 2.11 Lod

score for markers D11S987 and D11S916 on 11q13-23 (See section 2.10). Therefore the linkage region was defined as being within this area.

3.9 Mutation Detection / Sequencing

3.9.1 The polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) can be performed *in vitro* and uses heat stable DNA polymerases to amplify a region of isolated DNA that is selected and bound by two short strands of complementary DNA sequences (the primers), chosen specifically to select the DNA regions of interest. DNA polymerase then copies the two strands by inserting complementary bases as it reads both strands of the DNA in a 3' to 5' direction from the bound initiating primers, producing two new strands of double stranded DNA. The original target sequences as well as the copies are then replicated repeatedly in subsequent tandem reactions resulting in an exponential increase in the number of copies of the region (Figure 6). In general the size of the region that can be amplified by this method does not exceed 2000 base pairs.

The reaction mix contains individual nucleotides together with primers, template DNA and Taq polymerase. Taq DNA polymerase amplifies template DNA by attaching nucleotides to the 3' hydroxyl group of the primer and extends along each template strand by adding complementary nucleotides. DNA synthesis proceeds in a 5' to 3' direction while template DNA strands are 'read' in a 3' to 5' direction. This ultimately gives rise to multiple copies of double stranded DNA. Exponential amplification is obtained by cycling of the above conditions for 20-40 cycles.

PCR amplifies short sequences of DNA approximately one billion times. Thus in a standard PCR, picograms of template DNA can be amplified to produce micrograms of DNA product. The resultant product will be multiple copies of a selected DNA region that can be routinely sized by comparison with known size markers using agarose gel electrophoresis.

Electrophoresis is a separation technique based on the movements of charged molecules in an electric field. Molecules of different lengths move at different rates and the different sized components of a mixture will be separated when an electric field is applied. Several factors have important effects on the mobility of DNA fragments in agarose gels and can be used to advantage, as in this study, to optimize separation of DNA fragments, including agarose concentration (2%) and voltage (100W).

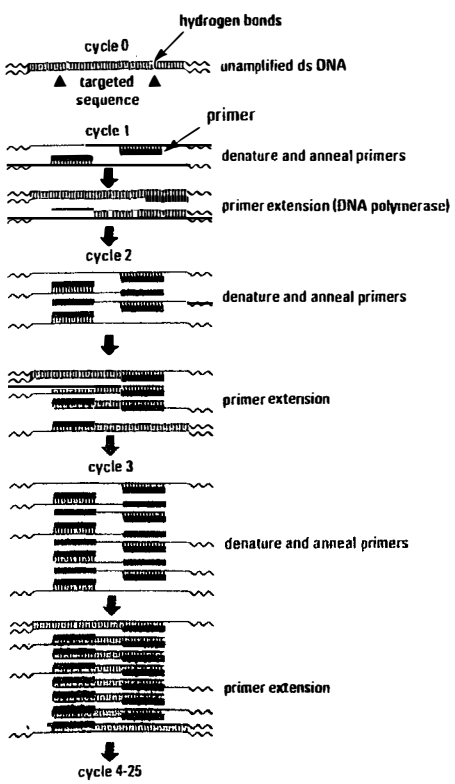


Figure 6 Schematic representation of the polymerase chain reaction (PCR)

3.9.1.1 Primers

Primers are chemically synthesised oligonucleotides, usually 20-25 nucleotides in length obtained commercially from several Biotechnology companies such as Geneworks, Research Genetics and Applied Biosystems. For this project, primers were designed to be complementary to both the 5' and 3' ends of the intronic sequences surrounding each exon under examination (Figure 7), as identified in GenBank. They were designed using primer design programs to comply with the criteria outlined in Section 3.9.1.2. Oligo's were suspended in water to a concentration of 100ng/μL and a working solution of 50ng/μL was prepared for each PCR reaction.

3.9.1.2 Primer design

The primer annealing temperature varies according to the primer melting temperatures (T_M). The annealing temperature is the temperature at which the primers will anneal with single strand DNA sequences and this is based on their length and *G/C* content. Those primers with the same annealing temperatures could be run at the same time in a thermocycler, so for the sake of time efficiency, primers for a single reaction were designed to have similar annealing temperatures. Similar annealing temperatures can be achieved by designing primers that are the same length and display similar *G/C* base content. Generally primers that are complementary to sequences displaying lower *G/C* content can be compensated for by an increase in length. Primers ideally comprised 40-60% *G/C* content. The approximate annealing temperature was calculated as follows; $T_A(°C) = [2(A+T) + 4(G+C)] - 5$ (Rozen & Skaletsky, 1996).

UVRAG Exon 2

Forward Primer →

tctggctcaataaataaatgactgagacaacttattacatttcatttttg
aactgcaacttccagaaaatttttataggaggaaaaatctgaatgccttct
ctttttttgttgttatttttcag*CGGCGTCTTCGACATCTTCGGA*
ACATTGCTGCCCGGAACATTGTTAATAGAAATGGCCA
TCAGCTCCTTGATACCTACTTTACACTTCACTTGTGTA
*GTACTGAAAAGATATATAAAG*gtaaggggatctgtggcctta
ctacccacagattgttatatttttattttttttgtaacacaagtgattctca
gtgcctcctgcttttgtggcttctc**ggaaacctaaattgttatgctgtc**

← Reverse Primer

Forward Primer: 3' - tctggctcaataaataaatgact - 5'
Reverse Primer: 3' - ctttggatttaacaatacgacag - 5'

Figure 7: An example of Primers Designed to amplify an exon of interest

UVRAG exon 2 sequence from GenBank. Primers were designed to be complementary to intronic sequences 100bp on either side of the exon. Primers are depicted in bold case. Exonic sequence is depicted in uppercase, intronic sequence is in lower case.

Primers were designed to be approximately 22 bases long and complementary in sequence to intronic sequences located within 100 bases on either side of the exonic sequence. Furthermore, to reduce the chance of primer-dimer formation, primers themselves did not contain any sequences that were complementary to sequences within the primer or its partner; these include complementary triplets or complementary sequences within or between primer pairs particularly at the 3'-end of each primer. The 3'-end of each primer was designed so as to avoid ending with a 'T' as this increases the chance of mismatch as it does not bind with sufficient strength to complementary DNA. Furthermore, the forward and reverse primers were designed to be as similar in annealing temperature as possible. In specific instances where this was not possible, an average annealing temperature was used in the PCR. Again, in certain instances where this approach was not successful, either forward

or reverse primers were chosen from exonic sequences and designed to have a similar annealing temperature to that of the intronic primer of the pair. Where the exons were very large, primers were designed to amplify adjacent, overlapping exonic sequences, such as in UVRAG exon 15 where the exon was too large to sequence using one set of exclusively intronic primers (Appendix 16).

Primers were also run through the BLAST program on GenBank to check the complementarity to other known gene sequences, particularly at the 3'-end of each primer. This ensured that the designed primers would not anneal elsewhere in the genome.

3.10 PCR Thermocycling

The thermocycling of a polymerase chain reaction (PCR) comprises three components; denaturing, annealing and extension. The denaturing step consists of 10-30 seconds at 94°C and is responsible for the separation of all double-stranded DNA. The annealing step generally varies between 10-30 seconds although this is dependant on the size of the DNA region to be amplified and the primer sequence. As previously described the annealing temperature is deduced from the melting temperature of the primer. The extension step involves 1 minute at 72°C, the optimal synthesis temperature for many thermostable polymerases.

For this project, PCR cycling conditions were generally as follows: initial denaturing step of 5 minutes at 94°C followed by 30 cycles of 94°C for 15 seconds, 55°C (T_A) for 30 seconds, 72°C for 1 minute followed by a chasing step of 72°C for 10 minutes (Table 3.7, p94). The cycle primer annealing, DNA amplification and denaturation steps were

repeated, anywhere up to 40 times following the initial denaturation. This yields continuous denaturing and amplification of the template DNA, producing new products that act as templates for the following cycle. Copies of template DNA flanked by the primers multiply exponentially providing a final mix of multiple copies of double stranded DNA of a defined length and sequence. DNA synthesis occurs at a rate of approximately 1000 bases/min thereby achieving a billion fold increase in template DNA following 30-35 PCR cycles.

Table 3.7 Typical PCR Thermocycling Conditions

Steps		Time
Temperature		
Initial denaturation	2:15	94°C
Amplification cycle X 30 cycles		
Initial denaturation	45sec	94°C
Annealing	60sec	*
Extension	60sec	72°C
Final Extension	10min	72°C
Storage	infinity	4°C

* approximately 60°C but depends on primer melting temperature

3.11 PCR Reaction

PCR reactions generally comprised 2.0μL of 10 X Qiagen amplification buffer, 0.3μg of each dNTP, 1μL of each primer at 50ng/μL, 0.1μL of Taq DNA polymerase (at 5u/μl), 5μL of target DNA (at 1ng/μl) in a total volume of 20μL (Table 3.7, p94). "HotStar" Taq DNA polymerase was used to amplify exons which could not be amplified using standard Taq DNA polymerase. HotStar reactions contain 250 units of Taq polymerase, 10mM of each dNTP, 5x Q-solution, 25mM $MgCl_2$ and 10X

PCR Buffer per reaction, premixed in a 10μL HotStar master mix, 1μL each of 50ng/μL primers, 3μL ddH_2O and 5μL template DNA at 1ng/μl (Table 3.8, p95).

HotStar Taq DNA polymerase is supplied as an enzyme antibody complex that exhibits no polymerase activity until heated at 95°C for 15 minutes which denatures the antibody and activates the HotStarTaq DNA Polymerase (Table 3.9, p95). This ensures highly specific amplification by preventing mispriming at non-specific sites which tends to take place at lower temperatures i.e. preventing extension of non-specifically annealed primers and primer-dimer extensions which characteristically form during the reaction setup and initial heating period.

Table 3.8 Hot Star PCR Reaction Mix

10μL Hot Star master mix
1μL forward / reverse primer
3μL ddH_2O
5μL template DNA
20μL PCR reaction

Table 3.9 Hot Star PCR Thermocycling conditions

Steps	Time	Temperature
Hot Star	14:30	94°C
Amplification cycle	30 times	
Initial denaturation	30 sec	94°C
Annealing	60 sec	*
Extension	60 sec	72°C
Final Extension	10 min	72°C
Storage	infinity	4°C

* annealing temperature, approximately 60°C

Table 3.8 depicts a typical Hotstar PCR reaction that was used to amplify DNA regions that could not be amplified using standard Taq DNA polymerase.

Table 3.9 depicts Thermocycling conditions for Hotstar PCR.

3.12 Electrophoresis

All Amplification products were visualized by electrophoresis on a 2% agarose gel. DNA grade agarose powder was added to 1X TAE Buffer

to achieve a concentration of 2%. The solution was shaken and melted in a 700W microwave until the DNA grade agarose powder was completely dissolved. The gels were poured in open-ended horizontal trays, with wells formed at one end of the gel by a suspended comb inserted into the gel.

Agarose gels were run under 1 X TAE buffer in horizontal tanks, with DNA loaded into the submerged wells using a high density loading dye containing bromophenol blue and ficoll. A 100V current was applied across the tank for 60 minutes. The DNA moved through the gel towards the positive electrode (cathode) due to the negative charge carried by phosphate groups in DNA alongside a 1kB+ ladder (Figure 8).

Figure 8 1Kb Plus Ladder

900ng/lane; 0.9% agarose gel stained with ethidium bromide

DNA in the agarose gel was visualized by staining the gel in 100ml of ethidium bromide solution (1μg/100mL) for 10 minutes and gels were then photographed over UV light on a transilluminator. Alternatively, EtBr was added to the melted agarose (1μL EtBr was added per 100mL agarose). PCR products, visualized by gel electrophoresis, were assessed for size and purity by reference to a DNA size ladder as well as to positive and negative PCR controls.

3.13 Purification of PCR Products

To purify PCR products for subsequent sequencing reactions, QiaQuick buffers and spin-columns (Qiagen) were used (protocol summary, Table 3.10, p99). The QiaQuick PCR purification kit separates DNA strand by size and allows the efficient binding of double stranded PCR products as small as 100bp to a silica membrane on a spin column while allowing (99.5%) of primers up to 40 nucleotides to pass through the column. During the DNA absorption step, unwanted primers and impurities, such as salts, enzymes, unincorporated nucleotides, agarose, dyes, ethidium bromide, oils and detergents which do not bind to the silica membrane, flow through the column and are discarded. Salts are quantitatively washed away by the ethanol-containing buffer PE. Any residual buffer PE, which may interfere with subsequent enzymatic reactions, is removed by additional centrifugation steps.

A volume of Buffer PB supplied by the manufacturer was added to each completed PCR reaction in the ratio 5:1 Buffer:PCR volume. This mixture was then loaded onto a spin-column placed in a collection tube and centrifuged at 13 000 rpm for 30-60 seconds. Any flow-through following centrifugation was discarded, as DNA binds to the spin-column

matrix. 700μL of 20% PE buffer (in ethanol), supplied by the manufacturer, was dispensed onto each spin column and these were similarly centrifuged for 30-60 seconds. The flow-through was again discarded and the columns centrifuged for a further 30-60 seconds to collect and discard any remaining PE buffer.

Spin-columns containing bound DNA were transferred into clean labeled 1.5mL eppendorf tubes. Depending on the quantity of DNA, as assessed by PCR product band intensity, 30μL elution buffer (for lower DNA concentrations) or 50μL ddH_2O (for higher DNA concentrations) was added to the center of each spin-column. The columns were centrifuged for the final time at 13 000 rpm for 30-60 seconds. DNA bound to the spin-column matrix dissolves in either elution buffer or ddH_2O and moves through the column into the supporting eppendorf tube in the final centrifugation. This purification method removes all PCR reagents and DNA fragments smaller than 100 bases in length providing pure template DNA for sequencing reactions.

Table 3.10 QiaQuick protocol Summary

i)	5 volumes of buffer PB were added to 1 volume of PCR sample & mixed
ii)	A spin column was added into a 2mL collection tube
iii)	To bind DNA, the sample was applied to the QiaQuick column and centrifuged at 13 000rpm for 30-60 seconds
iv)	Flow-through was discarded and the QiaQuick column was replaced in the same tube
v)	To wash, 0.75mL Buffer PE was added to QiaQuick column and centrifuged at 13 000rpm for 30-60 seconds
vi)	Flow-through was discarded and the QiaQuick column was placed back into the same tube. Columns were centrifuged for an additional 1 minute at max speed (13 000rpm)
vii)	QiaQuick column was placed in a clean 1.5mL centrifuge tube
viii)	To elute DNA, 50μL Buffer EB or ddH_2O was added to the centre of the QiaQuick membrane and centrifuged at 13 000rpm for 1 minute. Alternatively, for increased DNA concentration, 30μL elution buffer was added to the centre of the QiaQuick membrane, the column was left to stand for 1 minute then centrifuged at 13 000rpm for 1 minute.
ix)	Isolated DNA was stored frozen prior to being sequenced.

3.14 Mutation Detection

The ideal outcome of a positional cloning project is the identification of the disease-causing mutation(s). The most suitable methods available for detecting a mutation in a short DNA strand include sequencing of a PCR product to detect small mutations, such as single base deletions, insertions or substitutions (such as trinucleotide repeat amplifications).

3.14.1 Heterozygote sequencing

In an autosomal dominant disorder, such as HFM, the presence of one copy of the mutated gene may cause the disease phenotype. Detection of a disease mutation in a candidate gene is complicated in these cases primarily because the sequencing reaction is performed on a combination of both alleles at a locus and the mutated allele will be

partially masked by the normal sequence of the normal allele. The fluorescent detection system most commonly used for sequencing in modern molecular biology can only distinguish a heterozygote by identifying a position at which the computer cannot distinguish whether the signal is due to one base or another at which point an N (unknown) is inserted into the sequence. This is complicated further by the consistency of the signal which depends not only on the dye being detected, but also on the surrounding sequence. A considerable effort has been made to reduce this noise from surrounding sequences by creating dye terminator and specialist dye systems (ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kits). The accuracy of heterozygote mutation detection is thus now greatly increased.

3.15 Sequencing Reactions

An ABI 373 DNA Sequencer was used for all sequencing. Generated sequences were analysed using the *Sequence Navigator* program. The analysed sequences were stored as data files on a G4 Macintosh computer in the Centre for Human Genetics, Edith Cowan University, in a file only accessed by this researcher so as to conform to ethical requirements.

3.16 Sequencing

Since the advent of thermostable polymerases, efficient fluorescent labeling systems and high throughput gel and capillary electrophoresis, sequencing is now done almost exclusively using the dideoxy method (Sanger, Nicklen & Coulson, 1977). Two variations of the dideoxy method are in general use, dye-terminator and dye-primer sequencing. The procedures vary in that the fluorescent label is

attached to either the dideoxy nucleotides (dye terminator) or the 5' end of the sequencing primer (dye primer).

Dye terminator sequencing is usually the method of choice for one-off sequencing of PCR products, and was the method employed in this project.

3.16.1 Direct sequencing of transcripts

Each dideoxy sequencing reaction contains a mixture of both deoxynucleotide triphosphates, dNTPs and dideoxynucleoside triphosphates (ddNTPs), dideoxy-adenosine triphosphate (ddATP), dideoxythymidine triphosphate (ddTTP), dideoxyguanosine triphosphate (ddGTP) and dideoxycytidine triphosphate (ddCTP) each ddNTP labelled with a different fluorophore. Also present in the reaction are a single forward or reverse primer, $MgCl_2$, 10X buffer and Taq DNA polymerase. Chain termination occurs when ddNTPs are incorporated into the growing chain because they lack a hydroxyl group at the 3' end of the carbon atom of the ribose sugar and therefore no subsequent nucleotides can be joined onto the growing chain. The concentration of ddNTPs is much lower than that of the dNTPs, consequently chain termination occurs randomly ultimately producing a series of products, each one nucleotide longer than the next. During electrophoresis of the sequenced products, a monitor records the four different fluorophore signals as they pass a detector. Data is recorded electronically as a chromatographic intensity profile with A, T, G, and C translation into a detailed chromatogram (Appendix 17).

3.16.2 Dye terminator sequencing reaction

Sequencing reactions were performed in 0.2mL microtubes, as per ABI protocol. Each tube contained 2-4μL purified PCR template dsDNA (at 1ng/μL), 1-3μL ddH_2O, 4μL Big Dye Terminator Mix and 1μL (of 50ng/μl) of the forward or reverse primer appropriate for the DNA template (Table 3.11, p102). The volume of template DNA and ddH_2O added to each sequencing reaction was adjusted according to initial DNA concentrations. When necessary, purified DNA PCR product was concentrated by vacuum centrifugation to a minimum volume of 10μL. The reaction mix was cycled on a thermocycler (Table 3.12, p102).

Table 3.11 Sequencing Reaction

2μL BigDye V3.1
1μL BigDye Buffer (10X)
0.5μL forward / reverse primer (50ng/μl)
2-4μL template DNA(10ng/μl)
4.5-2.5μL ddH_2O
10μL total volume

Table 3.11 summarizes the composition of a typical sequencing reaction

Table 3.12 Thermocycling Conditions for a Typical Sequencing Reaction

Step	Time	Temperature
Initial denaturation	1 min	96°C
Amplification cycle X 25 times		
	30 sec	96°C
	30 sec	50°C
	4 min	60°C
Storage	infinity	4°C

Table 3.12 indicates the cycles used for a sequencing reaction

3.16.3 DNA sequencing reaction cleanup protocol

Unincorporated ddNTPs, enzymes and salts were extracted from sequencing reaction products by ethanol / sodium acetate precipitation as follows; 2μL of 3M sodium acetate (pH 4.6) and 50μL of 95% ethanol were added to each completed sequencing reaction. Tubes were shaken and left to stand at room temperature for 5 minutes, then centrifuged for 45 minutes at 13 000 rpm. Supernatants were discarded and 150μL of 70% ethanol added to each pellet. Tubes were again centrifuged for a further 10 minutes at 13 000 rpm. Supernatants were discarded and the pellets briefly dried at 65°C. These were forwarded to the Australian Neuromuscular Research Institute (ANRI) for sequencing on an ABI 373 Automated DNA Sequencer.

3.16.4 Analysis of sequence data

Sequencing results and chromatograms received from ANRI were imported into and analysed using computer softwares *SeqEd*, Version 1.0.3 and *Sequence Navigator.* Sequencing results were compared with the equivalent normal sequence, both manually and by computer analysis using BLAST analysis online available at NCBI on GenBank. Any base discrepancies are evaluated as polymorphisms or as potential disease causing mutations by comparison with coding and non-coding sequences and translated amino acid sequences.

Chapter 4: Results: Identification, Amplification and Sequencing of Candidate Genes for Hemifacial Microsomia

4.0 Introduction

Identification of markers and candidate genes in the chromosome 11 linkage region thought to be associated with HFM in the West Australian family are discussed. Moreover the results of sequencing of four candidate genes in the chromosome 11 linkage region will be presented.

4.1 Marker Analysis to Refine the Chromosome 11 Linkage Region

The major result of this project has been to reduce the size of the chromosome 11 HFM region, linked to HFM in the WA family under investigation (Chandler, 2001) (Table 4.3, p108). Through the redefinition of marker positions and the employment of genome mining methodology, the chromosome 11 candidate region has been reduced in size from an estimated 18.8 million DNA bases to 13.5 million DNA bases. This considerable reduction in size of 5.3 million bases will simplify future studies. By regularly monitoring the positions of the markers in the Draft Human Genome, the candidate region was reduced by 28.2%. Essentially this involved continual inspection of the 5' and 3' chromosome 11 marker positions in the human genome sequence, the draft of which was regularly updated, thereby resulting in reorganization of markers and their positions on the genome in databases such as Ensembl. The sequencing of the genes themselves further reduced the linkage region by another 355,995 bases.

More specifically, the linkage region was assumed to be between bases 66,988,646-83,391,363 in September 2002 (D11S1883-D11S911), according to data from ENSEMBL. Yet, in August 2001, the marker D11S1883, the 5' limit of the HFM linkage region was located on the draft human genome

sequence at position 70,661,457-70,754,563. Furthermore, in December 2001 the same marker on the same database was positioned at 72,543,444-73,659,301 and in April 2002 was between 64,433,363-66,012,757. Alternatively, D11S4207, the 3' limit of the HFM linkage region was located in the draft human genome at position 84,621,967-84,822,321 (August 2001). Again, in December 2001, the marker was relocated to position 74,683,628-74,883,976. Due to inconsistency in the marker positions in April 2002 the locations of markers achieved in December 2001 were assumed to be correct at the commencement of experimentation, confirmed by Ensembl. Consequently, the linkage region was reduced to a length of 14,160,864 bases and the linkage region was amended to be Chr11: 64,533,373 - 75,971,382.

This was an extremely significant result and proved useful in reducing the region to be sequenced. As a consequence 100 known and predicted genes were excluded as potential candidate genes for HFM, since genes previously located within the chromosome 11 HFM linkage region were excluded, thus reducing the number of potential candidate genes in this region. The next step, the choice of candidate genes by genome mining, was therefore simplified by reducing the number of possible candidates in the linkage region between the markers D11S1883 and D11S911.

Table 4.1 Icelandic Map Chromosome 11 Marker positions

4th September 2002

D11S1765	66,988,646
D11S1883	(not currently mapped)
D11S987	77,403,003
D11S4178	77,801,817
D11S4136	79,387,634
D11S4139	(not currently mapped)
D11S1314	82,286,854
D11S4207	(not currently mapped)
D11S916	83,161,640
D11S911	(not currently mapped)
D11S937	89,391,363

The Icelandic Map (Weber, 2002) is based on the genotyping of 5,136 microsatellite polymorphisms in 146 nuclear families, all from Iceland, containing 1,257 meioses and is considered to be of substantially improved resolution compared to previous human genetic maps. The Icelandic Map on the 4th September 2002, was explored for the aforementioned genetic markers (Table 4.1, p107). Four of the markers previously mapped on the Ensembl database, namely D11S1883, D11S4139, D11S4207 and D11S911 were not found on the Icelandic Map. The D11S911 marker was of particular relevance as it is the 3' marker flanking the HFM linkage region. Nevertheless the genes between markers D11S1765 and D11S916 were consistent with those identified by Ensembl on December 2001 and thus selection of candidate genes was confirmed.

The positions of the markers on chromosome 14 surrounding the chromosome 14 HFM linkage region markers were assessed in July 2002 (Table 4.2, p 108) and found to be consistent with previous marker positions for chromosome 14 identified on Ensembl thereby indicating that changes were minimal and therefore that chromosome 11 markers may have stabilised.

Table 4.2 Chromosome 14 markers and their positions on ENSEMBL as at 22 July 2002

Marker	Position
D14S1142	92 880 607 - 92 880 769
D14S1143	93 119 794 - 93 119 956
GATA168F06	93 507 497 - 93 507 718
D14S987	95 080 876 - 95 081 183
D14S65	96 108 679 - 96 108 841
	96 108 697 - 96 108 847
D14S267	97 711 801 - 97 712 038
	97 711 831 - 97 712 043

Consequently, the positions of markers for the chromosome 11 HFM linkage region on the draft human genome were mapped in July 2002 (Table 4.3, p 110) and the linkage region for chromosome 11 was amended to be Chr11: 64,533,373 - 75,971,382 (Table 4.3, p109) as the final positions.

Table 4.3 Chromosome 11 Genetic markers and locations as at 3rd July 2002

D11S1883	(64,533,373 - 64,533,624)
D11S987	(not currently mapped)
D11S4178	(not currently mapped)
D11S4136	(69,221,479 - 69,221,668)
D11S4139	(69,946,835 - 69,946,983)
D11S1314	(not currently mapped)
D11S4207	(not currently mapped)
D11S911	(75,971,191 - 75,971,382)

The final positioning of the above markers were assumed to be correct and were used in candidate gene selection and experimentation

4.2 Genome Mining Results

Genes in the chromosome 11 linkage region were subsequently identified using the NCBI database. The sequences of these genes were then compared to sequences from genome databases in GenBank and GDB. Using databases such as NCBI, GenBank and GDB a list of 131 genes was compiled for the chromosome 11 HFM linkage region (Appendix 9). Specific information on each of these genes was then obtained through sequence and database analysis from data on PubMed and LocusLink (Appendix 12). Candidate genes were selected based on their compliance with selection criteria. The more criteria a particular gene fulfilled, the more its candidate status was enhanced. From this information, four genes were identified as prime candidates, based on selection criteria outlined in Section 3.6.

4.3 Sequencing of *UVRAG, ARIX, CLNS1A* and *GARP*

Four genes, namely *UVRAG, ARIX, CLNS1A* and *GARP*, were chosen as the most likely candidate genes based on their compliance with the three selection criteria, of embryonic expression, association with axial asymmetry and possession of / association with trinucleotide repeats.

4.3.1 *UVRAG* (UV Radiation Resistance Associated Gene)

UVRAG was selected as the prime candidate gene for chromosome 11 HFM. It is located in position 74,047,886-74,376,304 on chromosome 11q13.5 (Map Viewer, Appendix 11). This gene fulfilled all three components of the selection criteria (see Section 1.5, p8). *UVRAG* is expressed embryonically and is associated with axial asymmetry (Ensembl gene ID ENSG00000137493) (Perelman *et al.*, 1997). Furthermore *UVRAG* contains two (aac) trinucleotide repeats and an (aat) trinucleotide repeat (positions Chr11:74,105,993-74,106,007, 74,272,978-74,272,998 & 74,287,745-74,287,759 respectively) (Figure 9), thereby making it a prime candidate for chromosome 11 HFM.

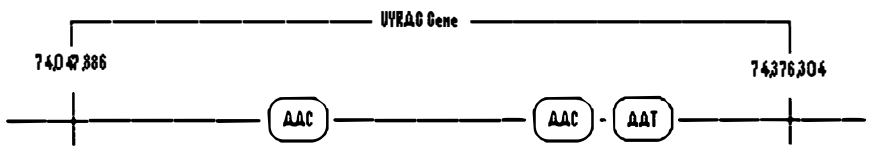


Figure 9 Position of trinucleotide repeats in relation to *UVRAG*

4.3.2 *ARIX* (Aristaless homeobox)

ARIX is located on chromosome 11 at position 75,883,451-75,887,841 (Map Viewer, Appendix 11). *ARIX* gene ontology, according to LocusLink suggests regulation of downstream target gene expression, morphogenesis and differentiation (Johnson *et al.*, 1996) (LocusID 116648). It is a DNA binding protein involved in neuronal development, neurotransmitter synthesis and storage. More specifically it was identified as a candidate gene for chromosome 11 HFM as it is associated with left-right asymmetry determination and craniofacial abnormalities (Pattyn *et al.*, 1997). Specifically *ARIX* is expressed in noradrenergic cell types of the sympathetic nervous system, brain and adrenal medulla (Johnson, Smith, Johnson, Rhodes, Rinchik, Thayer & Lewis, 1996). Furthermore two (gtt) trinucleotide repeats (positions Chr11:75,820,692-75,820,715 & 75,934,321-75,934,344 respectively) are situated on either side of the gene (Figure 10) and these repeats could be in regulatory regions of the gene which normally lie 5' and 3'. Regulatory regions may be located in close proximity to the gene and others may be several kilobases away from the actual gene.

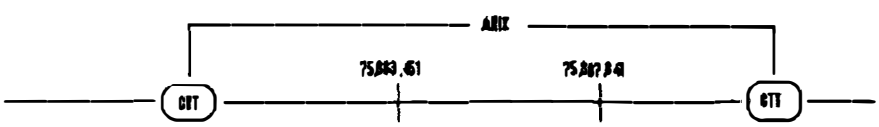


Figure 10 Position of trinucleotide repeats in relation to *ARIX*

4.3.3 *CLNS1A* (Chloride channel, nucleotide-sensitive, 1A)

CLNS1A is located on chromosome 11, position 75,889,882-75,911,533 (Map Viewer, Appendix 11). Gene ontology for *CLNS1A* as indicated in Ensembl suggests involvement in vision, circulation, small molecule transport and photoreception (Ensembl gene ID ENSG00000074201) (Bekri *et al.*, 1997; Buyse *et al.*, 1996). It encodes an integral membrane protein and the gene is alternatively referred to as *CLC1, ICLN, ICln* or *CLNS1B*. More specifically *CLNS1A* contains an (att) trinucleotide repeat and a (gtt) trinucleotide repeat at positions Chr11: 75,893,893-75,893,916 and 75,900,770-75,900,787 respectively (Figure 11). These repeats may lie in regulatory regions generally located 5' and 3' of the gene.

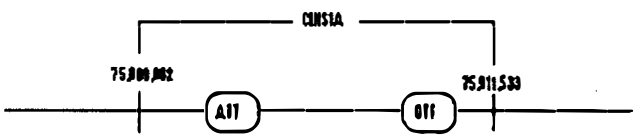


Figure 11 Position of trinucleotide repeats in relation to *CLNS1A*

4.3.4 *GARP* (Garpin complex)

The *GARP* gene is located on chromosome 11 at position 74,891,183-74,903,592 (Map Viewer, Appendix 11). Gene ontology according to Pruitt, Katz, Sicotte and Maglott (2000) proposes that *GARP* encodes an integral plasma membrane protein. *GARP* is situated near an aat and a gat trinucleotide repeat (position Chr11: 74,858,600-74,858,623 & 74,909,997-74,910,011 respectively (Figure 12).

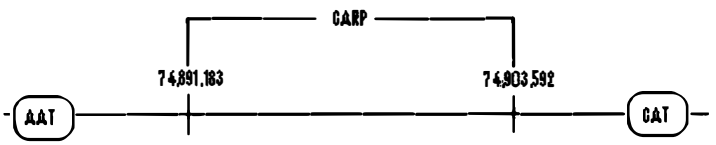


Figure 12 Position of trinucleotide repeats in relation to *GARP*

As mentioned previously, the functions of *GARP* are diverse and include receptor interactions, enzyme inhibition, cell adhesion and cellular trafficking. It has been reported to be involved in early mammalian development, neural development, cell polarization, regulation of gene expression, the cytoskeletal framework and apoptosis signalling (Pruitt, Katz, Sicotte & Maglott, 2000). Pruitt *et al.* (2000) also identified the homologous relationship between *GARP* and the *Drosophila Aristaless* gene and the human gene *ARIX*. Human *GARP* and *ARIX* genes are specifically expressed in noradrenergic cell types of the sympathetic nervous system, brain and adrenal medulla (Pruitt *et al.*, 2000), so alterations to these genes may affect craniofacial development by affecting neural function.

4.4 Results of Homology Investigations

Despite homologies of *ARIX* and *GARP* to regions on mouse chromosome 7, no genes associated with characteristics pertaining to HFM were identified on the mouse chromosome at these regions. No other candidate genes on chromosome 11 showed homology to relevant mouse genes (Appendix 15).

A number of animal models have been described, including *far* and *Hfm* but other genes are homologous to the genes identified on chromosome 11 and both present different aetiologies to that of human HFM (Naora *et al.* 1994). These models therefore, were not responsible for additional candidate gene identification in this project. None of the genes within the chromosome 11 linkage region demonstrated homology to the genes within the chromosome 14 linkage region although this was considered an essential step in the identification of candidate genes for HFM. No relevant data was obtained using this method so PCR and sequencing of the three candidate genes was adopted as the most sensitive.

Despite no homologies being identified through these comparative methods, they were nonetheless considered relevant analytical techniques for candidate gene identification and selection.

4.5 Sequencing of HFM candidate genes in DNA from a HFM patient

A search of NCBI Genome Browser identified genomic sequence for each of these genes. Primers were specifically chosen to amplify the intronic and exonic regions of all these genes and were designed in accordance with parameters detailed in Section 3.9.1.2, p91 (Appendix 7). Primers were designed to amplify the exons including all intron/exon boundaries (Figure 6, p90) so that any mutations at intron/exon boundaries, likely to cause splicing errors and thus errors in the final gene product, would also be identified along with mutations in the coding exonic regions.

The patient tested in this project is individual D02-517 (III:6). The patient phenotype was described in Section 3.5.1 (p77) as characteristic of an affected member of the WA family (see WA pedigree – Figure 1). This patient was diagnosed with HFM and therefore the DNA from this patient was isolated from blood and utilised to find mutations in the chromosome 11 linkage region. Furthermore the DNA of this patient was already accessible for research purposes. Primers were then designed and used to sequence the exons of the four candidate genes in to identify possible mutations in the HFM linkage region.

4.6 PCR amplification of *UVRAG* and sequencing results

4.6.1 Exon 1 – Amplification and sequencing

Forward primer	74 049 814F	5'-CTAAGCCAATGAGCGCTCC-3'
Reverse primer	74 050 433R	5'-AGCTGGCTGGCTGCTTGTC-3'

An annealing temperature of 55°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products incorporating Q-solution were electrophorised at 100V for 1hr. The PCR products were then QIAquicked and eluted with 50μL ddH_2O. PCR products were sequenced using 3.5μL template DNA. No mutations were detected in affected or unaffected DNA, when assessed by comparison of sequences between each sample and GenBank (See Appendix 17 for sequencing chromatogram).

4.6.2 Exon 2 – Amplification and sequencing

Forward Primer	74 086 399F	5'-GAGACTGGGTTATGGTTTCTGG-3'
Reverse Primer	74 086 741R	5'-AAGCAGGAGGCACTGAGAATCA-3'

An annealing temperature of 61°C with 35 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products incorporating Q-solution were electrophorised at 80V for 1hr20min (Figure 13). In an attempt to cleanup the PCR product another PCR was ran using an annealing temperature of 61°C with 25 cycles. This was also successful and these PCR products were consequently QIAquicked and eluted with 30μL elution buffer. Sequencing using 2μL template DNA was successful. However no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank. (See Appendix 17 for sequencing chromatogram).

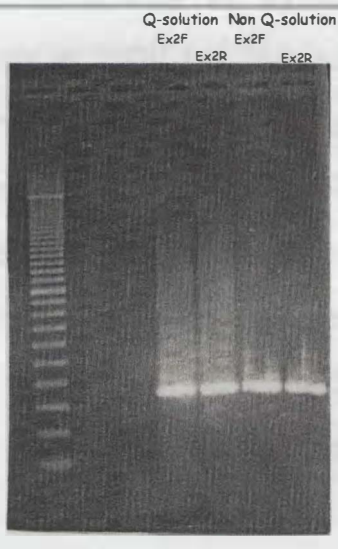


Figure 13 - UVRAG exon 2 PCR Amplification visualized on 2% EtBr stained agarose gel

4.6.3 Exon 3 - Amplification and sequencing

Forward Primer	74 095 976F	5'-cactagcaaccaggtaacctaca-3'
Reverse Primer	74 096 685R	5-'gtatacatgtgccatgtcggtgtg-3'

Due to amplification difficulties, an alternative technique was used to amplify this exon. Three differing concentrations of template DNA were used in three separate PCR reactions: 1μL, 1/100μL and 1/1000μL. Higher concentrations (1μl) of template DNA gave the best result when cycled at 68°C for 45 cycles. However sequencing was unsuccessful, therefore new primers were designed using Web-based primer design program [http://seq.yeastgenome.org/tmp/sorted.tmp.26024.html].

Forward primer	74 094 747F	5'-TGAGTTTTGTGACCTTTTCCA-3'
Reverse primer	74 096 084R	5'-TTCTTTGTTTTCAACACAGCC-3'

Using these new primers, a PCR reaction was performed and an annealing temperature of 47°C with 30 cycles proved unsuccessful as assessed by electrophoresis at 100V for 1hr, consequently a HotStar PCR was prepared using control DNA (unaffected/unrelated male and female DNA) (Figure 14). Interestingly, the control DNA amplification was successful when amplified. Similar conditions to these were used for the affected DNA (i.e. HotStar Taq DNA polymerase). PCR amplification was achieved by running a Hot Star PCR at 50°C for 30 cycles (Figure 14). Thus, the PCR was successfully amplified for affected and unaffected DNA and the PCR fragment sequenced and analyzed but it revealed no mutations in affected or unaffected DNA when sequences were compared relative to each other and GenBank.

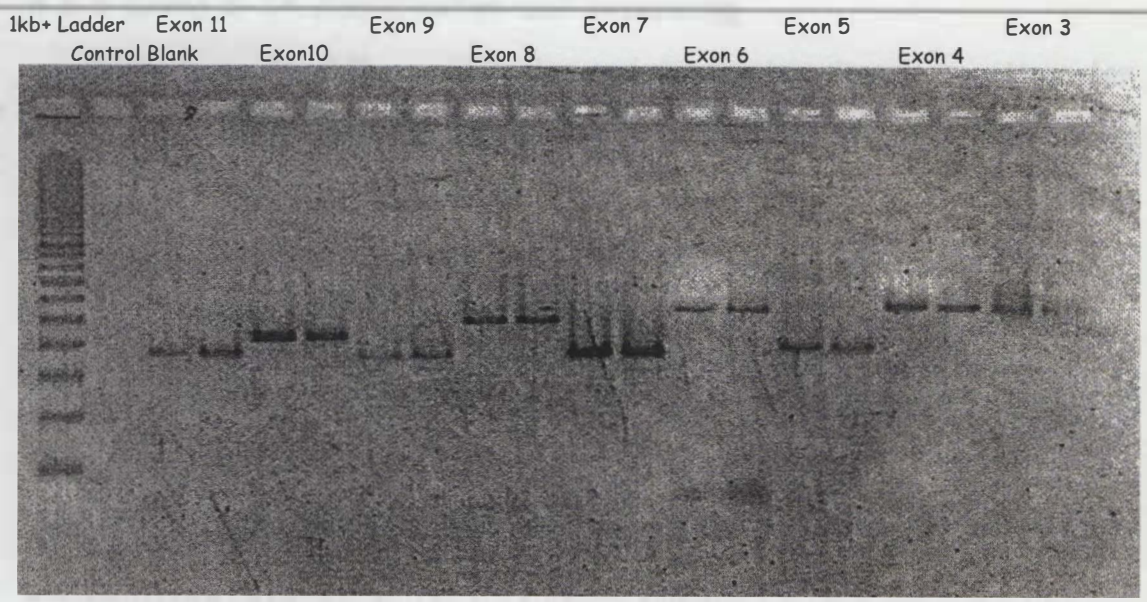


Figure 14 - PCR amplification using a 2% agarose gel of *UVRAG* exons 3-11

Since all gel pictures are similar, the gel pictures shown are used as typical example of gels obtained through experimentation throughout this project

4.6.4 Exon 4 - Amplification and sequencing

Forward Primer	74 114 287F	5'-ccatgtaagtgagtgatagg-3'
Reverse Primer	74 114 840R	5'-catcaaggtctctcttagtg-3'

A Hot-Start PCR using an annealing temperature of 53°C for 25 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1hr. PCR products were QIAquicked and eluted with 40µL ddH_2O. PCR products were sequenced using 2µL template DNA but again no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.6.5 Exons 5-14 Amplification and sequencing

Conditions for PCR amplification and sequencing cleanup reactions for exons 5-14 are displayed in Table 4.4, p119. All PCR reactions were performed as for standard PCR reactions as described in Sections 3.9 and 3.10. PCR products were electrophorised at 100V for 1hr (Figure 14). PCR products were then QIAquicked and eluted with volumes given in Table 4.4, p119. PCR products were sequenced using 3μL template DNA. For all these exons, no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

Table 4.4 Amplification and Sequencing of UVRAG exons 5-14

Exon	Primer	Primer sequence	Annealing Temperature	Number of Cycles	Elution
5	Forward - 74 123 374F Reverse - 74 123 748R	5'-gccggtttctcagtattgaag-3' 5'-atggactggatggttatgtgc-3'	57°C	30	30μL elution buffer
6	Forward - 74 146 376F Reverse - 74 146 915R	5'-gctccttctccactcttgtcag-3' 5'-ttctctcccactgtccagag-3'	60°C	45	30μL elution buffer
7	Forward - 74 196 035F Reverse - 74 196 397R	5'-ctgaggctctttgttgag-3' 5'-gcaggcacttcatgcgttac-3'	53°C	30	30μL elution buffer
8	Forward - 74 217 800F Reverse - 74 218 291R	5'-ttgtgtagtgtctgttgggtc-3' 5'-ccctgtcattaagtgatgtgtg-3'	58°C	30	40μL ddH_2O
9	Forward - 74 238 567F Reverse - 74 238 924R	5'-ctctcacagtcagggatttggc-3' 5'-tggcacagtgcatgacacacag-3'	63°C	30	30μL elution buffer
10	Forward - 74 242 073F Reverse - 74 242 494R	5'-ctgtggcctattacagacag-3' 5'-tgatgagtgcatcccacca-3'	54°C	30	40μL ddH_2O
11	Forward - 74 243 311F Reverse - 74 243 678R	5'- caccatgtactcatcatccagc-3' 5'-atctcccatcaaggagtctg-3'	55°C	30	25μL elution buffer
12	Forward - 74 251 478F Reverse - 74 251 635R	5'-gatggaagcattgctgttgc-3' 5'-tctctctccttttccgtcag-3'	55°C	30	40μL ddH_2O
13	Forward - 74 300 291F Reverse - 74 300 550R	5'-gcctctgttgacaacttggag-3' 5'-aaatgcgagggtcaactgcag-3'	61°C	30	30μL elution buffer
14	Forward - 74 350 386F Reverse - 74 350 755R	5'- gtgtgcatactcatacacgtgc-3' 5'-tgtggagcaatgtggaagtcag-3'	61°C	30	50μL ddH_2O

4.6.6 Exon 15 - Amplification and sequencing

Due to the size of *UVRAG* exon 15 (1,940bp) and for ease of sequencing, this exon was divided into four sections - *UVRAG* exon 15(i)-(iv) by choosing primers suitably located within the exon. Primers were selected so that regions to be sequenced overlapped (Figure 15).

4.6.6.1 Exon 15 (i) Exonic Primers - Amplification and sequencing

Forward Primer 74 375 366F 5'-gatccagaaggcttacactgag -3'

Reverse Primer 74 375 968R 5'-*CAGAAGAAATCATCGGGCTGGT* -3'

An annealing temperature of 61°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products incorporating Q-solution were electrophorised at 100V for 1hr. These PCR products were then QIAquicked and eluted with 40μL ddH_2O. Sequencing was achieved using 3μL template DNA and was not definitive. Despite several attempts to complete sequencing were unsuccessful. The sequence contained several repeat nucleotide sequences which are difficult to sequence. [No difference in PCR product size was observed when assessed by gel chromatography thereby suggesting no obvious deletions or insertions].

4.6.6.2 Exon 15(ii) Exonic Primers - Amplification and sequencing

Forward Primer 74 375 842F 5' - *AATGGCACTCTCCTACCCAG* - 3'

Reverse Primer 74 376 425R 5' -*GAGCTTACAGCTCGAGTCACCT*- 3'

An annealing temperature of 60°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products incorporating Q-solution were electrophorised at 100V for 1hr.

(i) *UVRAG* 74 375 366F 5' - gatccagaaggcttacactgag - 3'
UVRAG 74 375 968R 5' - ACCAGCCCGATGATTTCTTCTG - 3'
(ii) *UVRAG* 74 375 842F 5' - AATGGCACTCTCCTACCCAG - 3'
UVRAG 74 376 425R 5' - AGGTGACTCGAGCTGTAAGCTC - 3'
(iii) *UVRAG* 74 376 320F 5' - CAGTGGGCCAAAGTTAGA - 3'
UVRAG 74 376 860R 5' - TTCATCTAATCTGAGGGCCG -3'
(iv) *UVRAG* 74 376 238 F 5' - AGCTGTGACGTTCCATCTCTTC - 3'
UVRAG 74 376 502 R 5' - GTCCAGTTGTCATTCCCGCATT - 3'

tcatgaaatccttctcactgcctcactgagcatgggcaagtcatgtaaattctctgccagatgaggatgcctgcttttcaaagctaaaataaga
attgaattgaatactaaaacaagaatgagattatggatagcccttcagaaatatgaagttagttatatccttcacatgttttcttgctagatc
cagcctgtccagttacttcacctcttccccataggacagggtttctgttccctgccctgccttgggccctcttctctaaatatattgcagtgtg
tctgtatcctgctggaaggaagctcccaaagttggacccacatacctcatgttgtttttgcattgagacactgcaataactattgtctaagat
tgattcttatctatttttattagctcagacctaagtatgctgttccctgtctctg**gatccagaaggcttacactgag**ttctgaagtccaaagta
acttcttgttttgtttctcttctagTGACAGACATCACACCTCCAGTGCAATCCCTGTTCCTAAGAGACAAAGCTCC
ATATTTGGGGGTGCAGATGTAGGCTTCTCTGGGGGGATCCCTTCACCAGACAAAGGACATCGAAAACGGG
CCAGCTCTGAGAATGAGAGACTTCAGTACAAAACCCCTCCTCCCAGTTACAACTCAGCATTAGCCCAGCCT
GTGACCACCGTCCCCTCCATGGGAGAGACCGAGAGAAAGATAACATCTCTATCCTCCTCCTTGGATACCTC
CTTGGACTTCTCCAAAGAAAACAAGAAAAAAGGAGAGGATCTAGTTGGCAGCTTAAACGGAGGCCACGC
GAATGTGCACCCTAGCCAAGAACAAGGAGAAGCCCTCTCCGGGCACCGGGCCACAGTC**AATGGCACTCTCC**
TACCCAGCGAGCAGGCCGGGTCCGCCAGTGTCCAGCTTCCAGGCGAGTTCCACCCAGTCTCAGAAGCTGAG
CTCTGCTGTACTGTGGAGCAAG**CAGAAGAAATCATCGGGCTGGA**AGCCACAGGTTTCGCCTCAGGTGATC
AGCTAGAAGCATTTAACTGCATCCCAGTGGACAGTGCTGTGGCAGTAGAGTGTGACGAACAAGTTCTGG
GAGAATTTGAAGAGTTCTCCCGAAGGATCTATGCACTGAATGAAAACGTATCCAGCTTCCGCCGGCCGCG
CAGGAGTTCCGATAAGTGAAGTGAGCAGGTCAACAGTAGGACTGGGGCAGAAGCTCTGCCTAAAATGAA
GTGAAAGCTGCACTTAACCCTTTGTGATAATGATGACACAAAATGAATATTAATGGAGGATATTCCTCG
GAAAAACAGACTTTGGGAATGAAGGAGGGACTCAGGATCATTGTTAT**CAGTGGGCCAAAGTTAGA**TTT
TGCTTTCAAGATTTGCTTTTCGGGCCTGATGATTTTAAAGCAAAAATCACCCTCTAGTTGAAA**GAGCTT**
ACAGCTCGAGTCACCTTTTAGCTATTTGTCTGCTTTTTATTTACCCTTGTATGTTATCCTCAGAGGGAAG
ATGATAATATATAAATAATATAATGAACACACCCTTAGTTTCTCATAAGCATTTGCCCTCACCATGGTT
TATAAAACTTTGGGAAAACGGAATATTCAGAAATAGGTTTCCGCCATGTACTGAAAGGTCTGTGGCCAT
CTGTGAGGTAGATGAAGAAGCAGCATAGTGGTCTCCTTACATCTAGGCCTAACTGTCCCTCTTCCTGCCCC
CGGGTACCACAGTCCACCTTTAGACCCTACTGTCGCCCCATCTTCTCCGTGGATGGGCCATGCGTCCTGAA
AACAGGACATCAGATTCACTGGTTCTGTAACCCAGT**AGCTGTGACGTTCCATCTCTTC**TAACCAGCCATG
GCCTTCCCCTCCTCTGCCATACCCTTAATG**CGGCCCTCAGATTAGATGAA**AAACTTGCTCCTGGTGGATCC
CAAGGGACCCTCAAGGACCTCGAGGTTACTGCAGTCAGATGCCATCTCATCCCCTGTGGGGGCCAAAGTTT
TTATGTGGGCAGATGCTGTGGTCAGGAACTAGGCATGCTTTCTGGCAATGCACTCACCAGACAAAAATCC
CTTGATGTAAATCCCATGTTAATTTATTAAATTTCAGTCAGAAGGTCAGCATTTACATGACAGAATGTA
TGTAGAGAGTTGGGGTGTCTGGTAGGCAAACTGCAAGGCAGTTGAGATAGTTGGATTAAGAGGCTAGAC
GAGACATAGAATACTATTGGTATGTGTGCAATTTCATGAATATTAAATTATGTTTCGAA**GTCCAGTTG**
TCATTCCCGCATTCAGATTTCATTTGCTGTTGCTTTATACGTTACGTACCCAAGGACATTGCCTCAGGGT
TGCAAACTCTTTAAAGGAAAATTTATCCATATATCCATGTATTATATAGAAGAATAAAAATTGAGTTT
ACTTC

Figure 15 *UVRAG* Exon 15 and positions of primers selected for PCR and sequencing reaction

Forward primers are depicted in red text and reverse in blue text

The PCR products were QIAquicked and eluted with 40μL ddH_2O. Sequencing was achieved using 2μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.6.6.3 Exon 15 (iii) Exonic Primers - Amplification and sequencing

Forward Primer	74 376 320F	5'-CAGTGGGCCAAAGTTAGA-3'
Reverse Primer	74 376 860R	5'-CGGCCCTCAGATTAGATGAA-3'

An annealing temperature of 52°C with 30 cycles amplified this portion of exon 15 from DNA of an unaffected and affected patient. PCR products incorporating Q-solution were electrophorised at 100V for 1hr. These PCR products were QIAquicked and eluted with 40μL ddH_2O. PCR products were sequenced using 3μL template DNA. Again sequencing analysis was unsuccessful, and despite numerous attempts this portion of exon 15 was unable to be excluded as a causative sequence for HFM. However it was noted that there were no differences apparent on gel chromatography therefore suggesting no obvious deletions or insertions on this portion of the exon.

4.6.6.4 Exon 15 (iv) Exonic Primers - Amplification and sequencing

Forward Primer	74 376 238F	5'- AGCTGTGACGTTCCATCTCTTC -3'
Reverse Primer	74 376 502R	5'-AATGCGGGAATGACAACTGGAC-3'

An annealing temperature of 61°C with 30 cycles amplified this portion of exon 15 from DNA of an unaffected and affected patient. PCR products incorporating Q-solution were electrophorised at 100V for 1hr. These PCR products were QIAquicked and eluted with 40μL ddH_2O. PCR products were

sequenced using 3μL template DNA. Again sequencing analysis was unsuccessful, and despite numerous attempts this portion of exon 15 was unable to be excluded as a causative sequence for HFM. However it was noted that there were no differences apparent on gel chromatography therefore suggesting no obvious deletions or insertions on this portion of the exon.

To summarize, 14.5 out of 15 exons for this gene was sequenced, and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank. Therefore sequencing of this gene has reduced the region to be assessed for future studies of chromosome 11 HFM mutation analysis.

4.7 PCR amplification of *ARIX* and sequencing results

4.7.1 Exon 1 - Amplification and sequencing

Forward Primer	75 883 245F	5'-ccacacctctgagccctaagacg-3'
Reverse Primer	75 883 905R	5'-cattggagggtctggccaaggca-3'

The PCR was thermocycled using an annealing temperature of 69°C for 30 cycles. The annealing temperature was reduced to 68°C to improve quality of amplification but with little success. New primers were designed and ordered.

Forward Primer	75 883 245F	5'-ccacacctctgagccctaaga-3'
Reverse Primer	75 883 915R	5'-tctggccaaggcaggaat-3'

The PCR was thermocycled using an annealing temperature of 56°C for 30 cycles, without success. A HotStar was prepared and ran at 50°C for 30 cycles. A HotStar PCR reaction was prepared using control (unaffected,

unrelated male and female) DNA. This too was unsuccessful, implying that the primers were not annealing successfully. Primers were redesigned in the exonic sequence at the beginning and end of the exon. Those primers were manually designed and then compared with on-line databases as to specificity of region to be sequenced. Moreover, their design complied with suggested parameters for success (See section 3.9.1.2).

Forward Primer	71 676 429F	5'-CTGAGTGCGGCCGCGAC-3'
Reverse Primer	71 676 808R	5'-GCGCTCACCTGCCGAGTA-3'

The PCR was thermocycled using an annealing temperature of 55°C for 30 cycles, without success. Control DNA and the affected DNA were used. The control DNA yielded multiple bands while the affected DNA did not amplify. Despite numerous attempts this exon was unable to be excluded as a candidate for chromosome 11 HFM.

4.7.2 Exon 2 - Amplification and sequencing

Forward Primer	75 886 137F	5'-ggctgccgggaccaagacga-3'
Reverse Primer	75 886 575R	5'-gccttcgggctgcatctgcc-3'

The PCR was thermocycled using an annealing temperature of 63°C for 30 cycles. Multiple bands were evident but faint. Two separate PCR reactions were prepared and run, one at 62°C with 30 cycles, and the other at 64°C with 30 cycles. The former proved to be more successful and a sequencing reaction using 4μL template DNA was prepared and run. This was not accurate enough to definitively compare the affected sequence to the sequence available on GenBank. Consequently primers were redesigned and ordered.

Forward Primer	75 886 107F	5'-ttgagtaatagggaggacgct-3'
Reverse Primer	75 886 709R	5'-tcggaactttctgccctaaga-3'

The PCR was thermocycled using an annealing temperature of 57°C for 30 cycles. PCR reaction without Q-solution was superior in its amplification however all PCR products were faint. A new PCR was prepared and run at 52°C for 30 cycles. This was unsuccessful and the annealing temperature was dropped to 50°C for 30 cycles. A HotStar PCR reaction was prepared,. Again, efforts were fruitless. Control DNA was used to prepare two separate reactions, a standard PCR reaction and a HotStar PCR reaction, both of which were run at 50°C for 30 cycles. This too was unsuccessful. Exonic primers were redesigned and ordered. These primers were manually designed and then compared to on-line databases as to specificity of region to be sequenced. Moreover their design complied with suggested parameters for success (See section 3.9.1.2).

Forward Primer	71 679 299F	5'-TGCCCTACAAGTTCTTC-3'
Reverse Primer	71 679 486R	5'-CTGCACGCGAGCCTCA-3'

The PCR was successfully thermocycled using an annealing temperature of 48°C for 30 cycles. PCR reaction without Q-solution was superior in its amplification. PCR products were QIAquicked and eluted with 40μL ddH_2O. Sequencing however was not clean enough to compare with sequences available on GenBank and therefore this region cannot be excluded as a candidate region for HFM. The PCR reaction amplified by electrophoresis displayed no size difference and hence suggests no obvious changes (see Appendix 17 for chromatograms).

4.7.3 Exon 3 – Amplification and sequencing

Forward Primer	75 887 278F	5'-tccgaaccaggatctcactcgag-3'
Reverse Primer	75 887 995R	5'-gagtggccctgacttggtctcc-3'

The PCR was thermocycled using an annealing temperature of 67°C with 30 cycles. The annealing temperature was reduced to 65°C without success. New primers were ordered.

Forward Primer	75 887 280F	5'-cgaaccaggatctcactcga-3'
Reverse Primer	75 887 982R	5'-ttggtctccaaagttgggga-3'

Two PCRs were thermocycled using annealing temperatures of 65°C and 66°C for 30 cycles. This proved unsuccessful so a PCR was run at 50°C for 30 cycles. The PCR reaction without Q-solution added was superior in its amplification although multiple bands were evident. A HotStar PCR was prepared and run at 50°C for 30 cycles. Given the difficulties with the first and second exon of this gene, exonic primers were designed and ordered.

Forward Primer	71 680 387F	5'-AGGTCTGGTTCCAGAACCG-3'
Reverse Primer	71 680 843R	5'-GCAGCTAGAAGAGATTGGTC-3'

The PCR was thermocycled using an annealing temperature of 55°C for 30 cycles successfully. PCR reaction with Q-solution added was superior in its amplification. The PCR products were QIAquicked and eluted with 25μL ddH_2O. Similarly to exon 2, sequencing was unsuccessful and we were unable to analyze this exon accurately enough to definitively claim this exon does not contain a mutation.

To summarize, sequencing was not definitive with the three exons of ARIX. Therefore this project is unable to exclude ARIX as a candidate gene for chromosome 11 HFM. It is important to note however that PCR

reactions amplified by electrophoresis displayed no size difference and hence suggests no obvious changes.

4.8 PCR amplification of GARP and sequencing results

4.8.1 Exon 1 – Amplification and sequencing

Forward Primer	74 891 069F	5'-taagtcagctgaggccgagag-3'
Reverse Primer	74 891 257R	5'-ctccagcacatgctgagccg-3'

An annealing temperature of 61°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR reactions with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL elution buffer. Sequencing was achieved using 2μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.8.2 Exon 2 – Amplification and sequencing

Forward Primer	74 894 838F	5'-ttccagctccagccgtgctc-3'
Reverse Primer	74 895 150R	5'-cctgactcttctaacttctggc-3'

An annealing temperature of 61°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL elution buffer. Sequencing was achieved using 4μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.8.3 Exon 3 – Amplification and sequencing

4.8.3.1 Exon 3 (i) Exonic primers

Forward primer	74 899 173F	5'-ccttcatcgagttccttcctt-3'
Reverse primer	74 899 598R	5'-TATGCAGGTCAAGCTGCTCCA-3'

An annealing temperature of 51°C with 30 cycles was unsuccessful in amplification of this exon. A Hot Star PCR using control (unrelated/unaffected male and female) DNA was thermocycled using an annealing temperature of 50°C with 30 cycles. This identified the primers as being unable to anneal and thus new exonic primers were chosen that were more complementary to sequences at the beginning and end of the exon.

Forward primer	74 891 334F	5'-gGTGGACAAGAAGGTCTCGTG-3'
Reverse primer	74 891 794R	5'-CTATGCAGGTCAAGCTGCTCC-3'

An annealing temperature of 61°C with 30 cycles was unsuccessful and despite numerous attempts this exon was unable to be excluded as a causative sequence for HFM. However it was noted that there were no differences apparent on gel chromatography suggesting no obvious deletions or insertions in this portion of the exon.

4.8.3.2 Exon 3 (ii) Exonic primers

Forward Primer	74 899 713F	5'-CCTTCATCGAGTTCCTTCCTT-3'
Reverse Primer	74 900 289R	5'-TATGCAGGTCAAGCTGCTCCA-3'

An annealing temperature of 58°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 40μL ddH_2O. Sequencing

was achieved using 3μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank

4.8.3.3 Exon 3 (iii) Exonic primer

Forward Primer	74 900 181F	5'-AGGACAGCAAGGGCATCC-3'
Reverse Primer	74 900 809R	5'-CCTCCAAGGAGGCCTCC-3'

An annealing temperature of 53°C and 55°C with 30 cycles were both unsuccessful in amplification of this exon. Alternative exonic primers were ordered.

Forward primer	74 900 181F	5'-AGGACAGCAAGGGCATCCA-3'
Reverse primer	74 901 326R	5'-TCCCGGCTTCTTTAGGCTTTA-3'

An annealing temperature of 52°C with 30 cycles was unsuccessful and despite numerous attempts this exon was unable to be excluded as a causative sequence for HFM. However it was noted that there were no differences apparent between gel chromatography therefore suggesting no obvious deletions or insertions on this portion of the exon.

4.8.3.4 Exon 3 (iv) Exonic primers

Forward Primer	74 901 157F	5'-GAGCCACGTGCGTCCTGAGGAC-3'
Reverse Primer	74 901 764R	5'-CGGGGCCTGCCGAGCTCTGGA-3'

An annealing temperature of 69°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 40μL ddH_2O. Sequencing was achieved using 3μL template DNA and no mutations were detected in

affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.8.3.5 Exon 3 (v) Exonic primers

Forward primer	74 901 457F	5'-CCTGCTGCATCAGTGGGTGA-3'
Reverse primer	74 902 423R	5'-TTCCTGTCCACTTAACGGCCT-3'

An annealing temperature of 59°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 40μL ddH_2O. Sequencing was achieved using 3μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.8.3.6 Exon 3 (vi) Exonic primers

Forward Primer	74 902 197F	5'-TAGGAGAGAGTGCTGCAGAG-3'
Reverse Primer	74 902 796R	5'-CAGACACAAGGCTTGGATTCA-3'

An annealing temperature of 57°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL ddH_2O. Sequencing was achieved using 2μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.8.3.7 Exon 3 (vii) Exonic primers

Forward Primer 74 902 297F 5'-GCACCCAGCTTGGCAGATGTG-3'

Reverse Primer 74 902 898R 5'-AGAGAGGACATCACTCTGGTCC-3'

An annealing temperature of 63°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL ddH_2O. Sequencing was achieved using 2μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.8.3.8 Exon 3 (viii) Exonic primers

Forward Primer 74 902 782F 5'-CTGAGGCTTAGGAAGAGAATG-3'

Reverse Primer 74 903 490R 5'-tgccatgatgattgaacgacc-3'

An annealing temperature of 57°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL ddH_2O. Sequencing was achieved using 2μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

To summarize, the first two exons and the majority of exon three (75%) for this gene was sequenced, and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to

each other and GenBank. Therefore sequencing of this gene further reduced the region to be assessed for chromosome 11 HFM linkage.

4.9 PCR amplification of *CLNS1A* and sequencing results

4.9.1 Exon 1 – Amplification and sequencing

Forward Primer	75 889 781F	5'-tccacacgttcttagccgacctc-3'
Reverse Primer	75 890 255R	5'-gagccttccacccgctacaggta-3'

An annealing temperature of 67°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 50μL ddH_2O. Sequencing was achieved using 3μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.9.2 Exon 2 – Amplification and sequencing

Forward primer	75 897 695F	5'-gtggtcttacatgaggatttac-3'
Reverse primer	75 898 022R	5'-gattacaggcgtgagccac-3'

An annealing temperature of 57°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL ddH_2O. Sequencing was achieved using 4μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.9.3 Exon 3 – Amplification and sequencing

Forward Primer	75 901 707F	5'-ctaggatgctcactgtataatc-3'
Reverse Primer	75 902 094R	5'-caatagaatggaggtaatggtg-3'

An annealing temperature of 57°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 50μL ddH_2O. Sequencing was achieved using 3μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank.

4.9.4 Exon 4 – Amplification and sequencing

Forward Primer	75 902 472F	5'-gcttcaaccgctttcaag-3'
Reverse Primer	75 902 813R	5'-gaactataatcctctaccc-3'

An annealing temperature of 51°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL ddH_2O. Sequencing was achieved using 2μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.9.5 Exon 5 – Amplification and sequencing

Forward Primer	75 904 858F	5'-gtgccaccacgcctggctaa-3'
Reverse Primer	75 905 214R	5'-gtgctcatctgactgttcacac-3'

An annealing temperature of 61°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products with Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 35μL ddH_2O. Sequencing was achieved using 3μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.9.6 Exon 6 – Amplification and sequencing

Forward Primer	75 907 890F	5'-cactagcttccttctggtaga-3'
Reverse Primer	75 908 195R	5'-atgccactgcactgtagcct-3'

An annealing temperature of 57°C with 30 cycles successfully amplified this exon from DNA of an unaffected and affected patient. PCR products without Q-solution were electrophorised at 100V for 1 hr. These PCR products were QIAquicked and eluted with 30μL ddH_2O. Sequencing was achieved using 3μL template DNA and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank (see Appendix 17 for chromatograms).

4.9.7 Exon 7 – Amplification and sequencing

4.9.7.1 Exonic Primer (i)

Forward Primer	75 910 899RF	5'-ggcactcgtgactaggcatt-3'
Reverse Primer	75 911 374R	5'-TGGTAAAACATGGCGAAAGG-3'

An annealing temperature of 55°C with 30 cycles was unsuccessful and despite numerous attempts this exon was unable to be excluded as a causative sequence for HFM. However it was noted that there were no

differences apparent between gel chromatography therefore suggesting no obvious deletions or insertions on this portion of the exon.

4.9.7.2 Exonic Primer (ii)

Forward Primer	75 911 301F	5'-GCACAGGAGCTTGGTAG-3'
Reverse Primer	75 911 691R	5'-gccttactaaatacttgcc-3'

An annealing temperature of 49°C with 30 cycles was unsuccessful and despite numerous attempts this exon was unable to be excluded as a causative sequence for HFM. However it was noted that there were no differences apparent between gel chromatography therefore suggesting no obvious deletions or insertions on this portion of the exon.

To summarize, 6 out of 7 exons for this gene was sequenced, and no mutations were detected in affected or unaffected DNA when sequences were assessed relative to each other and GenBank. Therefore sequencing of this gene has reduced the region to be assessed for chromosome 11 HFM linkage.

In those genes sequenced, no sequence variations were detected in exonic sequences of affected or unaffected individuals (Appendix 18).

4.10 Summary

To summarize, although no mutations were found, sequencing of these genes has contributed significantly in reducing the region to be analyzed by 355,995 bases. Unfortunately no further results could be achieved with the time available for this project, however, publication of these results (paper

in preparation) will markedly improve available data for HFM. These sequencing results, together with marker analysis, reduced the overall region to be analyzed by 355,995 bp.

Chapter 5: Discussion

5.0 Discussion

The overall purpose of this project was to explore the underlying pathogenesis of HFM, investigate treatment and management options, identify likely candidate genes and screen candidate genes for mutation(s) causing the form of HFM segregating on chromosome 11 in a West Australian family. Outcomes of this project have significantly advanced investigations into the molecular pathology of hemifacial microsomia and will provide those affected with chromosome 11 HFM some hope that a genetic mutation may be found in the near future which could lead to better diagnosis, prognosis, prevention and treatment.

Hemifacial microsomia (HFM) is a congenital developmental disorder involving malformation of the first and second branchial arch derivatives. Variable axial asymmetry is the characteristic feature of this disorder with craniofacial malformations as the phenotypic manifestation of the disease. As any derivative of the branchial arches may be affected, the clinical picture varies considerably.

Embryologically, asymmetry is a relatively common occurrence in nature. Examination of frog embryos has indicated that a cascade of signaling molecules is responsible for the generation of asymmetry, beginning with the asymmetric expression of *Shh* along the dorsal-ventral axis and FGF and RA along the rostro-caudal axis and forms part of the normal specification and formation of axes, crucial developmental events, for determining the orientation of the body plan. However, left-right asymmetry is essential for formation of the two sets of a symmetrical body

plan and is generated by signaling molecules and genes (*lefty* and *nodal*). Asymmetry across the left-right axis leads to abnormalities which have fatal consequences if they severely affect organ and skeletal structures. HFM is a relatively mild manifestation of left-right asymmetry.

The aetiology of HFM has been subject to much debate and is yet to be sufficiently explained. Literature surrounding this phenomenon has suggested that HFM results from a culmination of causal factors, including defective genes, teratogenic substances or vascular anomalies. This leads to the conclusion that within the HFM spectrum there is an amalgamation of both incidental and inherited disorders without any true prediction of the pathogenesis of the syndrome that results. Alternatively, variable expression of an autosomal dominant gene may explain the reported HFM pedigrees as it could also account for the range of deformities associated with HFM. Numerous chromosomal abnormalities have been associated with HFM and studies involving both mutant and transgenic mice have been discovered that appear to be models of HFM.

This project, based on the anticipation described in the family under investigation, and the autosomal dominant nature of this disorder, identified and sequenced four candidate genes at the chromosome 11 linkage locus, previously identified by Dr David Chandler (2001). Despite unsuccessful attempts at mutational analysis both here and previously, performing this study has greatly added to the general understanding of this disease and its phenotype, and provides a basis for further studies

aimed at better diagnosis, treatment and identification of causative genes and other factors.

5.1 Diagnosis recommendations

HFM is typically not diagnosed until late infancy or early childhood, as recognizable features such as facial asymmetry are subtle, in newborn infants. Accurate facial anomaly assessment is essential for effective diagnosis, classification and treatment of HFM. The two most common classification systems adopted to analyse HFM are the SAT and OMENS classification systems. Alternative systems, including the Pruzansky classification system, Teconi and Hall's phenotypic classification system and Munro and Lauritzen's system were alternative classification systems devised to guide treatment planning and diagnosis of HFM, although not as widely accepted as the SAT and OMENS classification system. This author, through review of the classification systems determined that SAT and OMENS are the two classification systems of preference however individual phenotypes must be taken into account to support accurate treatment and management. A combination of the two is recommended for future studies.

5.2 Treatment and Management Suggestions

Treatment has historically concerned the jaw or ear, the choice being dependant on the patient, family concerns and the surgeon's expertise and experience. Every facial structure affected in patients with HFM shows evidence of bilaterally diminished growth potential. This is a key feature of HFM and worthy of consideration in treatment regimes. Giving

consideration to the fact that the inborn morphogenetic error cannot be treated, the more appropriate approach is to create an environment where normal facial growth is encouraged and secondary distortion is minimised (Kearns, Padwa, Mulliken & Kaban, 2000; Murray, Kaban & Mulliken, 1984). Determining a patient's degree of skeletal deformation is one of the first steps in treating a patient with HFM. The severity of the defect can be appreciated and the progression rate predicted through the investigation of mandibular reconfiguration. This project has contributed by providing an overall literature review on treatment strategies, as well as comparing the benefits and risks associated with early or late surgical intervention and pulication of this material as a result may assist with diagnosis in the future.

Reconstruction of the mandible is one of the key elements in the skeletal rehabilitation of patients with HFM (Polley, Figueroa, Liou & Cohen, 1996). Aforementioned, severe grades of HFM present with significant dysplasia of both skeletal and soft tissues, posing challenging problems for the surgical reconstructive team (Polley & Figueroa, 1997). Central skeletal deformity usually centers on the temporomandibular region. A variety of orthodontic regimes have been recommended for treating HFM, including functional appliances, used singularly, or in conjunction with surgery (Chate, 1995).

The question of facial growth potential in HFM is important when developing treatment strategies for patients (Cobourne, 2000). This is central to timing of surgical intervention which has been subject to much

debate (Cousley & Calvert, 1997; Kaban, Moses & Mulliken, 1988; Murray, Kaban & Mulliken, 1984). In general, the treatment plan is considered most beneficial if formulated once the soft-tissue and skeletal deformities are classified and the mandibular type and anatomy are appropriately analysed (Cousley & Calvert, 1997). To date, there is no universal protocol to be followed when treating patients with HFM.

Despite the deformation of other structures aside from facial asymmetry it is the face that is considered of prime aesthetic importance. Hence, many treatment reviews have concentrated on improving the symmetry of the mandible and its associated deformities.

5.2.1 Arguments Supporting Early Intervention and Recommendations

Based on this author's review of the literature and treatment, it is suggested that early intervention is recommended in terms of psychological benefit to the patient however it is apparent that a long-term management plan, possibly involving further surgical intervention, may be required as skeletal structures develop and deformation becomes more apparent.

Through clinical observation Murray, Kaban and Mulliken (1984) deemed the mandible responsible for growth restriction of the maxilla, zygoma and orbits. Early mandibular elongation therefore provides more symmetrical growth of the midface and subsequent psychological benefit to the young patient (Kaban, Moses & Mulliken, 1988). This author deemed the most beneficial approach to treatment involves mandibular repositioning in a more physiological position which will trigger the growth potential of

adjacent structures. Results of this strategy have included minimised secondary deformity, improved function and appearance of the mandible and a greater benefit of the patient's skeletal and psychological growth (Kearns, Padwa, Mulliken & Kaban, 2000). As a result of continual asymmetric growth, the clinical defects worsen, even though the morphogenetic damage within the first and second branchial arches occurs *in utero*. The earliest skeletal manifestation characteristic of HFM is mandibular hypoplasia which compromises otherwise normal downward growth of the maxilla. Thus mandibular distortion worsens with asymmetric skeletal growth and can cause secondary deformation of the mid-face.

Operations involving bone or cartilage transplantation can be used to enhance facial symmetry. Despite the aesthetically pleasing results, transplanted bone is not expected to grow or improve the facial contour (Polley & Figueroa, 1999; Kazanjian, 1939). Mechanical devices used to control normal mandible movements must take into consideration the patients age and the substitution of temporary teeth with permanent ones. Following on from the proposition that the affected side of a patient exhibits diminished growth potential, it can be argued that with time the skeletal defect would progressively worsen. A more extensive mandibular procedure, maxillary osteotomy, is frequently required by adult HFM patients, proceeded by several bone grafts, whereas in the developing child the operative correction can be limited to the mandible.

Those children suffering from severe HFM are most likely to benefit psychologically from early intervention and treatment (Poole, 1989). Reports

are insufficient on the long-term result of major facial osteotomies, although several authors have observed the preservation of symmetry up to 6 years post-op. Moreover appropriately timed mandibular construction and/or elongation in children with HFM is safe and effective (Polley & Figueroa, 1997). The optimized facial growth eventuates from the reduction in secondary deformity on the affected side.

There are several advantages associated with advocating early skeletal correction in children with HFM. Primarily, the mandibular procedure is easier for the patient and the surgeon. Secondly, children seem to benefit from an improved self-image while growing, thereby diminishing the psychological trauma of their asymmetrical deformity (Kaban, Moses & Mullike, 1988; Stark & Saunders, 1962). While difficult to measure, the psychological benefits are imperative to both the patient and their family. Finally, in early operative intervention, complications are few.

Consequently, early intervention/operation is advisable to assist in maintaining normal soft tissue development on an affected side, since the bone defect cannot be completely or permanently rectified at the operative stage (Stark & Saunders, 1962). However, those who advocated early surgery recognised that further definitive correction was required at the end of growth (Poole, 1989; Vargervik, 1983a,b).

5.2.2 Arguments Supporting Late Intervention and suggestions

Surgery prior to the conclusion of adolescence has been criticised because of morbidity; potential growth disruption from the operation and

subsequent scarring; the underlying asymmetry of the soft tissue functional matrix; the effects on patient compliance and the need for post-growth corrections (Renzi, Carboni, Perugini & Becellim, 2002; Poswillo, 1974). Consequently it has been argued that definitive soft tissue reconstruction should be delayed until after the completion of both growth and skeletal surgery (Cousley & Calvert, 1997). Ultimately it can be reasoned that complete correction can be successfully carried out in the adult patient with one major skeletal operation (Kaban, Moses & Mulliken, 1988).

Kaban, Moses and Mulliken (1988) acknowledged concerns regarding surgical intervention in children, primarily that young patients severely affected by HFM would require numerous operations throughout the course of facial growth. Therefore until full growth is acquired, changes in the contour of the face would be constantly taking place (Kazanjian, 1939). This author also acknowledged that premature surgical treatment could accentuate a deformity by decreasing already compromised growth potential. This would be compounded if multiple procedures were required due to repeat tissue damage and scar formation. Similarly Murray and colleagues (1984) advocated late intervention in patients with HFM.

There are two opposing views in the literature concerning the progressive nature of facial asymmetry in HFM patients. Both Rune *et al.* (1981) and Polley *et al.* (1997) emphasized that asymmetry did not evolve in HFM patients and that the deformity changes in proportion to the child's overall size but that it did not become increasingly severe with time. Therefore, it is recommended that the facial asymmetry be addressed, at

the completion of growth, by surgical correction of the end-stage deformity. Similarly Rune *et al.* (1981) and Polley *et al.* (1997) concluded that growth on the affected side parallels that of the non-affected side with the degree of mandibular asymmetry remaining relatively constant throughout craniofacial development. However, several authors suggest alternatively that mandibular asymmetry is progressive and patients exhibit increasing vertical mandibular asymmetry with growth (Kearns, Padwa, Mulliken & Kaban, 2000; Padwa, Mulliken, Maghen & Kaban, 1998; Polley & Figueroa, 1997; Polley, Figueroa, Liou & Cohen, 1996; Rune *et al.*, 1981).

HFM patients are best treated in multidisciplinary centres by competent specialists with the necessary experience and skills. This effectively involves individual patient assessment in relation to reconstructive surgery (Guichard & Arnaud, 2001), orthodontics (Chate, 1995), otolaryngology, audiology, speech and language therapy, opthalmology, paediatrics, genetics and psychology (Poole, 1989).

Prevention of craniofacial malformations is unlikely, based on knowledge of their etiology or pathogenesis, although identification of causal mechanisms may play a role in the future clinical management. The ultimate goal of treatment of HFM is to improve facial symmetry and mandibular function upon completion of craniofacial growth. More specifically, the objectives of treatment include optimizing function (hearing, speech, mastication, swallowing and respiration) and aesthetics (symmetry and balance) (Cousley & Calvert, 1997). The need to recognise hearing loss early and provide for a sensory prosthesis and auditory training

where indicated, is a matter of early priority in patient care (Figueroa & Pruzansky, 1982). However, recognition that a child's psychosocial confidence may suffer from an obvious abnormality is important (Cousley & Calvert, 1997; Dryland, 1996). The decision whether to perform surgery or not should only be dictated by the possibility of an optimal long-term result.

Despite the varying methods of treatment available diagnosis and treatment of HFM will be greatly enhanced by discoveries of genetic mutations linked to the disease. Treatment and management of HFM essentially relies on an increased understanding of the pathogenesis of HFM. Insufficient numbers of severe HFM patients have been observed over the long-term to growth completion to predict with certainty how many will require maxillary osteotomy. This is an important consideration when analysing treatment and management modalities. Additionally, there is a lack of understanding surrounding the ramifications of embryonic neural crest cell destruction on structure development and long-term growth potential of the affected derivatives of the first and second branchial arches.

Additional assessment of the probands' family members for defects associated with the HFM complex is essential. Chromosome analysis and genetic counselling can be offered, where appropriate (Cousley & Calvert, 1997; Epstein, Curry, Packman, Sherman & Hall, 1979).

Previously, linkage studies were hampered by the lack of large families with HFM, despite the identification of numerous small families. Non-penetrance and the difficulties associated with classifying HFM make it

both unnecessary and unwise to combine small families for the purpose of linkage studies.

5.3 Mutation analysis in this thesis

The identification and co-operation of a large West Australian family with HFM has allowed the phenotypic and genotypic study of the disorder presented in this thesis. Phenotypic analysis demonstrates that Goldenhar and HFM symptoms are both present in this family, indicating that these two disorders may be caused by mutations in a single gene (Jacobsson & Granstrom, 1996; Kaye *et al.*, 1992; Yanagihara, Yanagihara & Kabasawa, 1979). Although large, this family does not have sufficient affected individuals to provide significant linkage in an affected-person only analysis. It was however large enough to give significant linkage using all members of the pedigree and to exclude a large portion of the genome as being in linkage with HFM (Chandler, 2001; this thesis).

Professor Robin Winter's group in London, has performed a genome screen on another large family with HFM. This yielded a Lod score of 3.00 on chromosome 14 in the vicinity of markers D14S267 and D14S987. Typing of the Western Australian family for these markers showed that the West Australian family does not link to this region. Conversely, work on the English family for markers in the chromosome 11 linkage region for the West Australian family excluded the English family from being linked to a region on chromosome 11. This confirmed the probability of different genetic loci associated with these disorders in two separate families and heightens the possibility of genetic association with an autosomal dominant

disease gene. Moreover, these results provide a clue to the suspected genetic heterogeneity for this disorder, based on phenotypic heterogeneity thereby emphasising the importance of careful and accurate phenotypic classification.

Once a linkage had been established and phenotype carefully assed, this project adopted a novel strategy known as 'genome mining' (Houle, Cadigan, Henry, Pinnamaneni & Lundahl, 2000) to identify hemifacial microsomia candidate genes. This technique was successful and identified four genes, namely *UVRAG, ARIX, CLNS1A* and *GARP* from the chromosome 11 linkage region as being likely candidates for the West Australian family segregating HFM. All four genes were thoroughly investigated (for association with known HFM regions in patients and animal models) using several databases. Moreover, the majority of three genes were sequenced. Despite the implementation of numerous laboratory techniques sequencing of all for genes could not be completed. Nonetheless, this project has contributed to enhancing the understanding and assessment of HFM through the adoption of novel strategies; genome mining and bioinformatics. While data-mining and sequencing of these three genes in an affected and unaffected individual identified no differences from the published sequence, these results assist future studies by minimising the regions and genes required to be sequenced. Furthermore, results presented here exclude *UVRAG, CLNS1A* and *GARP* as candidates for the disease in this family and suggests further sequencing of *ARIX* as it is a strong candidate for future research and experimentation.

Within the chromosome 11 linkage region, there are 131 known genes. From genome mining studies performed for this study, we have reduced this list to the most promising candidate genes. In this way, genome mining further reduced the list of candidate genes and provides a useful resource for future studies aimed at identifying likely candidate genes for HFM.

5.3.1 Triplet repeat expansion

The observation that anticipation may be involved in this family in the inheritance of HFM prompted the investigation of triplet repeats for a disease causing expansion. Future examination of candidate genes should take into account the possibility of a perhaps small repeat expansion either in or near the gene as a causative mutation. A full list of repeat regions in or near genes is provided in Appendix 10 and is a useful resource for future genome mining studies of this region.

5.4 Contribution of this study to the treatment and diagnosis of dysmorphology

An important aim of this project was to add to the understanding and concepts and aid in decisions that optimise treatment. The identification of disease genes is crucial for the understanding and treatment of the disease. Moreover optimal management of HFM is dependant on appreciation of the complete genotype and phenotype. In this thesis, I provide a comprehensive appraisal of genetic and interdisciplinary approaches along with long-term coordinated treatment planning. Aforementioned, this author endorses a co-ordination of both early and late intervention and management. I also indicate effective treatment strategies specific to the

individual that can be formulated once the soft-tissue and skeletal deformities are classified. The use of bioinformatics to assess genetic information has also proven to be partially successful and provides a framework for future studies.

5.5 Future Research

Through identification of genes associated with HFM, it will be possible to accurately diagnose, treat and manage patients with HFM, providing them with optimal life expectancies and better quality of lie. Results presented in this thesis contribute to this cause.

It is implausible to test every gene in this linkage region for this project. Foremost, further familial studies are necessary in order to confirm this linkage region on chromosome 11 as associated with HFM and so as to refine the linkage region. From studies presented here, it becomes evident that the intricacy of the developmental system implies that many molecules will be involved in HOX code communication to the myriad of tissues and organs reliant on positional and patterning information. Through future studies on the interaction of nuclear proteins with the HOX transcription factors, a number of genes, including the zinc-finger genes, will prove to play a role in craniofacial development. Should these or related genes, be discovered in the linkage region previously described, they should become prime candidates and tested accordingly in the affected family to determine if they are in fact mutated. Furthermore, *ARIX* has been identified by this project as a prime candidate for chromosome 11 HFM.

Due to time restraints and lack of success in sequencing, this project has identified *ARIX* as a candidate for future testing.

5.6 Concluding Comments

A model of human development will in time be built through the discovery of inherited malformations. In turn this will assist in the understanding of causative factors of randomly occurring malformations. The knowledge and scope of dysmorphology has greatly expanded over the last decade through the identification and localization of disease genes. Every gene discovery is another stich in the intricate tapestry that is the development of the human body. This thesis has helped by contributing strategies and sequencing information that will assist in the search for genes associated with hemifacial microsomia.

REFERENCES

REFERENCES

Abdelhak, S., Kalatzis, V., Heilig, R, Caompain, S., Samson, D., Vincent, C., Weil, D., Cruaud, C., Sahly, I., Leibovici, M., Bitner-Glindzicz, M., Francis, M., Lacombe, D., Vigneron, J., Caharachon, R., Boven, K., Bedbeger, P., Van Regemorter, N., Weissenbach, J., Petit, C. (1997). A human homologue of the drosophila eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. Nature Genetics, 15, 157-164.

Basch, M.L, Garcia-Castro, M.I., & Bronner-Fraser, M. (2004). Molecular mechanisms of neural crest induction. Birth Defects Research, 72(2), 109-23.

Bassila, M.K., & Goldberg, R. (1989). The association of facial palsy and/or sensorineural hearing loss in patients with hemifacial microsomia. Cleft Palate Journal, 26(4) 287-291.

Bekri, S., Adelaide, J., Merscher, S., Grosgeorge, J., Caroli-Bosc, F., Perucca-Lostanlen, D., Kelley, P.M., Pebusque, M.J., Theillet, C., Birnbaum, D., Gaudray, P. (1997). Detailed map of a region commonly amplified at 11q13-q14 in human breast carcinoma. Cytogenetic Cell Genetics, 79(1-2), 125-131.

Bisgrove, B.W., & Yost, H.J. (2001). Classification of left-right patterning defects in zebrafish, mice and humans. American Journal of Medical Genetics, 101(4), 315-23.

Boles, D.J., Bodurtha, J., & Nance, W.E. (1987). Goldenhar complex in discordant monosygotic twins: a case report and review of the literature. American Journal of Medical Genetics, 28, 103-109.

Bonaventure, J. (1998). Skeletal Development in Human: a Model for the Study of Developmental Genes[on-line]. Available WWW: http://www.infobiogen.fr/services/chromcancer/IntroItems/GenDevelShortEngl.html [2003, August 6].

Brady, A.F., Winter, R.M., Wilson, L.C., Tatnall, F.M., Sheridan, R.J., & Garrett, C. (2002). Hemifacial microsomia, external auditory canal atresia, deafness and Mullerian anomalies associated with acro-osteolysis: a new autosomal recessive syndrome? Clinical Dysmorphologies, 11(3), 155-161.

Brook, J.D., McCurrach, M.E., Harle, et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3-prime end of a transcript encoding a protein kinase family member. Cell, 68, 799-808.

Burck, U. (1983). Genetic aspects of hemifacial microsomia. Human Genetics, 64, 291-296.

Buyse, G., De Greef, C., Raeymaekers, L., Droogmans, G., Nilius, B., & Eggermont, J. (1996). The ubiquitously expressed pI(Cln) protein forms homomeric complexes in vitro. Biochem, Bioophys, Research and Communication (218), 822-827.

Carvalho, G.J., Song, C.S., Vargervik, K., & Lalwani, A.K. (1999). Auditory and facial nerve dysfunction in patients with hemifacial microsomia. Archives of Otolaryngology Head and Neck Surgery, 125(2), 209-212.

Casey, H.D., Braddock, S.R., Haskins, R.C., Carey, J.C., & Morales, L.Jnr. (1996). Frontonasal malformation and the oculauriculovertebral spectrum: the oculoauriculofrontonasal syndrome. Cleft Palate - Craniofacial Journal, 33(6), 519-523.

Cavaliere, C.M., & Buchman, S.R. (2002). Mandibular distraction in the absence of an ascending ramus and condyle. Journal of Craniofacial Surgery, 13(4), 527-532.

Chandler, D. (2001). Molecular genetics of three dominant dysmorphologies. Doctoral dissertation, University of Western Australia, Perth, Western Australia.

Chate, R.A.C. (1995). The propellant unilateral magnetic appliance (PUMA): a new technique for hemifacial microsomia. European Journal of Orthodontics, 17, 263-271.

Chen, H., & Johnson, R.L. (2002). Interactions between dorsal-ventral patterning genes *lmx1b*, *engrailed-1*, and *wnt-7a* in the vertebrate limb. International Journal of Developmental Biology, 46, 937-941.

Cobourne, M.T. (2000). Construction for the modern head: current concepts in craniofacial development. Journal of Orthodontics,

Dec, 27(4), 307-14.

Cohen, M.M. Jnr. (1991). A critique of the OMENS classification of hemifacial microsomia. Cleft Palate - Craniofacial Journal, 28, 77.

Connor, J.M., & Fernandez, C. (1984). Genetic aspects of hemifacial microsomia. Human Genetics, 68, 349.

Converse, J.M., Coccaro, P.J., Becker, M., & Wood-Smith, D. (1973). On hemifacial microsomia; the first and second branchial arch syndrome. Plastic and Reconstructive Surgery, 51(3), 268-279.

Converse, J.M., Wood-Smith, D., McCarthy, J.G., Coccaro, P.J., & Becker, M.H. (1974). Bilateral facial microsomia: diagnosis, classification, treatment. Plastic and Reconstructive Surgery, 54(4), 413-423.

Cousley, R.R. (1993). A comparison of two classification systems for hemifacial microsomia. British Journal of Oral and Maxillofacial Surgery, 31(2), 78-82.

Cousley, R.R., & Calvert, M.L. (1997). Current concepts in the understanding and management of hemifacial microsomia. British Journal of Plastic Surgery, 50(7), 536-551.

Cousley, R., Naora, H., Yokoyama, M., Kimura, M., & Otani, H. (2002). Validity of the Hfm transgenic mouse as a model for hemifacial microsomia. Cleft Palate Craniofacial Journal, 39(1), 81-92.

Cousley, R.R., & Wilson, D.J. (1992). Hemifacial microsomia: developmental consequence of perturbation of the auriculofacial cartilage model?. American Journal of Medical Genetics, 42(4), 461-466.

Cummings, M.R. (2000). Human Heredity: Principles and Issues (5th ed.). Brooks/Cole, United States of America.

David, D.J., Mahatumarat, C., & Cooter, R.D. (1987). Hemifacial microsomia: a multisystem classification. Plastic and Reconstructive Surgery, 80(4), 525-533.

Dear, N., Matena, K., Vingron, M., & Boehm, T. (1997). A new subfamily of vertebrate calpains lacking a calmodulin-like domain: implications for calpain regulation and evolution. Genomics, 45, 175-184.

Dryland, K. (1996). Functional therapy in hemifacial microsomia: therapeutic protocol for growing children (Discussion). Journal of Oral and Maxillofacial Surgery, 54, 278-280.

Dupont, S., Hani, E.H., Cras-Meneur, C., De Matos, F., Lobbens, S., Lecoeur, C., Vaxillaire, M., Scharfmann, R. & Froguel, P. (n.d.) No Evidence for Linkage or for Diabetes-Associated Mutations in the Activin Type 2B Receptor Gene (ACVR2B) in French Patients with Mature-Onset Diabetes of the Young or Type 2 Diabetes. Diabetes Journal [on-line serial], 50(5). Available WWW: http://diabetes.diabetesjournal.org/cgi/content/full/50/5/1219 [2003, August 6].

Ekker, S.C. (1993). Determination of the Animal Body Plan; Embryonic Patterning; Transposable Elements for Gene Discovery in Vertebrates [on-line]. Available WWW [2003, August 6]: http://www.cbs.umn.edu/mcdbg/faculty/Ekker.html.

Epstein, C.J., Curry, C.J.R., Packman, S., Sherman, S., & Hall, B.D. (1979). Risk, communication and decision making in genetic counselling. Birth Defects, 25(5) 178-179.

Evans, K. L., Fantes, J., Simpson, C., Arveiler, B., Muir, W., Fletcher, J., van Heyningen, V., Steel, K. P., Brown, K.A., Brown, S.D.M., St. Clair, D., & Porteous, D. J. (1993). Human olfactory marker protein maps close to tyrosinase and is a candidate gene for Usher syndrome type I. Human Molecular Genetics, 2, 115-118.

Fassler, J., Nadel, C., Richardson, N., McEntyre, J., Schuler, G., McGinnis, Pongor, S., & Landsman, D. (2000). Blast information - NCBI [on-line]. Available WWW [2002] ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html.

Figueroa, A.A., & Friede, H. (1985). Craniovertebral malformations in hemifacial microsomia. Journal of Craniofacial Genetics in Developing Biology Supplement, 1, 167-178.

Figueroa, A.A., & Pruzansky, S. (1982). The external ear, mandible and other components of hemifacial microsomia. Journal of Maxillofacial Surgery, 10(4), 200-211.

Finnerty, J.R. (1994). A Case Study: Axial Patterning [on-line]. Available WWW: http://www.bu.edu/cecb/Faculty/finnerty.html [2003, August 6].

Fischer, C.E., & Prahl-Andersen, B. (1996). Hemifacial microsomia: a review. Ned Tijdschr Tandheelkd, 03(10), 392-395.

Gale, M. Jnr., Blakely, C.M., Hopkins, D.A., Melville, M.W., Wambach, M., Romano, P.R., Katze, M.G. (1998). Regulation of interferon-induced protein kinase PKR: modulation of P58(IPK) inhibitory function by a novel protein, P52(rIPK). Molecular Cell Biology, 18, 859-871.

Gavalas, A., Davenne, M., Lumsden, A., Chambon, P., & Rijli, F.M. (1997). Role of Hoxa-2 in axon pathfinding and rostral hindbrain patterning. *Development, 124* (3693-3702).

Gibson, J.N.A., Sillence, D.O., & Taylor, T.K.F. (1996). Abnormalities of the spine in Goldenhar's syndrome. Journal of Pediatric Orthopaedics, 16, 344-349.

Gilbert (1998). Hox Genes and Nematode Development [on-line]. Available WWW: http//zygote.swarthmore.edu/cyto6.html [2003, August 6].

Gilbert, S.F. (2000). Developmental Biology, 6th Edition, Ed. Carol Wigg, Sinauer Associates Inc, Sunderland Massachusetts, USA, 508-516.

Gorlin, R.J. (1963). Chromosomal abnormalities and oral anomalies. Journal of Dental Research, 42, s1297-1306.

Gorlin, R.J., Pindborg, J.J., Cohen, M.M. (Eds.) (1976). Oculoauriculovertebral dysplasia. In, Seow, W.K., Urban, S., Vafaie, N. & Shusterman, S. (1998). Syndromes of the Head and Neck (2nd ed.). New York, McGraw-Hill, 546-552.

Grabb, W.C. (1965). The first and second branchial arch syndrome. Plastic and Reconstructive Surgery, 36(5), 485-508.

Graham, J.M, Hixon, H., Bacino, CA., Daack-Hirsch, S., Semina, E., Murray, J.C. (1995). Autosomal dominant transmission of a Goldenhar-like syndrome with linkage to the branchial-oto-renal syndrome. Pediatric Research, 37, 83A.

Graham, A., & Smith, A. (2001). Patterning the pharyngeal arches. Bioessays, 23(1), 54-61.

Guichard, S., & Arnaud, E. (2001). [Reconstructive surgery of the soft tissue in hemifacial Microsomia]. Annual Chiropractic and Plastic Esthetics, 46(5), 551-563.

Hodes, M.E., Gleiser, S., DeRosa, G.P., Yune, H.Y., Girod, D.A., Weaver, D.D., & Palmer, C.G. (1981). Trisomy 7 Mosaicism and Manifestations of Goldenhar Syndrome with Unilateral Radial Hypoplasia, Journal of Craniofacial Genetics and Developmental Biology, 1, 49-55.

Horgan, J.E., Padwa, B., LaBrie, R.A. & Mulliken, J.B. (1995). OMENS-plus: analysis of craniofacial and extracraniofacial anomalies in hemifacial microsomia. Cleft Palate - Craniofacial Journal, 32(5), 405-412.

Houle, J. L., Cadigan, W., Henry, S., Pinnamaneni, A., Lundahl, S. (2001). White Paper: Database Mining in the Human Genome Initiative [on-line]. Available http://www.biodatabases.com/whitepaper01.html [2003]

Ingham, P.W. & McMahon, A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Genes & Development [on-line serial], 15(23), 3059-3087. Available WWW: http://www.genesdev.org/cgi/content/full/15/23/3059?maxtosho

Jacobsson, C., & Granstrom, G. (1997). Clinical appearance of spontaneous and induced first and second branchial arch syndromes. Scandanavian Journal of Plastic Reconstructive Hand Surgery, 31, 125-136.

Ji, Y., Li, T., Shamburger, S., Jin, J., Lineaweaver, W.C., & Zhang, F. (2002). Microsurgical anterolateral thigh fasciocutaneous flap for facial contour correction in patients with hemifacial microsomia. Microsurgery, 22(1), 34-38.

Johnson, K.A., Fairhurst, J., & Clarke, N.M.P. (1995). Oculoauriculovertebral spectrum: new Manifestations. Paediatric

Radiology, 25, 446-448.

Johnson, K.R., Smith, L., Johnson, D.K., Rhodes, J., Rinchik, E.M., Thayer, M., & Lewis, E.J. (1996). Mapping of the ARIX homeodomain gene to mouse chromosome 7 and human chromosome 11q13. Genomics, 33, 527-531.

Jones, N.C., & Trainor, P.A. (2004). The therapeutic potential of stem cells in the treatment of craniofacial abnormalities. Expert opinions in Biology, 4(5), 645-57.

Jorde, L.B., Carey, J.C., Bamshad, M.J., & White, R.L. (2000). Medical Genetics (2nd ed). Mosby Inc., United States of America.

Jurliff, D.M. & Harris, M.J. (1983). Abnormal facial development in the mouse mutant first arch. Journal of Craniofacial Genetics and Developmental Biology, 3, 317-337.

Kaban, L.B., Moses, M.H., & Mulliken, J.B. (1988). Surgical correction of hemifacial microsomia in the growing child. Plastic and Reconstructive Surgery, 82(1), 9-19.

Kallen, B., Harris, J., & Robert, E. (1996). The epidemiology of orofacial clefts - two associated malformations. Journal of Craniofacial Genetics and Developmental Biology, 16, 242-248.

Kay, E.D., & Kay, C.N. (1989). Dysmorphogenesis of the mandible, zygoma, and middle ear ossicles in hemifacial microsomia and

mandibulofacial dysostosis. American Journal of Medical Genetics, 32(1), 27-31.

Kaye, C.I., Martin, A.O., Rollnick, B.R., Nagatoshi, L., Israel, J., Hermanoff, M., Tropea, B., Richtsmeier, J.T., & Morton, N.E. (1992). Oculoauriculovertebral anomaly: segregation analysis. American Journal of Medical Genetics, 43, 913-917.

Kearns, G.J., Padwa, B.L., Mulliken, J.B. & Kaban, L.B. (2000). Progression of facial asymmetry in hemifacial microsomia. Plastic and Reconstructive Surgery, 105(2), 492-498.

Kelberman, D., Tyson, J., Chandler, D.C., McInerney, A.M., Slee, J., Albert, D., Aymat, A., Botma, M., Calvert, M., Goldblatt, J., Haan, E.A., Laing, N.G., Lim, J., Malcolm, S., Singer, S.L., Winter, R.M., & Bitner-Glindzicz, M. (2001). Hemifacial microsomia: progress in understanding the genetic basis of a complex malformation syndrome. Human Genetics, 109(6), 638-645.

Keusch, C.F., Mulliken, J.B., & Kaplan, L.C. (1990). Craniofacial anomalies in twins. Plastic and Reconstructive Surgery, 87(1), 16-23.

Kushnick, T., & Colondrillo, M. (1975). 49, XXXXY patient with hemifacial microsomia, Clinical Genetics, 7, 442-448.

Loescher, A.R. (1996). Facial asymmetry with severe unilateral hypoplasia of the muscles of mastication: a possible aetiology. British Journal of Oral and Maxillofacial Surgery, 34(5), 481.

Mansour, A.M., Wang, F., Henkind, P., Goldberg, R., & Shprintzen, R. (1985). Ocular findings in the facioauriculovertebral sequence (Goldenhar-Gorlin syndrome). American Journal of Opthalmology, 100, 555-559.

Marsh, J.L., Baca, D., & Vannier, M.W. (1989). Facial musculoskeletal asymmetry in hemifacial microsomia. Cleft Palate Journal, 26(4), 292-302.

Melton, K.R., Iulianella, A., & Trainor, P.A. (2004). Gene expression and regulation of hindbrain and spinal cord development. Front Bioscience, 1(9), 117-38.

Monahan, R., Sede, K., Patel, P., Alder, M., Grud, S., & O'Gara, M. (2001). Hemifacial microsomia: etiology, diagnosis and treatment. Journal of American Dental Association, 132(10), 1402-1408.

Mulliken, J.B., & Kaban, L.B. (1987). Analysis and treatment of hemifacial microsomia in childhood. Clinical Plastic Surgery, 14(1), 91-100.

Murray, J.E., Kaban, L.B., & Mulliken, J.B. (1984). Analysis and treatment of hemifacial Microsomia. Plastic and Reconstructive Surgery, 74(2), 186-199.

Naora, H., Kimura, M., Otani, H., Yokoyama, M., Koizumi, T., Katsuki, M., Tanaka, O. (1994). Transgenic mouse model of hemifacial microsomia: cloning and characterization of insertional mutation region on chromosome 10. Genomics, 23(3), 515-519.

Neu, K.W., Friedman, J.M., Howard-Peebles, P.N. (1982). Hemifacial microsomia in cri du chat (5p-) syndrome. Journal of Craniofacial Genetics and Developmental Biology, 2, 295-298.

Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.F. (1997). Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. Development, 124(20), 4065-4075.

Perelman, B., Dafni, N., Naiman, T., Eli, D., Yaakov, M., Feng, T.L., Sinha, S., Weber, G., Khodaei, S., Sancar, A., Dotan, I., & Canaani, D. (1997). Molecular cloning of a novel human gene encoding a 63-kDa protein and its sublocalization within the 11q13 locus. Genomics, 41(3), 397-405.

Politi, M., Sembronio, S., Robiony, M., & Costa, F. (2002). The floating bone technique of the vertical ramus in hemifacial microsomia: case report. Internation Journal of Adult Orthodontist and Orthognathic Surgery, 17(3), 223-229.

Poole, M.D. (1989). A composite flap for early treatment of hemifacial microsomia. British Journal of Plastic Surgery, 42, 163-172.

Poonawalla, H.H., Kaye, C.I., Rosenthal, I.M., & Pruzansky, S. (1980). Hemifacial microsomia in a Patient with Klinefelter Syndrome, Cleft Palate Journal, 17(3), 194-196.

Poswillo, D. (1973). The pathogenesis of the first and second branchial arch syndrome. Oral Surgery, 35, 302-328.

Pruitt, K.D., Katz, K.S., Sicotte, H., & Maglott, D.R. (2000). Introducing RefSeq and LocusLink: curated human genome resources at the NCBI, Trends Genetics, 16(1), 44-47.

Pruzansky, S. (1969). Not all dwarfed mandibles are alike. Birth Defects, 5, 120.

Renzi, G., Carboni, A., Perugini, M., & Becellim R. (2002). Intraoperative measurement of maxillary repositioning in a series of 30 patients with maxillomandibular vertical asymmetries. International Journal of Adult Orthodontic and Orthognathic Surgery, 17(2), 111-115.

Rinchik, E.M., Magnuson, T., Holdener-Kenny, B., Kelsey, G., Bianchi, A., Conti, C.J., Chartier, F., Brown, K.A., Brown, S.D.M., & Peters, J. (1992). Mouse chromosome 7. Mammalian Genome, 3, S104-S120.

Robinson, L.K., Hoyme, H.E., Edwards, D.K., & Jones, K.L. (1987). Vascular pathogenesis of unilateral craniofacial defects. The Journal of Pediatrics, 111(2), 236-239.

Rodgers, S.F., Eppley, B.L., Nelson, C.L., & Sadove, A.M. (1991). Hemifacial microsomia: assessment of classification systems. Journal of Craniofacial Surgery, 2(3), 114-126.

Rollnick, B.R., & Kaye, C.I. (1983). Hemifacial microsomia and variants: pedigree data. American Journal of Medical Genetics, 15(2), 233-253.

Rollnick, B.R., & Kaye, C.I. (1985). Hemifacial microsomia and the branchio-oto-renal Syndrome. Journal of Craniofacial Genetics in Developing Biology Supplement, 1, 287-295.

Rozen,S., & Skaletsky, H.J. (1996). Primer3. [Available online: WWW: genome.wi.mit.edu/genome_software/other/primer3.html].

Sanger, F. (1981). Determination of nucleotide sequences in DNA. Science, 214(4526), 1205-1210.

Satoh, K., Shibata, Y., Tokushige, H., & Onizuka, T. (1995). A mirror image of the first and second branchial arch syndrome associated with cleft lip and palate in monozygotic twins. British Journal of Plastic Journey, 48, 601-605.

Sensi, A., Cocchi, G., Martini, A., Garani, G., Trevisi, P., & Calzolari, E. (1996). Branchio-oto (BO) syndrome and oculo-auriculo-vertebral phenotype: overlapping clinical findings in a child from a BO family. Clinical Genetics, 49, 300-302.

Seow, W.K., Urban, S., Vafaie, N., & Shusterman, S. (1998). Morphometric analysis of the primary and permanent dentitions in hemifacial microsomia: a controlled study. Journal of Dental Research, 77(1), 27-38.

Shankland, M. (2004). Axial patterning in the leech: developmental mechanisms and evolutionary implications. Genetic Regulatory Networks in Embryogenesis and Evolution [on-line]. Available WWW: mbl.edu/CASSLS/shankland.html

Shimamura, K., & Rubenstein, J.L. (1997). Inductive interactions direct early regionalization of the mouse forebrain. Development, 124(14), 2709-2718.

Silvestri, A., Natali, G. & Fadda, M.T. (1996). Dental agenesis in hemifacial microsomia. Pediatric Dentistry, 18(1), 48-51.

Silvestri, A., Natali, G., & Ianetti, G. (1996). Functional therapy in hemifacial microsomia: therapeutic protocol for growing children. Journal of Oral and Maxillofacial Surgery, 54, 271-278.

Singer, S.L., Haan, E., Slee, J. & Golblatt, J. (1994). Familial hemifacial microsomia due to autosomal dominant inheritance. Case reports. Australian Dental Journal, 39(5), 287-291.

Sujansky, E. & Smith, A.C.M. (1981) Recombinant chromosome 18 in two male sibs with the first and second branchial arch syndrome, American Journal of Human Genetics, 92A.

Sze, R.W., Paladin, A.M., Lee, S., & Cunningham, M.L. (2002). Hemifacial microsomia in pediatric patients: asymmetric abnormal development of the first and second branchial arches. American Journal of Roenthenology (AJR), 178(6), 1523-1530.

Tiner, B.D., & Quaroni, A.L. (1996). Facial asymmetries in hemifacial microsomia, Goldenhar syndrome, and Treacher Collins syndrome. Atlas Oral Maxillofacial Surgery in Clinical North America, 4(1) p. 37-52.

Townes, P.L. & White, M.R. (1978). Inherited Partial Trisomy 8q (22→qter), American Journal of Dis Child, 132, 498-501.

Trainor, P.A., & Krumlauf, R. (2000). Patterning the cranial neural crest: hindbrain segmentation and Hox gene plasticity. Natural Rev Neuroscience, 1(2), 116-24.

Trumpp, A., Depew, M.J., Rubenstein, J.L.R., Bishop, J.M. & Martin, G.R. (1999). Cre-mediated gene inactivation demonstrates that FGF8 is required for cell survival and patterning of the first branchial arch. Genes in Development, 13, 3136-3148.

Vento, A.R., LaBrie, R.A., & Mulliken, J.B. (1991). The O.M.E.N.S. classification of hemifacial microsomia. Cleft Palate Craniofacial Journal, 28(1), 68-76.

Wang, X., Lin, Y., & Yi, B. (2001). [Application of distraction osteogenesis in correction of hemifacial microsomia]. Zhonghua Yi Xue Za Zhi, 81(5), 259-262.

Whyte, A.M., Hourihan, M.D., Earley, M.J., & Sugar, A. (1990). Radiological assessment of hemifacial microsomia by three-dimensional computed tomography. Dento Maxillo Facial Radiology,

19(3), 119-125.

Wilson, G.N., & Barr, M. Jnr. (1983). Trisomy 9 Mosaicism: Another Etiology for the Manifestations of Goldenhar Syndrome, Journal of Craniofacial Genetics and Developmental Biology, 3, 313-316.

Yanagihara, N., Yanagihara, H., & Kabasawa, I. (1979). Goldenhar's syndrome associated with anomalous internal auditory meatus. The Journal of Laryngology and Otology, 93, 1217-1222.

Yang, R.Y., Hsu, D.K., Yu, L., Ni, J., & Liu, F.T. (2001). Cell cycle regulation by galactin-12, a new member of the galactin superfamily. Journal of Biological Chemistry, 276, 20252-20260.

Zellweger, H., Bardach, J., Bordwell, J., & Williams, K. (1975) The Short Arm Deletion Syndrome of Chromosome 4 (4p- Syndrome), Arch Otolaryngol, 101, 29-32.

Zhang, J., Hagopian-Donaldson, S., Servedzija, G., Elsemore, J., Plehn-Dujowich, D., McMahon, A.P., Flavell, R.A., & Williams, T. (1996). Neural tube, skeletal and body wall defects in mice lacking transcription factor AP-2, Nature Genetics, 381, 238-241.

APPENDICES

APPENDIX 1

Branchial arch derivatives affected in HFM

APPENDIX 1

Branchial arch derivatives affected in HFM

AURICLE	First arch	anterior helix
		tragus
	Second arch	helix
		antihelix
		lobule
MUSCULATURE	First arch	muscles of mastication
		tensor palatine
	Second arch	muscles of facial expression
BONES	First arch	maxilla
		zygoma
		palatine bone
		mandible
		malleus
		incus
	Second arch	stapes
		Hyoid (lesser cornu)

Table to illustrate the derivatives of the first and second branchial arches that can be affected in HFM (Adapted from Figueroa & Pruzansky, 1982).

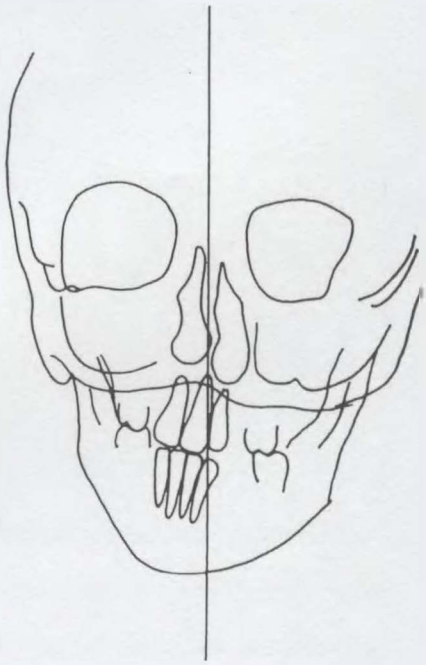
APPENDIX 2

Clinical presentation of an affected individual

APPENDIX 2

Clinical presentation of an affected individual

(Left) Front view of affected individual. Note the right-sided facial hypoplasia of the skeleton and soft tissue. Additionally, the chin deviates to the right.



(Right) PA cephalometric radiograph tracing of affected individual (right). Right-sided mandibular hypoplasia. Maxilla shows hypoplasia also. Lower and centre line displacement. Occlusal plane tilts to the right.

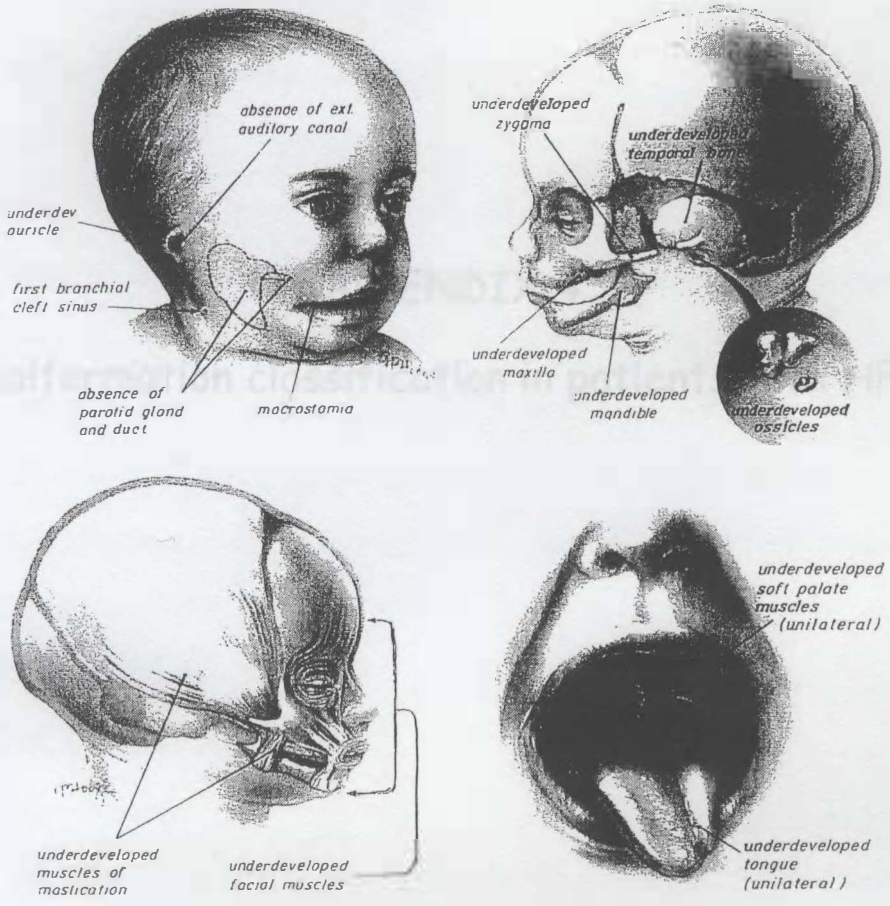
Adapted from Singer, Haan, Slee & Goldblatt (1994).

APPENDIX 3

Spectrum of congenital malformations which comprise the first and second branchial arch syndrome

APPENDIX 3

Spectrum of congenital malformations which comprise the first and second branchial arch syndrome



The above illustrations depict the spectrum of congenital malformations which comprise HFM (first and second branchial arch syndrome). A patient displaying HFM in its fullest expression would exhibit unilateral underdevelopment of the external ear, middle ear, mandible, zygoma, maxilla, temporal bone, facial, mastication and palatal muscles, tongue and parotid gland. In addition, macrostomia and a first branchial cleft sinus.

APPENDIX 4

Malformation classification in patients with HFM

APPENDIX 4

Malformation classification in patients with HFM

SKELETAL CATEGORIES

S1 – Small mandible with normal shape

S2 – Condyle, ramus, and sigmoid notch identifiable but grossly distorted; mandible strikingly different in size and shape from normal

S3 – Mandible severely malformed, ranging form poorly identifiable ramal components to complete agenesis of ramus

S4 – An S3 mandible plus orbital involvement with gross posterior recession of lateral and inferior orbital rims

S5 – The S4 defects plus orbital dystopia and frequently hypoplasia and asymmetrical neurocranium with a flat temporal fossa.

AURICLE CATEGORIES

A1 – Normal

A2 – Small, malformed auricle retaining characteristic features.

A3 – Rudimentary auricle with hook at cranial end corresponding to the helix

A4 – Malformed lobule with rest of pinna absent

SOFT TISSUE CATEGORIES

T1 - Minimal contour defect with no cranial nerve involvement.

T2 - Moderate defect

T3 - Major defect with obvious facial scoliosis, possibly severe hypoplasia of cranial verves, parotid gland, muscles of mastication; eye involvement; clefts of face or lips

---

(Adapted from Rodgers, Eppley, Nelson & Sadove, 1991 & Silvestri, Natali & Iannetti, 1996).

APPENDIX 5

Markers for Chromosome 11 Candidate Region

APPENDIX 5

Markers for Chromosome 11 candidate region

D11S1883 (64,533,373 - 64,533,624)

SYNONYMNS - RG74478, 31183, w6333, stSG34793, stSG34794, stCP2122, AFMB072WE5, D11S4146, SHGC-20653, AFMb072we5, Z53263, RH31164, AFN039XG3, AFM039xg3, STS1041, D11S1883, 039xg3, DBSTS:46439, DBSTS:31183, b072we5, HSB072WE5, RH74477

PRIMERS - 5' *AACACGAGGTTAAGCAGAG*

3' *GAATGAAGAATTTTCCAAACTAC*

D11S987 (Not currently mapped in Ensembl)

SYNONYMS - AFMa131ye5

PRIMERS - 5' *GACTCCAGTCTGGGCAATAAAAGC*

3' *GGTGGCAGCATGACCTCTAAAG*

D11S4178 (Not currently mapped in Ensembl)

SYNONYMS - DBSTS:46486, AFMB358XA9, D11S4178, AFMb358xa9, HSB358XA9

PRIMERS - 5' cgtgtccagatgaaagtg

3' caggcccagtctcttg

D11S4136 (69,221,479 - 69,221,668)

SYNONYMS - RH84412, AFMB032ZG5, RH15566, AFMb032zg5, STS16665, b032zg5, RH86312, DBSTS:31083, HSB032ZG5, stCP2113, D11S4136, Z53163, gbd:602708, SHGC-20680

PRIMERS - 5' *GAATCGCTTGAACCCAG*

3' *CCAGGTGGTCTTAACGG*

D11S4139 (69,946,835 - 69,946,983)

SYNONYMS - RH86094, AFMB038YB9, Z53184, SHGC-22368, DBSTS:46488, STS59405, AFMb038YB9, b038yb9, RH84110, HSB038YB9, stCP2116, D11S4139

PRIMERS - 5' TATAGACTTCAGCCCTGCTGC

3' CCTCTGTAGGATGCAGTTGG

D11S1314 (Not currently mapped in Ensembl)

SYNONYMS - SU1353, SHGC-2065, RH13524, 1549, D11S1314, AFM212xe3, RH37073, Z23617, RH973

PRIMERS - 5' ACAGACAGATCAAAAGGCAA

3' GAAATGTGACCTCCTTCACC

D11S4207 (Not currently mapped in Ensembl)

SYNONYMS - AFMA103ZF9, AFMa103zf9, a103zf9, HSA103ZF9, DBSTS:46492, D11S4207

PRIMERS - 5' gctgggtgttacacaggac

3' gagatcccgttcgacttg

D11S911 (75,971,191 - 75,971,382)

SYNONYMS - RH822, RH52094, AFM155XH10, STS15702, AFM155xh10, RH3099, 675, D11S911, DBSTS:675, gdb:593659, SHGC-780, RH13281, STS9925, RH72807, 155xh10, Z16690, HS155XH10, stCP1913

PRIMERS - 5' CTTCTCATGCTTGACCATTT

3' CTTCTGAACAATTGCCACAT

Above information was taken from ENSEMBL database (June 2002)

APPENDIX 6

Bioinformatics relating to genes identified on chromosome 14 between markers D14S1442-D14S267

APPENDIX 6

Bioinformatics relating to genes identified on chromosome 14 between markers D14S1442 – D14S267

1. KIAA1622

Alternate Symbols - HEAT-like repeat-containing protein, isoform 1/2
Alias – MGC4163 protein
Function – Gene ontology
known / inferred
Locus Link – 57718
Span – 93,917,777-94,023,200
Homologs – T2230 (*Caenorhabditis elegans - CE*)

NE (%)	Tissue
34.81	testis, cell line
14.54	germ cell
13.00	muscle
10.49	pool, lung, testis and b-cell
8.86	pool, melanocyte, heart and uterus
7.03	breast
6.95	eye
3.22	uterus
1.08	brain

2. SERPINA10

Alternate Symbols – PZI, ZPI
Alias – protein Z-dependent protease inhibitor precursor
Function – Gene ontology
Blood clotting, serpin, Plasma protein, inhibitor/repressor
Locus Link – 51156
Span – 94,026,778-94,036,489
Homologs - AA238242 (*Mus musculus - MM*)
U55765 (*Rattus norvegicus – RN*)
T47462 (*Arabidopsis thaliana – AT*)
T16119 (*CE*)
1617172A (*MM*)
JC4841 (*RN*)

NE (%)	Tissue
42.11	liver
26.26	corresponding non-cancerous liver tissue
25.69	hepatocellular carcinoma
5.96	pool, liver and spleen

3. SERPINA6

Alternate Symbols – CBG, CBG decreased in, CBG elevated, corticosteroid-binding globulin precursor

Alias – corticosteroid-binding globulin, alpha-1 antiproteinase, antitrypsin, serpina6 corticosteroid-binding globulin deficiency, transcortin deficiency

Family – Serpin family

Mouse Homology Maps
12 51.00cM
12 950.75cR
12 953.79cR
12 20157.15cR

Function – Gene ontology

Steroid binding, serpin, inhibitor / repressor

Locus Link – 866

Span – 94,047,714-94,066,801

Homologs – X70533 (*MM*)
BF388405 (*RN*)
T02362 (*AT*)
T16119 (*CE*)
Q06770(*MM*)
P31211 (*RN*)

Disease associations – defect in CBG (leuven) is linked with reduced cortisol binding affinity

Tissue Specificity – plasma, synthesized in liver, also identified in a number of glycocorticoid responsive cells

NE (%)	Tissue
60.90	hepatocellular carcinoma
21.77	corresponding non-cancerous liver tissue
4.91	whole embryo
3.18	pool, liver and spleen
2.50	liver
2.45	colon, 2 pooled adenocarcinomas
1.65	pancreas
0.92	kidney

0.59	heart
0.51	eye

4. SERPINA2

Alternate Symbols – ATR, PIL, ARGS, alpha-1-antitrypsin-related gene sequence, alpha-1-antitrypsin-related protein precursor

Alias – Protease inhibitor 1-like, protease inhibitor 1 (alpha-1-antitrypsin)-like

Family – Serpin family

Function – Gene ontology

Serine protease inhibitor / repressor

Locus Link – 5299

Span – 94,107,179-94,110,163

NE (%)	Tissue
100	testis

5. SERPINA1

Alternate Symbols – PI, AAT, PI1, A1AT, alpha-1-antiproteinase, PRO0684/PRO2209, alpha-1 protease inhibitor, alpha-1-antitrypsin precursor

Alias – Protease inhibitor (alpha-1-antitrypsin), protease inhibitor 1 (anti-elastase), alpha-1-antitrypsin, serpinA1 alpha-1-antitrypsin deficiency, autosomal recessive

Family – Serpin family

Mouse Homology Maps 12 51.00cM

Function – Phenotype

emphysema, emphysema-cirrhosis, hemorrhagic diathesis due to 'antithrombin' Pittsburgh

Gene ontology

Serine proteinase inhibitor, plasma glycoprotein, inhibitor / repressor, proteinase inhibitor, connective tissue development and maintenance, gas exchange, extracellular matrix maintenance, serpin acute phase response

Locus Link – 5265

Span – 94,121,848-94,126,716

Homologs – U78975 (*BT*)

AB019366 (*RN*)

T00972 (*AT*)

T16119 (*CE*)

1617172A (*MM*)

ITRT (*RN*)

Tissue Specificity - plasma

NE (%)	Tissue
19.62	gall bladder
14.35	liver
12.11	liver, hepatocyte
10.11	corresponding non-cancerous liver tissue
9.96	uterus, endometrium
8.80	hepatocellular carcinoma
4.77	thymus
3.65	marrow
2.96	pool, liver and spleen
2.26	spleen

6. SERPINA4

Alternate Symbols – KAL, KST, P14, KLST, kallistatin precursor

Alias – protease inhibitor 4 (kallistatin), tissue kallikrein inhibitor

Family – Serpin family

Function – Gene ontology

Serine protease inhibitor / repressor, plasma protein, protein binding

Locus Link – 5267

Span – 94,306,957-94,313,069

Homologs - T00972 (*AT*)
T25504 (*CE*)
S23675 (*MM*)
P35577 (*RN*)

Tissue specificity – secreted from liver cell lines

NE (%)	Tissue
21.93	ovary
15.89	pool, liver and spleen
13.27	spleen
11.09	liver
8.99	pancreas, exocrine
6.91	corresponding non-cancerous liver tissue
5.90	pool, melanocyte, heart and uterus
3.66	whole embryo
3.38	colon
3.27	mixed

7. SERPINA5

Alternate Symbols - PCI, PAI3, PROCI, PLANH3, acrosomal serine protease inhibitor

Alias - protein C inhibitor, protein C inhibitor (plasminogen activator inhibitor-3), protein C inhibitor (plasminogen activator inhibitor-III), plasma serine protease inhibitor precursor

Family - Serpin family

Mouse Homology Maps 12cM

Function - Phenotype

Protein C inhibitor deficiency

Gene ontology

Peptidase, plasma glycoprotein, protein degradation, serine protein inhibitor, extracellular (excluding cell wall)

Locus Link - 5104

Span - 94,330,818-94,335,707

Homologs - NM_008785 (*MM*)
AB013128 (*RN*)
T00972 (*AT*)
T25504 (*CE*)
P70458 (*MM*)
P09006 (*RN*)

Tissue Specificity - synthesised in liver, secreted in plasma

NE (%)	Tissue
29.97	uterus, endometrium
8.46	adrenal cortico adenoma for Cushing's syndrome
7.52	hepatocellular carcinoma
5.70	pheochronocytoma
5.16	adrenal gland
3.90	ovary, epithelium
3.71	testis, cell line
3.68	pool, liver and spleen
3.67	gall bladder
2.92	pancreas, exocrine

8. SERPINA3

Alternate Symbols - ACT, AACT, alpha-1-antichymotrypsin precursor

Alias - actichymotrypsin, alpha-1-antichymotrpsin, antichmotrypsin alpha-1

Family - Serpin family of serine protease inhibitors

Mouse Homology Maps 12 52.00cM

12 957.87cR
16 509.87cR

Function - Phenotype
Alpha-1-antichymotrypsin, cerebrovascular disease, occlusive
Gene ontology
Serine protease inhibitor, plasma protein, proteinase inhibitor, acute-phase response
Locus Link - 12
Span - 94,357,899-94,367,522
Homologs - T00972 (*AT*)
T25504 (*CE*)
JH0494 (*MM*)
P09006 (*RN*)
Disease associations - chronic obstructive pulmonary disease / occlusive
cerebrovascular disease
Tissue Specificity - plasma

NE (%)	Tissue
17.84	prostate, stroma
17.53	gall bladder
11.58	uterus, epithelium
10.83	brain, pituitary
6.4	liver
5.36	pancreas, islet
3.31	blood cd34+/cd38- hematopoietic
2.82	hepatocellular carcinoma cells
2.76	sciatic nerve
2.68	choroid

9. KIAA0928
Alternate Symbols - HERNA, K12H4.8-LIKE
Alias - helicase-moi, dicer drosophila homolog of, helicase with RNAase motif
Family - dead box helicase family, deah subfamily
Function - Gene ontology
ATP binding, DNA helicase
Locus Link - 23405
Gene lynx - 9775
Span - 94,833,138-94,885,433

Homologs - T48946 (*AT*)
S44849 (*CE*)
P40562 (*Sacharomyces cerevisiae - SC*)

NE (%)	Tissue
12.83	whole embryo, mainly head
12.17	pancreas, islet
11.87	tongue
9.04	colonic mucosa with ulcerative colitis
7.06	pheochromocytoma
5.15	human skeletal muscle
5.06	hypothalamus
4.06	prostate, metastatic prostate bone lesion
3.51	brain
2.73	macular retina

10. calmin

Alternate Symbols - FLJ12383, KIAA1188, calponin like transmembrane domain protein
Alias - hypothetical protein FLJ12383
Family - 130p family of ribosomal proteins
Mouse Homology Maps 12cM
Function - unknown
Locus Link - 79789
Span - 94,935,243-94,947,905
Homologs - P14148 (*MM*)
JC4320 (*RN*)

NE (%)	Tissue
26.30	mammary gland
16.77	ovary, tumor tissue
11.86	melanoma (neuro cell line)
9.55	human eye anterior segment
6.44	oesophagus
2.72	corresponding non-cancerous liver tissue
2.52	muscle
2.82	breast
2.07	prostate epithelium
1.90	eye, retina

11. FLJ21276

Function - unknown

Locus Link - 79686

Span - 95,152,296-95,154,718

NE (%)	Tissue
35.95	subcondral bone
12.28	CNS, ms lesions
10.39	human fetal eyes
9.48	tonsil, enriched for germinal centre B-cells
6.58	kidney
4.42	pool, melanocyte, heart and uterus
4.27	muscle
3.93	chondrosarcoma
3.85	bone
3.56	foreskin melanocyte

12. FLJ20034

Function - unknown

Locus Link - 54792

Span - 95,205,402-95,209,243

13. LOC51218

Alternate Symbols - PRO1238

Alias - clone FLB4739, PRO1238

Family - cyclic nucleotide-gated cation channel family,

Mouse Homology Maps 12103.20cR

Function - unknown

Locus Link - 51218

Span - 95,289,008-95,289,648

Tissue specificity - rod cells in the retina and inner medulla of kidney

NE (%)	Tissue
20.48	connective tissue
14.13	testis, epididymus
12.71	thymus
7.75	muscle, leg skin muscle
3.19	bone
2.80	hepatocellular carcinoma
2.70	thyroid
2.48	bone, mixed marrow stroma
2.40	human fetal eye

2.11	human lung, epithelial cell lines in treated 1ps 6hr to 1ps

14. TCL6(i)

Alternate Symbols - TNG1, TNG2, T-cell leukemia / lymphoma 6
Alias - TCL1-neighbouring gene 1 / 2, T-cell leukemia / lymphoma 6, isoform TCL
Family - dead box helicase family, deah subfamily
Function - Gene ontology
Known / inferred
Locus Link - 27004
Span - 95,408,286-95,416,877

NE (%)	Tissue
25.25	leukopheresis
18.91	placenta, human 8wk
16.29	placenta
5.40	tonsil, enriched for germinal centre b-cells
5.11	b-cells germinal
4.16	lymph
3.79	pool, liver and spleen
3.74	aorta
3.43	blood, lymphocyte
3.13	chondrosarcoma

15. TCL6(ii)

Alternate Symbols - TNG1, TNG2
Alias - TCL1-neighbouring gene 1 / 2
Function - Gene ontology
Known / inferred
Locus Link - 27004
Span - 95,396,208-95,416,877

16. TCL6(iii)

Alternate Symbols - TNG1, TNG2
Alias - TCL1-neighbouring gene 1 / 2
Function - Gene ontology
Known / inferred
Locus Link - 27004
Span - 95,408,286-95,418,482

17. TCL6(iv)

Alternate symbols – TNG1, TNG2
Alias – TCL1-neighbouring gene 1 / 2
Function – Gene ontology
Known / inferred
Locus Link – 27004
Span – 95,396,208-95,418,482

18. TCL6(v)

Alternate symbols – TNG1, TNG2
Alias – TCL1-neighbouring gene 1 / 2
Function – Gene ontology
Known / inferred
Locus Link – 27004
Span – 95,408,286-95,424,863

19. TCL6(vi)

Alternate symbols – TNG1, TNG2
Alias – TCL1-neighbouring gene 1 / 2
Function – Gene ontology
Known / inferred
Locus Link – 27004
Span – 95,408,286-95,424,863

20. TCL1B

Alternate Symbols – TML1/SYN-1/syncytio trophoblast-specific protein, T-cell leukemia / lymphoma protein 1B, TCL1B oncogene
Alias – Leukemia 1B/T-cell lymphoma
Expression – mouse embryonic expression
Family – tcl1 family
Function – Gene ontology
Oncogenesis, cell growth and/or maintenance
Locus Link – 9623
Span – 95,431,471-95,437,658
Homology – P56844 (*MM*)

NE (%)	Tissue
66.89	lymph
16.18	tonsil, enriched for germinal centre B-cells
8.16	germ cell
5.42	pool, lung, testis and B-cells
3.36	colon

21. **TCL1A**

Alternate Symbols – TCL1 / TCL1-PEN

Alias – T-cell lymphoma-1 / T-cell leukemia/lymphoma1A, T-cell lymphoma-1A, lymphoma/leukemia T-cell, P14TCL1 protein, TCL-1 protein, TCL1 oncogene

Family – TCL1 family

Function – Phenotype

Leukemia / lymphoma, T-cell

Developmental process

microsome

Differentiation

Cell growth and/or maintenance

Locus Link – 8115

Span – 95,454,995-95,459,139

Homologs – IJNP (*MM*)

Tissue specificity – restricted in the T-cell lineage to immature pharmocytes and activated peripheral lymphocytes. Preferentially expressed early in T- and B- lymphocyte differentiation

NE (%)	Tissue
60.49	lymph, b-cell
15.67	lymph
15.62	b-cells
2.45	b-cells from Burkitt lymphoma
1.41	muscle
0.95	blood, lymphocyte
0.94	testis, cell-line
0.59	leukocyte
0.45	tonsil, enriched for germinal centre b-cells
0.35	pool, liver and spleen

22. **DKFZp761F2014**

Alternate Symbols – DKFZP761F2014

Function – Gene ontology

unknown

Locus Link – 56967

Span – 95,835,457-95,838,917

Homologs – T19201 (*CE*)

S71512 (*MM*)

A57514 (*RN*)

S78475 (*SC*)

NE (%)	Tissue
23.93	tongue
17.49	brain, hippocampus
17.19	whole embryo, mainly body
10.96	brain, cerebellum
4.03	human fetal eye
3.39	uterus, epithelium
3.31	normal lung tissue epithelial cells
3.06	rpe and choroid
3.01	brain
2.24	eye, retina

23. BDKRB2

Alternate Symbols – B2R, BK3, BKR2, BRB2, B2 Bradykinin receptor, BK-2 receptor, Brady Kin receptor 2

Family – family 1 of g-protein coupled receptors

Mouse Homology Maps 12 53.00cM
12 968.01cR

Function – Gene references into function

Two cysteine residues located in the carboxyterminal domain of the bradykinin B2 receptor are palmitocylated & play a crucial role in the response of the repressors to ligand binding

Gene ontology

Virulence, circulation, invasive growth, plasma membrane, sensory perception, Bradykinin receptor, smooth muscle contraction, integral plasma membrane protein, cytostolic calcium ion concentration elevation, cell surface receptor linked signal transduction, phosphatidylinositol-4,5 biphosphate hydrolysis, G-protein linked receptor protein tyrosine kinase signalling pathway

Locus Link – 624

Span – 95,949,888-95,989,480

Homologs – T30999 (*CE*)
JC7209 (*DM*)
P32299 (*MM*)
P25023 (*RN*)

NE (%)	Tissue
28.86	lung, cell line
22.17	thyroid gland

20.69	colonic mucosa with ulcerative colitis
5.29	placenta human full-term
3.91	lung metatastic chondrosarcoma
3.47	kidney
3.10	pool, melanocyte, heart and uterus
2.88	germ cell
2.77	prostate, epithelium
1.29	mixed

24. BDKBR1

Alternate Symbols - BIBKR, BIR, BKR1, BRADY B1, BK-1 receptor
Alias - Bradykinin receptor B1, B1 Bradykinin receptor
Family - Family 1 of g-protein coupled receptors
Mouse Homology Maps 12cM
Function - Gene ontology
G-protein linked/coupled receptor, plasma membrane, bradykinin receptor, integral plasma membrane protein, cytostolic calcium ion concentration elevation, inflammatory response, pain sensation, endoplasmic reticulum, G-protein coupled / linked receptor protein signalling pathwat
Locus Link - 623
Span - 96,008,705-96,009,786
Homologs - NM_007539 (*MM*)
AJ132230 (*RN*)
T30999 (*CE*)
JC7209 (*DM*)
JC4681 (*MM*)
P97583 (*RN*)

NE (%)	Tissue
65.85	chondrosarcoma
15.50	unclassified
7.30	stomach
6.89	kidney
4.46	skin

25. FLJ10242

Family - Mouse Homology Maps 12 1003.20cR
Function - unknown
Locus Link - 55102
Span - 96,029,774-96,054,546

Homologs – T16637 (*CE*)
S63208 (*SC*)

NE (%)	Tissue
27.08	whole embryo, mainly head
17.33	foveal and mucular retina
9.16	small intestine
7.33	placenta, human full-term
6.35	human fetal eye
4.81	kidney
4.30	pool, melanocyte, heart and uterus
4.06	adrenal gland
3.52	eye, retina
2.67	uterus, epithelium

26. HSPC210

Family – Mouse Homology Maps 12cM
Function – unknown
Locus Link – 51527
Span – 96,127,276-96,130,759
Homologs – T26857 (*CE*)

NE (%)	Tissue
11.02	human optic nerve
6.70	pancreas, islet
6.02	human lens
5.35	leukopheresis
4.53	human skeletal muscle
4.40	human fetal eye
4.29	parathyroid
3.72	bone marrow
3.70	colon
3.41	pancreas, exocrine

27.MGC5378

Alternate Symbols – PAPOLA
Alias – poly (A) polymerase alpha
Family – poly (A) polymerase family
Function – mRNA processing, transcription, nucleus, RNA binding
Locus Link – 84718
Span – 96,247,453-96,277,724
Homologs – T10692 (*AT*)
T22140 (*CE*)

S19031 (*SC*)

NE (%)	Tissue
23.78	brain, frontal lobe
10.79	blood, white cells
9.37	testis cell line
4.97	human eye, anterior segment
3.67	human skeletal muscle
3.23	ear, cochlea
2.64	umbilical cord, endothelium
2.52	CNS, ms lesions
2.44	human lens
2.43	placenta human 8wks

28.VRK1

Alternate Symbols – Vaccinia virus BIR-related kinase 1

Alias – Vaccinia-related kinase-1

Expression - fetal

Family – Ser/Thr family of protein kinases

Mouse Homology Maps 12cM

Function – Gene ontology

Protein phosphorylation, protein serine/threonine kinase, transferase, ATP binding

Locus Link – 7443

Span – 96,542,405-96,626,629

Tissue Specificity – expressed in photoreceptor cells of the eyes as well as the region situated between the optic lobe and the central brain

Homologs – BC016676 (*MM*)
H359602 (*RN*)
C71405 (*AT*)
T16194 (*CE*)
O76324 (*DM*)
S29522 (*SC*)
1CKJ (*RN*)
S47616 (*MM*)

NE (%)	Tissue
27.34	normal gingiva (cell live from immortalised keratinocytes)
12.70	nose, olfactory epithelium
10.11	bone marrow
7.84	blood

6.34	bladder
4.12	whole embryo, mainly head
3.24	lung epithelial cells tissue
2.83	testis
2.65	bone, mixed marrow stroma
2.24	cervix

NE% - Normalised expression (%)

Indicators used to assess homology with chromosome 11 genes

APPENDIX 7

cDNA / Exonic Primers for Candidate Genes

APPENDIX 7

cDNA / Exonic Primers for Candidate Genes

UVRAG cDNA/Exonic Primers

Amplicon	Primers (Base numbers)	Primer Sequences
1	74 049 814F	5'-cagtaatgccagcgatggacag-3'
	74 050 433R	5'-tgagctggctggctgcttgtc-3'
2	74 086 399F	5'-gagactgggttatggtttctgg-3'
	74 086 741R	5'-aagcaggaggcactgagaatca-3'
3	74 094 747F	5'-cactagcaaccaggtaacctaca-3'
	74 096 084R	5'-gtatacatgtgccatgtcggtgtg-3'
4	74 114 287F	5'-ccatgtaagtgagtgatagg-3'
	74 114 840R	5'-catcaaggtctctcttagtg-3'
5	74 123 374F	5'-gccggtttctcagtattgaag-3'
	74 123 748R	5-atggactggatggttatgtgc-3'
6	74 146 376F	5'-gctccttctccactcttgtcag-3'
	74 146 915R	5'-ttctctcccactgtccagag-3'
7	74 196 035F	5'-ctgaggctctttgttgag-3'
	74 196 397R	5'-gcaggcacttcatgcgttac-3'
8	74 217 800F	5'-ttgtgtagtgtctgttgggtc-3'
	74 218 291R	5'-ccctgtcattaagtgatgtgtg-3'
9	74 238 567F	5'-ctctcacagtcagggatttggc-3'
	74 238 924R	5'-tggcacagtgcatgacacacag-3'
10	74 242 073F	5'-ctgtggcctattacagacag-3'
	74 242 494R	5'-tgatgagtgcatcccacca-3'
11	74 243 311F	5'-caccatgtactcatcatccagc-3'
	74 243 678R	5'-atctcccatcaaggcgtctg-3'

12	74 251 478F	5'-ggacatagtaagagttcttgc-3'
	74 251 635R	5'-aatgaggcactctacaaagg-3'
13	74 300 291F	5'-gcctctgttgacaacttggag-3'
	74 300 550R	5'-aaatgcgagggtcaactgcag-3'
14	74 350 386F	5'-gtgtgcatactcatacacgtgc-3'
	74 350 755R	5'-tgtggagcaatgtggaagtcag-3'
15 (i)	74 375 366F	5'-gatccagaaggcttacactgag-3'
	74 375 968R	5'-*CAGAAGAAATCATCGGGCTGGT* -3'
15 (ii)	74 375 842F	5'-*AATGGCACTCTCCTACCCAG*-3'
	74 376 425R	5'-*GAGCTTACAGCTCGAGTCACCT*-3'
15 (iii)	74 376 320F	5'-*CAGTGGGCCAAAGTTAGA*-3'
	74 376 860R	5'-*CGGCCCTCAGATTAGATGAA* -3'
15 (iv)	74 376 238F	5'-*AGCTGTGACGTTCCATCTCTTC*-3'
	74 376 502R	5'-*AATGCGGGAATGACAACTGGAC*-3'

UVRAG exonic base numbers used to identify primers. Lower-case identifies 5'-UTR and 3'-UTR primer sequences identified from GenBank

GARP cDNA/Exonic Primers

Amplicon	Primers	Primer Sequences
1	-114F	5'-taagtcagctgaggccgagag-3'
	74R	5'-ctccagcacatgctgagccg-3'
2	3655F	5'-ttccagctccagccgtgctc-3'
	3967R	5'-cctgactcttctaacttctggc-3'
3	7976F	5'-tccacccatggaaccttcatcga-3'
	8639R	5'-GGGTGAGGCGAGTCAGACTGTT-3'
	9235F	5'-gGTGGACAAGAAGGTCTCGTGC-3'
	8530F	5'-TGGACCTGTCTGGGAACA-3'
	9106R	5'-ATCCAGATTCAAGAGCTGG-3'
	8998F	5'-AGGACAGCAAGGGCATCC-3'
	9626R	5'-CCTCCAAGGAGGCCTCC-3'
	9635F	5'-ACTGACTGAGCTGGACC-3'
	10060R	5'-GGCAGAGACCAGTATGAA-3'
	9974F	5'-GAGCCACGTGCGTCCTGAGGAC-3'
	10581R	5'-CGGGGCCTGCCGAGCTCTGGA-3'
	10489F	5'-TAGGAGAGAGTGCTGCAGAG-3'
	11057R	5'-CAGACACAAGGCTTGGATTCA-3'
	11014F	5'-GCACCCAGCTTGGCAGATGTG-3'
	11619R	5'-AGAGAGGACATCACTCTGGTCC-3'
	11499F	5'-CTGAGGCTTAGGAAGAGAATG-3'
	12307R	5'-tgccatgatgattgaacgacc-3'

5' *GARP* exonic base numbers used to identify primers. Lower-case identifies 5'-UTR and 3'-UTR intronic primer sequences. Uppercase identifies exonic primer sequence

CLNS1A cDNA/Exonic Primers

Amplicon	Primers	Primer Sequences
1	-101F	5'-tccacacgttcttagccgacctc-3'
	373R	5'-gagccttccacccgctacaggta-3'
2	7740F	5'-ctcagtttccactactgaa-3'
	8140R	5'-gattacaggcgtgagcca-3'
	7810F	5'-gcataaaggtgtccaagc-3'
	7848F	5'-gtggtcttacatgaggatttac-3'
	8191R	5'-ctcgatctcctgaccttgtg-3'
3	11825F	5'-ctaggatgctcactgtat-3'
	12208R	5'-agaatggaggtaatggtg-3'
4	12590F	5'-gcttcaaccgctttcaag-3'
	12931R	5'-gaactataatcctctaccc-3'
5	14976F	5'-gtgccaccacgcctggctaa-3'
	15332R	5'-gtgctcatctgactgttcacac-3'
6	18008F	5'-cactagcttccttctggtaga-3'
	18313R	5'-atgccactgcactgtagcct-3'
7(i)	21001F	5'-tacctaagtcctctggca-3'
	21492R	5'-ctaccaagctcctgtgc-3'
7 (ii)	21419F	5'-gcacaggagcttggtag-3'
	21809R	5'-gccttactaaatacttgcc-3'

5' *CLNS1A* exonic base numbers used to identify primers. Lower-case identifies 5'-UTR and 3'-UTR primer sequences.

***ARIX* cDNA/Exonic Primers**

Amplicon	Primers	Primer Sequences
1	-170F	5'-ccacacctctgagccctaagacg-3'
	490R	5'-cattggagggtctggccaaggca-3'
	-170F	5'-ccacacctctgagccctaaga-3'
	421R	5'-tctgcaggaat-3'
	1F	5'-*CTGAGTGCGGCCGCGAC*-3'
	379R	5'-gcgctcac*CTGCCGAGTA*-3'
2	2722F	5'-ggctgccgggaccaagacga-3'
	3160R	5'-gccttcgggctgcatctgcc-3'
	2693F	5'-ttgagtaatagggaggacgct-3'
	3293R	5'-tcggaactttctgccctaaga-3'
	2870F	5'-*TGCCCTACAAGTTCTTC*-3'
	3057R	5'-*CTGCACGCGAGCCTCA*-3'
3	3863F	5'-tccgaaccaggatctcactcgag-3'
	4580R	5'-gagtggccctgacttggtctcc-3'
	3958F	5'-ag*GTCTGGTTCCAGAACCG*-3'
	4414R	5'-*GCAGCTAGAAGAGATTGGTC*-3'

5' *ARIX* exonic base numbers used to identify primers. Lower-case identifies 5'-UTR and 3'-UTR primer sequences.

APPENDIX 8

Gene list derived from location of markers on chromosome 14 D14S1442-D14S267 (2002)

APPENDIX 8

Gene list derived from location of markers on chromosome 14 D14S1442-D14S267

1. SERPINA1
2. SERPINA4
3. SERPINA5
4. SERPINA3
5. KIAA0928
6. calmin
7. FLJ21276
8. FLJ20034
9. LOC51218
10. TCL6
11. TCL1B
12. TCL1A
13. DKFZp761F2014
14. BDKRB2
15. BDKRB1
16. FLJ10242
17. HSPC210
18. MGC5378
19. VRK1

APPENDIX 9

Gene list derived from location of Markers on chromosome D11S1881-D11S911 (2002)

APPENDIX 9

Gene list derived from location of Markers D11S1881-D11S911

1. HRASLS3
2. ESRRA
3. HSPC152
4. PRDX5
5. PLCB3
6. BAD
7. LOC56834
8. KCNK4
9. LRP16
10. FLRT1
11. STIP1
12. HSPF2
13. MGC13045
14. MGC11134
15. MGC10966
16. COX8
17. FLJ13848
18. FLJ20113
19. HRLP5
20. LGALS12
21. RARRES3
22. HRASLS2
23. DKFZp434G0920
24. RPS6KA4
25. SLC22A11
26. NRXN2

27. RASGRP2
28. PYGM
29. SF1
30. MAP4K2
31. MEN1
32. EHD1
33. PPP2R5B
34. GPHA2
35. NAALADASEL
36. CAPN1
37. MRPL49
38. FAU
39. C11orf5
40. TM7SF2
41. C11orf2
42. ZFPL1
43. MGC16386
44. MGC10966
45. HSU79266
46. SNX15
47. ARL2
48. LOC116071
49. EMK1
50. POLA2
51. CEP2
52. REQ
53. FKSG44
54. SPRY4

55. NTKL
56. LTBP3
57. SF3B2
58. GSTP1
59. NDUFV1
60. MGC9740
61. CABP2
62. DOC-1R
63. PITPNM
64. AIP
65. ALDH3B2
66. UNC93B1
67. ALDH3B1
68. NDUFS8
69. TCIRG1
70. CHK
71. FLJ20039
72. CGI-85
73. C11orf24
74. LRP5
75. C11orf23
76. LOC51083
77. MTL5
78. CPT1A
79. IGHMBP2
80. CCND1
81. FGF19
82. FGF4

83. FGF3
84. FLJ10261
85. FADD
86. PPFIA1
87. EMS1
88. SHANK2
89. FLJ10661
90. FLJ11099
91. IL18BP
92. NUMA1
93. FLJ20625
94. DKFZP564M082
95. FOLR3
96. FOLR1
97. FOLR2
98. INNPL1
99. ARIX
100. SKD3
101. SDCCAG28
102. KIAA0769
103. P2RY6
104. KIAA0337
105. TNRSF19L
106. PHRET1
107. PME-1
108. DKFZP586N2124
109. FLJ22596
110. KIAA0102

111. NEU3
112. SLC21A9
113. ARRB1
114. RPS3
115. PP1665
116. SERPINH2
117. FLJ22644
118. DGAT2
119. **UVRAG**
120. WNT11
121. PRKRIR
122. C11ORF30
123. **GARP**
124. E2IG4
125. PHCA
126. CAPN5
127. OMP
128. MYO7A
129. PAK1
130. **CLNS1A**
131. HBXAP

Bold text depicts those genes identified in this thesis as candidate genes for chromosome 11 HFM.

APPENDIX 10

List of trinucleotide repeats present on chromosome 11 between markers D11S1881-D11S911 and identified in this project

APPENDIX 10

List of trinucleotide repeats present on chromosome 11 between markers D11S1881-D11S911 and identified in this project

1. Very large expansions of repeats outside coding sequences.

(i) Repeat $(CGG)_n$

Disease;	Fragile-X site A (FRAXA)
MIM No	309550
Inheritance	X-linked
Location	Xq27.3 / 5'UTR
Stable Repeat	6-54
Unstable Repeat	200-1000

CGGACACCCCGTTCTCCTGGTGCCCAGCGCGGGGGCCCCCCTACCTGCAGGCTGTCGGCGCAT
GGCTG**CGGCGGCGGCGGCGG**CTCGTCCGGCTTGAGGAAGACCTGCTGGCCCCGCCTGAGGAA
TTGGACAACAGCGATGAGGATGTGGTGCAGCACCAGGACCATGCTCGCAG

Base Numbers:	76,256,733 - 76,256,747
Number Repeats:	5
Nearest Gene:	PDE2A

CGGGCGTCAGGGTGCATTGTGGGAAGGAGGCGGCAGCGTCTCGGGCGGGC
GGGAGGTGCACCAG**CGGCGGCGGCGGCGG**TAAATCCTCCCGGCCGCTCAC
AGCACTGTGGAGCCGCGTCCCCAGCCCGGCCTCGGACCGCGGCACCCCTC

Base Numbers:	82,148,443 - 82,148,457
Number Repeats:	5
Nearest Gene:	Between GL002 & PRKRIR

CAGACTCAGGACCAGCGGACCCTGTGGCGGGGGGAGTCTCCTCACCGGTGAGCAGCTGCAG
GCGAAAA**CGGCGGCGGCGGCGG**CTGAGGCGGCGATGCCCTTGTCTTCTCCGGAGGCTACGG
CGGCGACGCTCTCGCCCGTGTCCAGTAGGGTGGC

Base Numbers:	84,487,220 - 84,487,234
Number Repeats:	5
Nearest Gene:	Between FLJ14993 & PHRETI

(ii) Repeat $(CCG)_n$

Disease;	Fragile-X site E (FRAXE)
MIM No	309548
Inheritance	X-linked

Location	Xq28 promoter
Stable Repeat	6-25
Unstable Repeat	200+

GTCCGGGTCTCTGGGCGAAGTGACTGAGGCGGTCGCGCCGCGAGAGAACAACAGTCCCAGA
TGTCTGGGTCCTGGCT**ccgccgccgccgccgccg**cccgccccgcctacggccccgccccgcgccttaggccccgc
cccttcccgcccTTGCTGATTTAAGTCCCCGAAGTTCTAGACTGGAGCCCGGGAGCGCGCTCCC

Base Numbers:	75,258,572 - 75,258,589
Number Repeats:	6
Nearest Gene:	Between PME-1 & FLJ10661

gttctcctcgcaccagtccagcgtggaggtcgtggggccccagtagccctctcggtccgcggccggagccatcacg**ccgc**
cgccgccgccgcccaggcgctccgctcactgtgccgggttaggctggcgaccaggccggctccggcccag

Base Numbers:	81,668,555 - 81,668,569
Number Repeats:	5
Nearest Gene:	PHCA

(iii) Repeat $(GAA)_n$

Disease;	Friederich's Ataxia (FA)
MIM No	229300
Inheritance	AR
Location	9q13-q21.1 intron 1
Stable Repeat	7-22
Unstable Repeat	200-1700

ccccctaccccgcacatctcgttttccggtgcgacagcaagttcaccgtctccaggacttggctctgctctcacaccttaa
acccttaaaagaaaaagctaagtttaagctatttgcctttaagtcataaagacaccaaaagtatttaaagtgcagatctag
aagaagaagaagaacgcctagatcaaactgacccagaagatctcaggctggctctagtcctcctccctcaatcttaaagc
tacagtaatgtagcaagtagtattaggtgttgta

Base Numbers:	73,965,721 - 73,965,735
Number Repeats:	5
Nearest Gene:	Between HREV107 & MRPL48

AAGAACAGGAAGCAGTTGTTACTGAAGGGTAGATACGGGAGGAACAAATGTGAGAAAGA
AGTGAT**gaaggaaaagaagaagaagaagaagaagaagaa**gaTGATGATGATGGTGATG

Base Numbers:	78,444,113 - 78,444,136
Number Repeats:	8
Nearest Gene:	DLG2

AGTCACACAAGATTGTCATGCGAGCTGAAGGGGGCTGAAATTATTCCCTCT**gaagaagaaga**
agaagaagaaAACATACTGCAGCTCCCGATTTGGTAAAAAGCCAGTGGCTTGCT

Base Numbers:	79,541,889 - 79,541,909

Number Repeats: 7
Nearest Gene: DLG2

AATGATCTGCCTTCTGGCCTAGCAATTTTCCGGAGATACTGTGCAATGTGATTGTTTCTT
TCCTGAGAAAAATGTTTAAAAAGGCAGCTTATTTAGGTCAAAG**GAAGAAGAAGAAGAA**
ATAGGAGAGGTATGAAAGAGAATGAAGATAAAGTTGTGTAGAGAAAAAGAGAGATTAC
TTGAGTCTTAAGAAAATGTGACTTCTTCCTTCTGCTACTT

Base Numbers: 81,282,464 - 81,282,478
Number Repeats: 5
Nearest Gene: Between PAK1 & MYO7A

agttgtgggcaGCcatagtggctcacatctgtaatcccagcactttgggaggacgagggacgagatgggaggatcacttg
tgcccaggaggtcaaggctgcagtgagcttcactccagcctggccaacagaatgagacactgtctcaaaaaaaaaaaaaa
aaa**gaagaagaagaaGAA**GTTCAAAGAAAAAGAAAGAAAAAGTCACTGAAATTAATGATCAT
TTTTAatgttgtcaatttgacatacggaatttggtttatttgtttctgttggcatacaattttagatgcctcccatttcc
tgtctgtcttgattttaactttctgaat

Base Numbers: 85,260,387 - 85,260,401
Number Repeats: 5
Nearest Gene: PTD015

attagttgggcatggtggcacatgcctgtaatcccggtactggggaggttggggcacgagaaccacttgaacccaggagg
tgaaggttgcagtgagttgagatcacaccagtgcattccagcctaggtgacggagtgagaccctgtctcaaagaaacaaa
aaaaaaaaaacaaaaaaaaAAAAGAAGAAGAAGAAGAAGGAAAAAATCTTGAATATAAATGGA
TTAAGTTCTTCAATTAAAAGGtacactacagcctgggcaacagagagacagacaccctgtctcaaagaaaga
aaaaaaaaaaaag

Base Numbers: 85,356,826 - 85,356,840
Number Repeats: 5
Nearest Gene: Between PTD015 & MGC16733

(iv) Repeat $(CTG)_n$

Disease;	Myotonic Dystrophy (DM)
MIM No	160900
Inheritance	AD
Location	19q13 3'UTR
Stable Repeat	5-35
Unstable Repeat	50-4000

Disease;	Spinocerebellar Ataxia 8
MIM No	606364
Inheritance	AD
Location	13q21 untranslated RNA
Stable Repeat	16-37
Unstable Repeat	110-500

TACATGCTGAATAG**CTGCTGCTGCTGCTG**CCAGCTTTTCAGGTCTGGAGAGAAAGGTTAT
TTCTCATAAGAAGTCACAAGGG

Base Numbers: 73,993,056 - 73,993,070
Number Repeats: 5
Nearest Gene: MRPL48

(v) Repeat $(CCCCGCCCCGCG)_n$

Disease; Juvenille Myoclonus Epilepsy (JME)
MIM No 254800
Inheritance AR
Location 21q22.3 promotor
Stable Repeat 2-3
Unstable Repeat 40-80

TGAGGCGGTCGCGCCGCGAGAGAACAACAGTCCCAGATGTCTGGGTCCTGGCTccgccgccgcc
gccgccgcccgccccgcctacgg**ccccgccccgcg**ccttaggccccgcccctttcccgcccTTGCTGATT

Base Numbers: 75,258,606 - 75,258,618
Number Repeats: 1
Nearest Gene: Between PME-1 & FLJ10661

2. Modest expansions of CAG repeats within coding regions.

Repeat $(CAG)_n$

Disease; Huntington Disease (HD)
MIM No 143100
Inheritance AD
Location 4p16.3 coding
Stable Repeat 6-35
Unstable Repeat 36-100

Disease; Kennedy Disease (SBMA)
MIM No 313200
Inheritance XR
Location Xq21 coding
Stable Repeat 9-35
Unstable Repeat 38-62

Disease; Spinocerebellar Ataxia 2 (SCA2)
MIM No 183090
Inheritance AD
Location 12q24 coding
Stable Repeat 14-31

Unstable Repeat	32-77

Disease;	Machado-Joseph Disease (SCA3, MJD)
MIM No	109150
Inheritance	AD
Location	14q32.1 coding
Stable Repeat	12-39
Unstable Repeat	62+86

Disease;	Spinocerebellar Ataxia 6 (SCA6)
MIM No	183086
Inheritance	AD
Location	19p13 coding
Stable Repeat	4-17
Unstable Repeat	21-30

Disease;	Spinocerebellar Ataxia 7 (SCA7)
MIM No	164500
Inheritance	AD
Location	3p12-p21.1 coding
Stable Repeat	7-35
Unstable Repeat	37-200

Disease;	Dentatorubral-Pallidoluysian Atrophy (DRPLA)
MIM No	125370
Inheritance	AD
Location	12p coding
Stable Repeat	3-35
Unstable Repeat	49-88

GAATAATAGGAAAGATGGAATTCCAAAGATAATGTAATAAAGGGAGGTATGATAggttag
gggaattaacaaggaataatgaagcactccggaacc**cagcagcagcagcagcagcagcagcag**caagaaaccttta
ctatccctaaggattagggcaagagagaagaaatagggagttacagtctgatgagaagaggagccttggaagaagtgttg
actgacag

Base Numbers:	79,623,417 - 79,623,446
Number Repeats:	10
Nearest Gene:	DLG2

GAGTGACTATGGATAAAGTCATCATAAAAATAAACTTATATCATCAATCTCCATGATCA
ATTTCTTTAAAGCAAAATATGATATG**CAGCAGCAGCAGCAG**AGTACAGAAGGTACACTG
GTCAGATAACCACAATTTTAAGCTAATTTTGGCCATGTGCTTGCTG

Base Numbers:	79,934,190 - 79,934,204
Number Repeats:	5
Nearest Gene:	Between DLG2 & HT007

TTGGCACCCGGGGAAGCATGGCCGGGTTGTCTGGGCCCCACTCACGGC**CAGCAGCAGCAGC**
AGGGGGCCACGGCATGGTGGGGGCTAGAAAGAGAGCCACAGGTGA

Base Numbers: 81,733,884 - 81,733,898
Number Repeats: 5
Nearest Gene: Between PHCA & E2IG4

ATCCCTGCCTGAGGTTTGCCCTACTTCATGGGATGCCCTGTGGCCTCCCCGCTCCTGTGTGG
GTTGGTAGGA**CAGCAGCAGCAGCAG**CTTCTGTGGGGCTCCAGGTCCGACCCAGCCCCAGCCC
CTCCCTCCCTGGCTCCAGGAAGTCTGGAGCCCCGAG

Base Numbers: 82,290,689 - 82,290,703
Number Repeats: 5
Nearest Gene: Between PRKRIR & WNT11

cagttctaactcagtttcctcctgttttagcagtctaagatttggcaagccatgttctctctctgggctttagtcttccatt
tataaaacaaatggactgaataagatgatatttaaggtccttttcagctacaagagtccatgttccaaga**GAAGATAA**
AAATTGCAGCAGCAGCAGCAGCAATATTCACTGAAGACTTAGCATGTGCCAGAATCTTGC
TAGGAACCA

Base Numbers: 83,710,316 - 83,710,330
Number Repeats: 5
Nearest Gene: Between KIAA0102 & FLJ22596

TGAGATGCTCCAGCAAAAGAAC**CAGCAGCAGCAGCAG**GAAGTTTCAATGGGGCCGGGCAG
AGCATGCGCCCTGCACCCGCGCGGTGGCCTTATGTGGAAGGAGCA

Base Numbers: 85,227,600 - 85,227,614
Number Repeats: 5
Nearest Gene: Between RAB6 & THRSP

3. Other short repeats within coding regions.

Repeat	$(GCG)_n$
Disease;	Oculopharyngeal Muscular Dystrophy
MIM No	9462747
Inheritance	AD / AR
Location	14q11.2
Stable Repeat	6
Unstable Repeat	7-13

CCGGACACCCCGTTCTCCTGGTGCCCAGCGCGGGGGCCCCCCTACCTGCAGGCTGTCGGCGCA
TGGCT**GCGGCGGCGGCGGCG**GCTCGTCCGGCTTGAGGAAGACCTGCTGGCCCCGCCTGAGGA
ATTGGACAACAGCGATGAGGATGTGGTGCAGC

Base Numbers: 76,256,732 - 76,256,746
Number Repeats: 5

Nearest Gene: PDE2A

AGAGGGCACCCCCGGGCGTCAGGGTGCATTGTGGGAAGGAGGCGGCAGCGTCTCGGGCGGG
CGGGAGGTGCACCA**GCGGCGGCGGCGGCG**GTAAATCCTCCCGGCCGCTCACAGCACTGTGG
AGCCGCGTCCCCAGCCCGGCCTCGGACCGCGGC

Base Numbers: 82,148,442 - 82,148,456
Number Repeats: 5
Nearest Gene: Between GL002 & PRKRIR

4. Triplet repeats without disease associations.

Repeat (AAC)

tattccaaaaaattagggaggagggtctcctccctaactcattctattaggccagcttcatcctgataccaaaacctggca
gagattacaaa**aacaacaacaacaacaac**aaaaaaacacaaaaacttcaggccaatatctttggtgaatatcaatgcaa
aaatcctcaacaaaatactggcaaaccaaatccagcagcatatcagaaaaattatccaccatgatcaagt

Base Numbers: 72,956,999 - 72,957,016
Number Repeats: 6
Nearest Gene: Between HRLP5 & SLC22A8

gcatgtgcctgtagtcccagctactcagaaagctgaggtgggggatcgcttgaacccaggagattgaggctgcagtgaa
ctatgaattcaatgaagtgccattgcccttcagccagggtgacagagtgagaccctatctcaa**aacaacaacaacaac**aa
attaaaaaaaaaaTTCCCTACCATGTTTCCTAAGAAGCAGATTTAACCAGGCAAAAGTTTTAC
TAATTTTT

Base Numbers: 73,417,540 - 73,417,554
Number Repeats: 5
Nearest Gene: Between SLC22A6 & HREV107

GGGAAGAAGGTAAGGATGAAAATGACCTTACAGATT**AACAACAACAACAACAAA**ATAT
GAACTTCAGCATGTACTGTGTTAACCCAATTAAAAACCTTTTATCAAAGAA

Base Numbers: 73,506,822 - 73,506,836
Number Repeats: 5
Nearest Gene: Between SLC22A6 & HREV107

ggatcgtttgaagtcaggagttccagactaccctgggcaacatagtaagatcctgtctctacaaaaaatacaaaaattagc
tgggcatggtatgtgcatctgtagccccagctactcgggaggctgaggtgggaggatcttgagcccaggagatggagact
gcagtgagccatgatcataccactgcactccagcctaggtgacagagtgagaccctgtctcTAAaacag**caacaacaa**
caacaacaataacaacaaGAGTACAGCCTTTACCTCCCTCAGTAGTTCCCACCTCAGTACAGAGT
TATCACTCCTTTC

Base Numbers: 74,396,542 - 74,396,556
Number Repeats: 5
Nearest Gene: KIAA0769

tggcgcatgcctgtaatcccagctacttgggaggctgaggcgggagaatcgcttgaacccaggaggtggaggttgcggtg
aaccgagattgcaccattgcactccagcctgggcaacaagagtgaaactccatctcaaaacaacaacaacaacaacaac
aacaacaacaacaacaacaacaacaaaagcatgtaaagcgcttaggtcagtgcctTCCTTTTAAGGCACTCTGTT
AATTGTTGTTGTTAATAGGAGAAGGGTTTTCTTGAGAAGGGAGGAAGGGACGCGAAGAG
GAATGGTC

Base Numbers: 74,397,612 - 74,397,653
Number Repeats: 14
Nearest Gene: KIAA0769

ACACATCCGCATTTAAAACAATAGTAggccaggtatggtggctcacgcctgtaatcatagcactttaggag
gccaagacgggagaatcacttgaggccaggagttcaatatcagcctgggcaacatagtgagaccccatctattaaaaaca
acaacaacaacaacaacaacaacaacaaaaaacttcttaaaaataaaccaggcatgatggtacacatctgtagtcctagcta
cttggaagtctgaggtgggagg

Base Numbers: 74,536,561 - 74,536,587
Number Repeats: 9
Nearest Gene: KIAA0769

CCTGTTGGgccgagcatggtgctgtgcacctgtaatcccagcccttcgggaagaccaggtgggaggattgcttaaacc
caggggtttaaggctgcaggagttTGAGGCTGCACTAAGGTTTTTTTGTGTGTCTCAAAAGCAACAA
CAACAACAACAAAAGTCACTCTGGCCATAGTAAGGGAAAGGATTGGATGGGGAGACTGC
AGGTCTAGAAGGTGGCCCACAGTGGCAGGAGGGTGGAGGAGTACTAG

Base Numbers: 74,666,789 - 74,666,803
Number Repeats: 5
Nearest Gene: Between KIAA0769 & UCP2

AAGGCAGAGGACCAAATTCCCTAGGCTAGGGAATCAGGAAAGTTGCAGTTGGAGCTGGGT
TCTTAAAGGATAACTAGGATGTAGATAAGAAACAAACAACAACAACAACGaaaaatagggaa
gggcattgtaagtacaaggaagggtatgcaaaggcccagagacagaaatgagtcccacgtgtcttgggaCAATTAAT
AGATAGGCTTGGTAAGATGCAGGTCCAAGTTGGAGAAGAAGAAG

Base Numbers: 75,554,480 - 75,554,494
Number Repeats: 5
Nearest Gene: Between FLJ10661 & FLJ11099

ggcgtgcctgtaatcccagctactcgggaggctgaggcaccagaatcgcttgaacccaggaagtggaggttgcagtgagc
tgagatcacgcctctgcattccagcctcggtgacaaagccagactctgtctcaaaaaaacaacaacaacaacaaaaaaa
aaCTTAGCTGAAGATAAACACTGTAACCTTCTTATCTGTCAGTTGCCAGTAAAAAGAAAT
TCTAGAG

Base Numbers: 76,125,973 - 76,125,987
Number Repeats: 5
Nearest Gene: Between SKD3 & PDE2A

ctgaggcaggagaatggcatgaacccgggagacagagcttgcagtgagccgagatcgtgccactgcactccagcctgggc
gacagagcaagactccgtctcaaacaacaacaacaacaacaacaaaatgatgatagcctcttcctgaaactaacccccctt
tttgcaaggggaccaaaactgcctttataaaactttataaaactaacaaattgtccataagtttagaattatggcttagaa

gttatgaagccaaaggtcacaagatttgtaacctccccaattgctcctatagataacaccactattgtaaaacctaagacc
ggtatttgagatattttcagaccttgcattctgagg

Base Numbers: 76,182,612 - 76,182,632
Number Repeats: 7
Nearest Gene: PDE2A

GAGAAAATATGGAAACTAAAAGTTGAACCTAATCTTTATAGAGAAGTAAATAATAATT
GAGTGATTAACAACAACAACAACAAAAAGGGACTGTAATACACAGCAAAATAGGCTTAG
GAACTTGAAAAAGAGTTGTGGCCCAGAGTTTTTCTGAAATGCTGGT

Base Numbers: 76,570,470 - 76,570,484; 76,714,587 - 76,714,601
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

ctactaaaaatacaaaaaaagaaattagccgggcgtggtggcgggcgcctgtagtcccagctactcgggaggctgaggca
ggagaatggcgtgaacccaggaggcagagcttgcagtgagtcgagatcacaccactgcactccagcctgggcgacagagc
gatactctgtctcaaaacaacaacaacaacaacaaaaaaaccctgaagtctgaagaccccatgtttggtaattgaagcttc
cacgagagaaattcctgggaaatggtagggcagaaaTGGCAtttgcctagaactaagggatttcttgttatgtggaact
tcaacgctaaaactgaaacagtctcaagcaa

Base Numbers: 77,455,596 - 77,455,613
Number Repeats: 6
Nearest Gene: Between FOLR1 & PRCP

GAAGGATGAGAAGCAAATGTTGTTTCACAGCTTTTAATTCTTTAGCGAAGTCTCAAAAG
TATCATAGTATAAACATGTCAGGACCATGGTTGAGAAAAGTAACCAGTTGCAGTTAAaac
aacaacaacaacaacaacaaAAAATTTTGGAAAAAAAGGAATGGCAAACTCTAGAAGATCTCTCCC
AGAGGCTAATAGAGCAATGATCCCTAAGTACTCATTACTATG

Base Numbers: 77,706,098 - 77,706,115
Number Repeats: 6
Nearest Gene: PRCP

AAAGATTCTTTGGAATAAAACAACTACTCAGCAACAACAAcaacaacaacccactgagtgcctgtt
attggccaggttgtgtgcaaagtac

Base Numbers: 78,276,463 - 78,276,480
Number Repeats: 6
Nearest Gene: Between MDS025 & DLG2

gcaccactacactccagcctgggtgaaagagtgagatcccgtctcaaaaacaaaaCTAAACAaacaacaacaacaac
aacaaAGCATCCCATGTCTACAACAAGCTTTCGTGAACTCGTAGAAGCATGAAGGATTTTTT
AGCATAGGTGAGGCATCATTTCAACCTTTTACATGAGGAGCATTTAG

Base Numbers: 78,941,556 - 78,941,573
Number Repeats: 6
Nearest Gene: DLG2

TATCCCTTAAGAGAGTTTCATCTGTGGAATcaacaacaacaacaacaacaaAAGTATATTGTAA
AGGATGTTGAATACCCTCATAGGCAATTGGAGGTAGAACAAGACAGTGTAAGACAGGAC
ATTGGTAGT

Base Numbers: 79,260,026 - 79,260,040
Number Repeats: 6
Nearest Gene: DLG2

CTAAAAAGTTTAGGTTTTATTGTTTTGTCATTGTATATATAGAAGAGCCCTGaacaacaac
aacaacaacaaGGACATAGCCCAGGGATCTTCAATGAGCTACTAAATATATTGAAT

Base Numbers: 79,565,214 - 79,565,231
Number Repeats: 6
Nearest Gene: DLG2

CTGGAGCCTGATCAAGTATGTGTCTGTCCCTTCCAAACCATTAAAaacaacaacaacaacaacca
caaaaaccacaacCCACAAACCCATGTGAGGAGACCAAAGCCAAAGCCAAAGCCAGAGCCAGA
TTCTCAGTCTT

Base Numbers: 79,687,340 - 79,687,357
Number Repeats: 6
Nearest Gene: DLG2

CTCAAAAACTATTTGAGAGTTGAAGAAGTTTCTCAGAAATAATGTTGTAATAAACAACA
ACAACAACAATACTGTGGCCCTCTCATTGATTCATATTTTCTTATTTATCCCTTCTCCATA
GAACACTGCAAA

Base Numbers: 79,709,066 - 79,709,080
Number Repeats: 5
Nearest Gene: DLG2

gtttgagcccaggagttcgagacctgcctgggcaatatagcgagattcccttccacaaaaaggaagaaaaaacaacaaca
acaacaacaacaacaacaacaacaacaacaGATACAGAATTCACATTCTGGTAAAGAAAAGCAGATAT
AAACATCTCCAAAACAAGGGATATACTAGTGTTACACT

Base Numbers: 79,731,813 - 79,731,848
Number Repeats: 12
Nearest Gene: DLG2

gtctaaaaaaaaaaaaacaacaacaacaaccacaacaaaaagtgtcttttacctcctgccatgattctgaggcctccc

Base Numbers: 80,309,449 - 80,309,463
Number Repeats: 5
Nearest Gene: Between DLG2 & HT007

atcatgccattgcactccagcctgggcgacagagtgagactctgtctcaaaaacaaaaacaaaaCAACAACAACAA
CAGAAACCCTTAACTATCCATTCTTCAATCCAAGGAGTGATAAAATTGCTCTTGTAATAA
TGAGAAAGCTA

Base Numbers: 81,224,261 - 81,224,275

Number Repeats: 5
Nearest Gene: PAK1

ggctaacacagtgaaaccccatctctactaaaaatacaaaaacaaa**aacaacaacaacaac**aaaattagctggatgtgtt
ggcatgcacctgtagtccccgctactcaagaggctgaggcaagataattgctcctgggaggcagagg

Base Numbers: 81,234,377 - 81,234,391
Number Repeats: 5
Nearest Gene: Between PAK1 & MYO7A

aagactctgtctcaaaaacaaacagacaaacaaaaaacc**aacaacaacaacaacaacaac**aaaaaactcccttcaacaa
ctacatatatctgtgtaaagcctaattttctttaaatacttcaaccaaaacaacagattacaacagattaaa

Base Numbers: 81,632,578 - 81,632,598
Number Repeats: 7
Nearest Gene: PHCA

ATACTATGATGTAAAAATTATAGGGCTT**aacaacaacaacaacaac**aaAGTTTGGAGCAGATC
ATTTAAAACATATGGTCTTTGTAAGCACAGTTCTAACAAAAACAGCACCAATCC

Base Numbers: 81,697,315 - 81,697,332
Number Repeats: 6
Nearest Gene: Between PHCA & E2IG4

caaaataagtaa**aacaacaacaacaac**aaaaatacaaaAGTCACCCTACagtagagaggacattgaactaggaacc
aaaagatgagtctcactctgccactaatc

Base Numbers: 82,042,629 - 82,042,643
Number Repeats: 5
Nearest Gene: Between GL002 & PRKRIR

aagaggcggaagttgcagtgagccgagattgcgccactgcattccaacctgggcaacagagtgaggctccatctAAAC
A**aacaacaacaacaacaac**aaAATGCTCCAGCCAAAATTTCATTTGGCCTTCTCTGCAAGCAAG
AAGGTTTCTTGC

Base Numbers: 82,106,906 - 82,106,923
Number Repeats: 6
Nearest Gene: Between GL002 & PRKRIR

CACGAGCCTCATTCAAGTCCTCCAGGGCGacaat**aacaacaacaacaac**aaaaacCTGTGAAAAAA
ATGTTAATGTAAAATGCCAGCAGCATGCAAGCTGCCCCTATTTTTCCCCCATGATCGAGT
AAGGAAGTGAC

Base Numbers: 82,471,596 - 82,471,610
Number Repeats: 5
Nearest Gene: UVRAG

ccaatgggcaatttttggtcctcctatttgatcactctctcctcttaaaaaaacaagcaaaaagacaaacaacaat**aacaa**
caacaacaacaacaacaaaaccctttctttctctggactttgagattctacacaaaactttgctctctttcccttgggcc
acttcttttcagtcatc

Base Numbers: 82,638,581 - 82,638,601
Number Repeats: 7
Nearest Gene: UVRAG

gcttgcagtgagccgagatcgcgccactgcattccagcctgggcgacagagtgagagtgaggctttgtctca**aaacaacaa**
caacaacaacaaaaaccaaaaacaaaaacaaaacacaaatggctcagatgggttataacacacatcaggctggtcac

Base Numbers: 83,484,106 - 83,484,123
Number Repeats: 6
Nearest Gene: Between SLC21A9 & NEU3

ttgcactctagcctgggtgaacagaggaagactccatctcaaaaaacaaacaaacaaacaaacaa**aacaacaacaacaac**
aaaaaa*CAAGCTGACTGGATTTAGACTAAATACAAATGTAGGCAAATGC*

Base Numbers: 83,567,984 - 83,567,998
Number Repeats: 5
Nearest Gene: Between NEU3 & KIAA0102

ggagaggaacccttgagctcaggagttcaaggctagcctggggaacatggtgagacctcatctcttaaaa**aacaacaac**
aacaacaacaacaacaacaacaaaaaacagaa*Cagacctgggtcctggaccctgatttgtcacttattggctgtgtgag*
ttgggacagaccacttct

Base Numbers: 83,801,002 - 83,801,031
Number Repeats: 10
Nearest Gene: Between KIAA0102 & FLJ22596

ggctgagatcacaccactgcactccagcttgggcaacagagtgacatcctgtctca**aacaacaacaacaac**aaaaaatg
tcgtgcttcaaagcacaccatcaagaaagtacaaagacaatccacagaatgggagaaaatttttgcaaataatttttcaat
cagggtcttgtattcaga

Base Numbers: 83,933,935 - 83,933,949
Number Repeats: 5
Nearest Gene: Between DKFZP58N2124 & KCNE3

atcacttgagcctgggaggtcgaggctgcagtgagccgtgactgtgaccaatgcattcaatcctgggcaacagagcaaga
tcctgt**aacaacaacaacaacaac**aaaatttaaaaagtaaaaa*CGGAAACA*atgcagatggaaccatataacgtgta
gtttttgtgtatgacttct

Base Numbers: 83,953,943 - 83,953,960
Number Repeats: 6
Nearest Gene: Between DKFZP58N2124 & KCNE3

actccatcctgggtgacagagtgagactccatct**caacaacaacaacaac**aaacaacaaTCTCGTTCTGTTATT
TAAAGAAAACACTTCAAACAGTTAGAAATTTTGATTGTTtgtggcaggcactgtactgagag

Base Numbers: 84,129,965 - 84,129,979
Number Repeats: 5
Nearest Gene: FLJ14993

ACAATGACCATTTTACAaaacaaacaaacaaacaaacaa**aacaaCAACAACAACA**AAAAACagaagttaa
gtgatttgttcaaggtcccatagatagtgggatgccagaaactcaaacctgggtctacagacacaagtccagtgttctttc
tatgatatcatgctg

Base Numbers: 84,369,995 - 84,370,009
Number Repeats: 5
Nearest Gene: Between FLJ14993 & PHRETI

gcttgcagtgagccgagatcgcgccactgcattccagcctgggcgacagagtgagagtgaggctttgtctca**aaacaacaa**
caacaacaacaaaaaccaaaaacaaaaaacaaaaacacaaatggctcagatgggttataacac

Base Numbers: 84,770,048 - 84,770,065
Number Repeats: 6
Nearest Gene: Between RAB6 & THRSP

atcctgt**aacaacaacaacaacaac**aaaatttaaaaagtaaaaaCGGAAACAATGCAGATGGAACCATAT
AACGTGTAGTTTTTGTGTATGACTTCtgtttt

Base Numbers: 84,784,898 - 84,784,915
Number Repeats: 6
Nearest Gene: Between RAB6& THRSP

cactgcagcctgggcaacaaagcgagatcctgtctcaaaaaaaaaaaaccaaaa**aacaacaacaacaacaac**aaaaaaa
CGGAAAACTGCGTGCAGGCATTAATCTACACAAGCATTTCCTTCTGAACAATTGCCACAT
ACACCATTTT

Base Numbers: 85,145,617 - 85,145,634
Number Repeats: 6
Nearest Gene: Between RAB6& THRSP

aggccagcctggccaacatgacgaaatcctgtctctac**aacaacaacaacaacaacaacaacaacaacaac**aaagcag
ggtgtggtggcgagcacctgtagtcccaactacttgggaggctgaggtgggaggatcgcttgagtctggga

Base Numbers: 85,584,976 - 85,585,008
Number Repeats: 11
Nearest Gene: Between MGC2840 & GAB2

ACATTCAAAGAGTcaaa**aacaacaacaacaacaac**aaAACCCAGAATCGAACACATCATTTCCCT
CATTTCCTCATTCAGGAAGCTTCGGGTGGACCCTGAACttttttctttc

Base Numbers: 85,601,890 - 85,601,907
Number Repeats: 6
Nearest Gene:

Repeat (GTT)

ACTTTTGCCTGGTTAAATCTGCTTCTTAGGAAACATGGTAGGGAAttttttttttaattt**gttgtt**
gttgttgtttgagatagggtctcactctgtcaccctggctgaagggcaatggcacttcattgaattcatagttcactgcag
cctcaatctcct

Base Numbers: 73,485,410 - 73,485,424
Number Repeats: 5
Nearest Gene: Between SLC22A6 & HREV107

CACAGTTTTGTTGGTATTTTTTATAGGTTCTTTCAGATATTTAttttt**gttgttgttgttgt**ttttgtttttgagagagagacggtcttgcttgtggcccaggcttgagtacagtggcatgatcatagctccctgcagcctcgaactcctgggttcaagc

Base Numbers: 74,026,398 - 74,026,412
Number Repeats: 5
Nearest Gene: MRPL48

TCAAGTCTTTACAAAGCAACAAACTTAAt**gttgttgttgttgttgttgttgtt**TTAAATTACCAAGCTACCTTCAACAACCAAGGAg

Base Numbers: 74,093,090 - 74,093,113
Number Repeats: 8
Nearest Gene: Between LOC51287 & SERPINH1

AAACTT**GTTGTTGTTGTTGTT**TGTTTGTTTGTTGTTTTGCTTTTCTTAGGCATCTGAAAGCACTAGCAATCTAGCAATCTAGGTTCACTTTAATCCAATTTCCAAGCCATCCAAATGAC

Base Numbers: 74,129,041 - 74,129,055
Number Repeats: 5
Nearest Gene: Between LOC51287 & SERPINH1

ttgtgatttttt**gttgttgttgttgttgtt**aaaaactggccatttgaatctaataaggtggtgactctggacatcagattctccctctttctgaggt

Base Numbers: 74,467,949 - 74,467,969
Number Repeats: 7
Nearest Gene: Between KIAA0769 & UCP2

GAGTGAACAATTTT**GTTGTTGTTGTTGT**tatttttagtagagacaaggtttcaacatgttgatcaggctggtctt

Base Numbers: 74,577,423 - 74,577,437
Number Repeats: 5
Nearest Gene: KIAA0769

aatctcccttgaccaatgtccagaagcatttcccctatgttttcttgtactagttctatgttttcttgtactagttctatagtgtgggtgttttttt**gttgttgttgttgttgt**tttctttttttttttttcgagacagggtatcactctgtcacccaggct

Base Numbers: 74,613,714 - 74,613,731
Number Repeats: 6
Nearest Gene: KIAA0769

atctcccttgaccaatgtccagaagcatttcccctatgttttcttgtactagttctatagtgtgggtgttTGTtt**gttgttgttgttgttgt**tttctttt

Base Numbers: 74,617,478 - 74,617,495
Number Repeats: 6
Nearest Gene: KIAA0769

GAAACTT**GTTGTTGTTGTTGTT**TGTTTGTTTGTTGTTTTGCTTTTCTTAGGCATCTGAAA
GCACTAGCAATCTAGGTTCACTTTAATCCAATTTCCAAGCCATCCAAATG

Base Numbers: 75,244,782 - 75,244,796
Number Repeats: 5
Nearest Gene: Between PME-1 & FLJ10661

cagtagctgtgattaaaggcgcctgccaccatgactggctaattt**tgttgttgttgttgttgtt**ttgagacggtttcactc
gtcaccccagctg

Base Numbers: 75,694,752 - 75,694,769
Number Repeats: 6
Nearest Gene: Between NUMA1 & FLJ20625

GTTGAGATGCAGGAACACTATGATGAATTTTTTTGAGGA**gttgttgttgttgttgttgttt**gagat
ggagttttgctttgttgcccaggctggagtgcagtggcgctatctcggctcactgcaagct

Base Numbers: 75,820,692 - 75,820,715; 75,904,244 - 75,904,267; 75,934,321 - 75,934,344
Number Repeats: 8
Nearest Gene: Between FOLR2 & FOLR3; Between ARIX & SKD3; Between ARIX & SKD3

cCCtgttgttgttgttttt**tgttgttgttgttgtt**TGTTTGTTTGTTTTACCTTAAGATGAAGCTCTAA
AAGTTTGGGAGGCCCGTGTCTTGAGAATTCCCATGCAGAATCCTTTCCTTCTCTCCTT

Base Numbers: 76,179,326 - 76,179,340
Number Repeats: 5
Nearest Gene: PDE2A

CAACAttgtttt**gttgttgttgttgttgtt**gAAAGTGACATTTGAAGGCAGGTAGATTTTAGTGAA
AACATGATTAGGAAAGTAAAGTTTCTTCCCAGG

Base Numbers: 76,816,361 - 76,816,378
Number Repeats: 6
Nearest Gene: Between FOLR1 & PRCP

attccctgcctctgtcatcaaccctagaaattactgatcatcctcttttctctgtagttctagttttacagaattgcatata
aatgcaggtgattt**gttgttgttgttgtt**attgttgtgttgttcttgagacggagtttcactcttgtttccca

Base Numbers: 76,935,406 - 76,935,420
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

ATGTGttttt**gttgttgttgttgtt**tttgagacagtctcacttattgcccaggatggagtgcagtagtgcgatcacaggc
tcactgcagcctcgacttcctgggctcaag

Base Numbers: 77,695,173 - 77,695,187
Number Repeats: 5
Nearest Gene: PRCP

TTTCCAAATGCCATCCTGTTGGGTGACTTTTTTT**GTTGTTGTTGTTGT**TTTTTTAATGGAAG
CAGGGCACTGAAGACATGGTAAAGTCACATCAATTATATAACTGAAAATT

Base Numbers: 79,247,010 - 79,247,024
Number Repeats: 5
Nearest Gene: DLG2

AAGGtgttttt**gttgttgttgttgttgttgttgttgtt**ttgagacggagtctcagtctatcactcaggctgaagtgc
aatggcatgatctcagctcactgcaacctcccctccc

Base Numbers: 80,354,060 - 80,354,089
Number Repeats: 10
Nearest Gene: Between DLG2 & HT007

accatgaccagctttttgttgttgttgttgctcct**gttgttgttgttgtt**gagacagggtctcacaatgttacccaggt
tggtcttaaactcctggcctcaagtgatcctcccacctc

Base Numbers: 80,404,403 - 80,404,420
Number Repeats: 6
Nearest Gene: Between DLG2 & HT007

ttt**gttgttgttgttgttgtt**cttgctgcttgagaaggagtctcactctgtcacccaggctggagtgcagtggtgcaa
tgtcggctcactgcaacctttacctcccaggttcaagagattctcct

Base Numbers: 80,618,890 - 80,618,910
Number Repeats: 7
Nearest Gene: SYTL2

ttcccatccaagagtatgatgtatctttccatttgttcagatcttgtgttctcttttttt**gttgttgttgttgttgttgt**
ttttgagac

Base Numbers: 81,878,489 - 81,878,512
Number Repeats: 8
Nearest Gene: Between E2IG4 & GARP

tgctaggattacaggtataagccactgcgtcagcctattgtggTTTTtt**gttgttgttgttgttgttgttgttgtt**tttg
agacaggatctcattctgtcacccaggctggagtgcagtggtgggatctcagctcacgatagcctctacctcccagggtc
aagtgat

Base Numbers: 82,235,089 - 82,235,115
Number Repeats: 9
Nearest Gene: Between PRKRIR & WNT11

tttttt**gttgttgttgttgtt**ttttttttgaggcagggtctcactctgttgcccaggctggagtgcagtgatatgatgatctt
ggctcattgcagcctctgcctcccaagcttggctcattgcagcctctgcctcccagcttgaccgatcc

Base Numbers: 82,246,819 - 82,246,833
Number Repeats: 5
Nearest Gene: Between PRKRIR & WNT11

tccaggatattatgcaggtattcatataacaagcgatgaaatgtgtttctacaagatatttgttgttgttgttgttgttgtt
tttagatggaatctcactctgttgcccaggctggagtgcagtggtgcgatctcggctcactgaac

Base Numbers: 82,786,763 - 82,786,783
Number Repeats: 7
Nearest Gene: Between DGAT2 & FLJ22644

CGAAATTAGttttttttgttgttgttgttgttttgagatggagtctcactcgttgccaggctggagtgcagtggtgtg
atctcggctcactgcaacctccgccccccaggttcaagtgattttcctgcc

Base Numbers: 82,948,435 - 82,948,449
Number Repeats: 5
Nearest Gene: Between FLJ22644 & SERPINH2

ccagccTATATATATATATTttgttgttgttgttgttgagatggagtctcactctgttgcccaggctggagtg
cagtggcatgatggtagtgactgcaacctctgcctgcctgattcaa

Base Numbers: 83,034,549 - 83,034,566
Number Repeats: 6
Nearest Gene: Between SERPINH2 & PP1665

acagaactcctcctagcttttatttcctcatttgttgttgttgttgttgttttgagatggagtctcactcttgttgcccag
gctggagtgcaatggtgcaatctcggctcactgcagcctct

Base Numbers: 83,374,997 - 83,375,014
Number Repeats: 6
Nearest Gene: SLC21A9

TAAACTACATTGAGGAGttttttgttgttgttgttgttgtctgttctgagaagaggccttgctctgtcacccagg
ctggagtgcagtggcgctatctcggct

Base Numbers: 83,441,907 - 83,441,921
Number Repeats: 5
Nearest Gene: Between SLC21A9 & NEU3

atgctcattatctagaatattacctgatacatgataggtgctcaataagtttgttagatgaatgaatGAttttttgttgttg
ttgttgttcagatggagtttcgctctgtagcccaggccggagtgcagtggtgcgatcttggctcactgcagcctccacctc
ccgggttcaagcaattctcctg

Base Numbers: 83,673,371 - 83,673,385
Number Repeats: 5
Nearest Gene: Between KIAA0102 & FLJ22596

ggagctgggattacaggcacccgccagcactcagctgatttttgtggggttttgttgttgttgttgttttgta

Base Numbers: 83,674,632 - 83,674,646
Number Repeats: 5
Nearest Gene: Between NEU3 & KIAA0102

ACATACTTCAACTCAATAGAACATTTTTTT**GTTGTTGTTGTTGTTTTG**TTTTTGAAACAT
ATTATGttttatgttgttgttgtgt

Base Numbers: 84,570,662 - 84,570,676
Number Repeats: 5
Nearest Gene: PHRETI

atggtgtgctttgaagcacagacatttttt**gttgttgttgttgtt**tgagacaggatgtcactctgttgcccaagctggagt
gcagtggtgtgat

Base Numbers: 84,783,053 - 84,783,067
Number Repeats: 5
Nearest Gene: Between RAB6 & THRSP

CGTTTTCAGGATtt**gttgttgttgttgttgtt**tttaacttctattttaagctcagaagtacatgtgcaggtttgctat
acaggtgaattgcatgtcacagggcttgggtttacagattattttgtcaccca

Base Numbers: 84,939,354 - 84,939,371
Number Repeats: 6
Nearest Gene: Between RAB6 & THRSP

TTCTACGTATTGGGAACTGTACAAGTGGCTTTCATGtttttgtttttttgtttgtttgactggttt**gttgtt
gttgttgttgttgttgttgttgtt**tttgagacggagtttcgctcttgttgcccaggctggagtacaatggcgcgatctagg
ctaccacaacctct

Base Numbers: 84,996,592 - 84,996,621
Number Repeats: 10
Nearest Gene: Between RAB6 & THRSP

aatt**gttgttgttgttgttgtt**ttctagagacagcatttagccatgttgcccagtctggtcctgaactcctgagctcaagct
atctgcctaccctggcctcctaaagtgttgggattacaagcgtaagccactgcaactggccaagaatagtattttcaacaa
atcgtgct

Base Numbers: 85,061,805 - 85,061,822
Number Repeats: 6
Nearest Gene: CLNS1A

Ggttttt**gttgttgttgttgttgttgttg**TTTTAACACAAATTATGGCGTGAATTCACAGGGAATAG
GCTCCAGCAACTCAGGCTCCTTCCCACTGGTTCTTACACAGTGTGCTTCTCTGG

Base Numbers: 85,276,822 - 85,276,845
Number Repeats: 8
Nearest Gene: PTD015

GCCTTTTGAACATGATTACTATCATTGCTATTTTGGGTTT**tgttgttgttgttgttgttgttgttgtt**
tgagacagagtcttgctcttgttgccctggctggagtgcaatggtgcaatctcacagccactgcaacctccgcctcccgg
gttcaa

Base Numbers: 85,409,013 - 85,409,039
Number Repeats: 9
Nearest Gene: MGC16733

AGGCAGTGCCTGCttattttattatttattttatttttt**gttgttgttgttgttgtt**gagatggaatcttgctctgtt
gcccaggcggga

Base Numbers: 85,468,387 - 85,468,404
Number Repeats: 6
Nearest Gene: Between MGC2376 & PRO2849

ctacacctagccTGATGACTAtttt**gttgttgttgttgttt**gtttttgagacggagtctcactctgttgccaggctg
aagtgcagtggt

Base Numbers: 85,577,825
Number Repeats: 5
Nearest Gene: Between MGC2840 & GAB2

Repeat (AGG)

GGCCGAGGGCCTGGGGGTCCAGGGGGCT**gaggaggaggaggaggagg**aagaag**aggaggaggaggaggg**g
gCCGTGTCCCAGCTTGCCCGCCCAAGGGCCCAGGTAATGCAAATGCAGATTCTGTACAGGC
AG

Base Numbers: 73,354,938 - 73,354,955; 73,354,962 - 73,354,976
Number Repeats: 6, 5
Nearest Gene: Between SLC22A6 & HREV107

gggggaggggacgaggaggaggacccat**aggaggaggaggaggag**tggggtgagcggtggggccagagcaggcggaga
gggagAGCCGGGCAGTGCTG

Base Numbers: 73,356,761 - 73,356,784
Number Repeats: 8
Nearest Gene: Between SLC22A6 & HREV107

aaggaggaggaggaggagagaa**ggaggaggaggaggaggag**tggcaactggaggatgaataccagctgagtgaggagga
ggagaa**ggaggaggaggaggaggaggag**CAACTGGAGGATGAATACCAGCTGAGTgaagagaa**ggaggaggag
aggaggag**caactggaggatgaataccagctgagtgaagaga**aggaggaggaggaggag**CAACTGGAGGATGAATA
CCAGCTGAGTgaagaga**aggaggaggaggaggaggg**gaagaaggagaaggagaaggaggagggcaagaaggagga
ggaggaaggggaggCGACCGGAGG

Base Numbers: 73,678,196 - 73,678,213; 73,678,217 - 73,678,234; 73,678,234 - 73,678,254; 73,678,332 - 73,678,346; 73,678,381 - 73,678,395; 73,678,430 - 73,678,447
Number Repeats: 6, 6, 7, 5, 5, 6

Nearest Gene: Between HREV107 & MRPL48

TACCAGCTGAGTgaggaggaggaggagaaggaggaggaggaggaggaggaggaggaggaggaggtggcaac
tggaggatgaataccagctgagtgaggaggaggagaaggaggaggaggaggaggaggaggCAACTGGAGGATGA
ATACCAGCTGATTgaagagaaggaggaggaggaggcaactggaggatgaataccagctgagtgaagagaaggag
gaggaggaggCAACTGGAGGATGAATACCAGCTGAGTgaagagaaggaggaggaggaggaggaagaa
ggagaaggaggagggcaagaaggaggaggaggaaggggaggCGACCGGAGGATGAATACCAGCAAAGCC
AGGAAAAAGGATACCTCTTCTCAC

Base Numbers: 73,690,028 - 73,690,066; 73,690,110 - 73,690,133; 73,690,168 - 73,690,182; 73,690,217 - 73,690,231; 73,690,266 - 73,690,283
Number Repeats: 13, 8, 5, 5, 6
Nearest Gene: Between HREV107 & MRPL48

gaagcagaagcagaagaaggagaagaagaaggaggaggaggaggagaaggaggagaagaaggaggaggagaagGa
ggagaaggaggaggaaattgtttt

Base Numbers: 77,535,167 - 77,535,181
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP
gaaagaaagaaaggaagaaagaaaaataaagaaaaggaggaggaggaggagaaaaacaagaagagggaac

Base Numbers: 79,812,611 - 79,812,625
Number Repeats: 5
Nearest Gene: Between DLG2 & HT007

CACACTGACGTTTCGAGGGCACTACAGCCCGAGCAATAGGAGGAGGAGGAGGAGGTACAC
GTGCTGGGTGACTGTACATCCTATCATAATAATCCCACT

Base Numbers: 79,842,296 - 79,842,313
Number Repeats: 6
Nearest Gene: Between DLG2 & HT007

gatcatgtcactgtactccagcctgggcaacacagtgagaccctgtctcaagaaggaggaggaggaggaggaggCACA
Gtgatgatga

Base Numbers: 80,324,986 - 80,325,006
Number Repeats: 7
Nearest Gene: Between DLG2 & HT007

AGGGACAGGAGGAGGAGGAGGCCTGGAGGAGAAGACAGTGCTCTTTTGAAGAAGAGGGT
GGCCTGGACAAGATAGTGGTTATTTGTGTGTAGAAAGAGAAAGATTATAG

Base Numbers: 81,119,440 - 81,119,454
Number Repeats: 5
Nearest Gene: Between CLTH & PAK1

gagtgagaccctgtcttaaaaaaaaaaaaaaaGGAGGAGGAGGAGGAGAAATGGCCTTTTCAGTGTCT
atttccacctgccagcatctgtacctttttctcaggacttttggagggaccggggctgtcaacccatgtgtcaggctgga

catgctggagaattcatatcccccgagcagccctcagccaatgactgactcccctagccccttgcctctaatgaggggt
aactctggatgtgtgtt

Base Numbers: 84,013,413 - 84,013,427
Number Repeats: 5
Nearest Gene: Between DKFZP58N2124 & KCNE3

gaaaacaaaaggaagaagaaagaaaaaggaggaggagtaggcagaggaaaaggcggggggg**aggaggaggaggagg**aa
gaagaAAGAAGAAGAGAAAGGAGGCAAAAAGAAGGAAaaaagaagaaatac

Base Numbers: 84,436,444 - 84,436,458
Number Repeats: 5
Nearest Gene: FLJ14993 - PHRETI

Repeat (CCT)

CCTCCGGTCGcctccccttcctcctcctccttcttgccctcctccttctccttcttcc**cctcctcctcctcctcct**tct
cttcACTCAGCTGGTATTCATCCTCCAGTTG**cctcctcctcctcct**tctcttcactcagctggtattcatcct
ccagttg**cctcctcctcctcct**tctcttcACTCAGCTGGTATTCATCCTCCAGTTG**cctcctcctcctcctcc**
tcctccttctcctcctcctcactcagctggtattcatcctccagttgcca**cctcctcctcctcctcct**tctcctcctcctcc
tcctccttctcctcctcctcctcACTCAGCTG

Base Numbers: 73,647,640 - 73,647,567; 73,647,692 - 73,647,706; 73,647,741 - 73,647,755; 73,647,790 - 73,647,813; 73,647,857 - 73,647,874; 73,647,878 - 73,647,895
Number Repeats: 6, 5, 5, 8, 6, 6
Nearest Gene: Between HREV107 & MRPL48

ctccttcttcc**cctcctcctcctcctcct**tctcttcACTCAGCTGGTATTCATCCTCCAGTTG**cctcctcctc**
ctccttctcttcactcagctggtattcatcctccagttg**cctcctcctcctcct**tctcttcACTCAGCTGGTATTC
ATCCTCCAGTTG**cctcctcctcctcctcctcctcct**tctcctcctcctcactcagctggtattcatcctccagtt
gcca**cctcctcctcctcctcctcctcctcctcctcctcctcct**tctcctcctcctcctcACTCAGCTGGTATTCAT
CCTCCAGTTGCCACCACCTCTTCCTTCTCTTCCTCCTCAATGCAG

Base Numbers: 73,845,263 - 73,845,280; 73,845,315 - 73,845,329; 73,845,364 - 73,845,390; 73,845,413 - 73,845,439; 73,845,483 - 73,845,521
Number Repeats: 6, 5, 5, 9, 13
Nearest Gene: Between HREV107 & MRPL48

CCCCTCCTCTCCGGC**CCTCCTCCTCCTCCT**GGTCCTTGCTAACGCTGCCGTCGG

Base Numbers: 74,356,049 - 74,356,063
Number Repeats: 5
Nearest Gene: KIAA0769

CCTCACCACAGACCAGCCGGAGCC**CCTCCTCCTCCTCCT**GCCACTTCACCCTCA

Base Numbers: 74,679,589 - 74,679,603
Number Repeats: 5

Nearest Gene: Between KIAA0769 & UCP2

GGCCAATTATTTGTGTATATTTTTctctt**cctcctcctcctcctcctcc**ACCATCTTCTTTTTTCTGT CCCTTAT

Base Numbers: 77,017,352 - 77,017,369
Number Repeats: 6
Nearest Gene: Between FOLR1 & PRCP

AAAGACTTACTCTGTAAATTACATTTCTATTCATCATA**cctcctcctcctcctcc**tttcttctgctcctt cttctcctccttct

Base Numbers: 77,376,957 - 77,376,971
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

TCTACTGCAATAAcatcagcagcatcatcgtcatcatcaccatcaccaACT**cctcctcctcctcctcctcctcc**cctTTCTGCCAAC

Base Numbers: 79,222,004 - 79,222,024
Number Repeats: 7
Nearest Gene: DLG2

cctttcacttctctttctttttctttctccttccacatctttgcctcttctt**cctcctcctcctcctcc**tttatcttttccttctCT TGTATTCAGAAGACA

Base Numbers: 79,392,422 - 79,392,436
Number Repeats: 5
Nearest Gene: DLG2

ATCT**CCTCCTCCTCCTCCT**AATTGATTTAATCTTCTCATACCAGAGAGTGGGGCCGGGTTT GGGGAGATGGTGTTGATCAGAA

Base Numbers: 83,311,146 - 83,311,160
Number Repeats: 5
Nearest Gene: Between ARRB1 & SLC21A9

catatagGTAcctgct**cctcctcctcctcctcc**ttcctccttctcATTACTATTGTGATTATTATTCTTAA ATGAATGTTTAGGATTACTGtgtgtttctca

Base Numbers: 83,357,574 - 83,357,588
Number Repeats: 5
Nearest Gene: SLC21A9

GCTAATTA**CCTCCTCCTCCTCCT**TCAGACCTCAGAGACACCTTTTACCTCTCTGACTAGGT CACATGACCCTGGTACTTCATGCACCTCTCTTG

Base Numbers: 83,676,246 - 83,676,260
Number Repeats: 5
Nearest Gene: Between KIAA0102 & FLJ22596

AACCAGATCT**CCTCCTCCTCCTCCT**AATTGATTTAATCTTCTCATACCAGAGAGTGGGGCC
GGTTT

Base Numbers: 84,765,219 - 84,765,233
Number Repeats: 5
Nearest Gene: Between RAB6 & THRSP

Repeat (AAG)

ACTGAAGGGTAGATACGGGAGGAACAAATGTGAGAAAGAAGTGATgaaggaa**aagaagaagaa
gaagaagaagaagaag**aTGATGATGATGGTGATGatggtagct

Base Numbers: 78,444,111 - 78,444,137
Number Repeats: 9
Nearest Gene: DLG2

GTCACACAAGATTGTCATGCGAGCTGAAGGGGGCTGAAATTATTCCCTCT**gaagaagaagaag
aagaag**aaAACATACTGCAGCTCCCGATTTGGTAAAAAGCCAGTGGCTTGCT

Base Numbers: 79,541,890 - 79,541,907
Number Repeats: 6
Nearest Gene: DLG2

AATCACTGGATTAAAAAAAAAAAA**GAAGAAGAAGAAG**GAGGGGGGAATGAGCTCACTC
AGGAAGGAG

Base Numbers: 79,744,482 - 79,744,496
Number Repeats: 5
Nearest Gene: DLG2

ggaggtcaaggctgcagtgagcttcactccagcctggccaacagaatgagacactgtctcaaaaaaaaaaaaaaaa**agaa
gaagaagaa**GAAGTTCAAAGAAAAAGAAAGAAAAAGTCACTGAAATTAATGAT

Base Numbers: 85,260,385 - 85,260,402
Number Repeats: 6
Nearest Gene: PTD015

aaaa**AAAAGAAGAAGAAGAAG**AAGGAAAAAATCTTGAATATAAATGGATTAAGTTCTTC
AATTAAAA

Base Numbers: 85,356,824 - 85,356,838
Number Repeats: 5
Nearest Gene: Between PTD015 & MGC16733

Repeat (CTT)

AACATGAAGTTCTCTTTCCTAAC**tttctttgtttccttattt**cttcttcttcttctctc**cttcttcttcttctt
cttcttcttcttcttcttcttcttcttcttcttcctcttcttcttcttcttcttcttcttcttcttcttcttcctttcct**

tcttctttttttgctgagacagggtctcactctgacagtacagtggtgccatcacagctcactgcagcctcgacctccagg
gctcaagcaatcctcccagctcaccctcccaa

Base Numbers: 75,732,016 – 75,732,030; 75,732,033 – 75,732,080; 75,732,084 – 75,732,104
Number Repeats: 5, 16, 7,
Nearest Gene: Between DKFZP564M082 & FOLR3

GGCCCAGtcttcttcttcttcttctttttttttttgagaccgagtctcattctgtcacccaggctggaatgcaatggc
acagtctcagctcactgcaacctctgcctcccaggttcaagcgattctcc

Base Numbers: 75,866,488 – 75,866,505
Number Repeats: 6
Nearest Gene: Between FOLR2 & INPPL1

ctaggatcttcttcttcttcttcttcttcttcttcttcttttttttttttttt

Base Numbers: 76,107,319 – 76,107,357
Number Repeats: 13
Nearest Gene: Between SKD3 & PDE2A

CTATAAAATTAATCATTTTTGTTTTGGATCTCTTGAGAGAGAATAATTttttccttcttcttct
tcttctacttcttcttcttctAAAtagagacagtgttttgctatgttgcccaggctagtcttgaactcctgggctcaa

Base Numbers: 78,909,269 – 78,909,283
Number Repeats: 5
Nearest Gene: DLG2

cctcccatatccccctcctcccctcctcctccctcttcccctcccctttccccCtcctcttcttcttcttcttcttttc
ttcctttctccttcttctttt

Base Numbers: 81,491,566 – 81,491,586
Number Repeats: 7
Nearest Gene: Between CAPN5 & PHCA

CCAGGGCCCAGCATCTTGGTACAGTCCTAGttcttcttcttcttcttttttttttgagacagacttttgct
cttgttgctcaggctagagtgcaatggcacaatctcggctcactgcaacc

Base Numbers: 83,281,737 – 83,281,757
Number Repeats: 7
Nearest Gene: Between ARRB1 & SLC21A9

GCCCCTCGACTTCCACGTCATGGGCAATGTTGTACATGtcttcttcttcctcttcttcttcttcttcctc
ttcctcttcttcctcttcgtcttcctcttcttcttcttctccttctccttcctcttcctctt

Base Numbers: 83,398,401 – 83,398,415
Number Repeats: 5
Nearest Gene: Between SLC21A9 & NEU3

acaattctccataa**cttcttcttcttcttcttcttcttcttcttcttcttcttcttct**attattattattattattatta
ttagagacagtctcactctgttgcccaggct

Base Numbers: 85,054,880 - 85,054,927
Number Repeats: 16
Nearest Gene: Between RAB6 & THRSP

TTT**Cttcttcttcttct**ttttttttttttttttgagatggagtcttgctctgtcgcccaggctggagtgcagtggtgcaat
ctcagct

Base Numbers: 85,461,054 - 85,461,068
Number Repeats: 5
Nearest Gene: Between MGC2376 & PRO2849

Repeat (ATC)

TATTTTACAAAGTGCTACATAATGTAATGGATATTGCCATATTCTt**catcatcatcatcatcat
catcacc****atcatcatcatcatcatc**accaccatcTTTCTTAGAGAAGGATGATAGTATAATGTCTTGA
GCT

Base Numbers: 74,476,098 - 74,476,118; 74,476,139 - 74,476,156
Number Repeats: 7, 6
Nearest Gene: KIAA0769

ATATCACTACCACCACAATGACTACACAAGGG**ATCATCATCATCATCATC**CTAATATGTA
ATAAGCTGCTactaatttatccttaaattgttataagataggtactattagtcctactttacaaataaggatagtgta
tctaagcaaaaataacttgactaagattgcataactaatTTAACATTTAACTCCTGAGTCTGTGA

Base Numbers: 79,425,063 - 79,425,080
Number Repeats: 6
Nearest Gene: DLG2

ACTGGAAATGGAAGAAGATGGTCTCCCTGTTTCTGTATCTCTC**ATCATCATCATCATCAC**
CGTGTTCTGTCTATACTTCTTATGG

Base Numbers: 81,003,400 - 81,003,414
Number Repeats: 5
Nearest Gene: Between CLTH & PAK1

GGCCTAAAGTGACTGCATCCAGATTt**catcatcatcatcatcatc**gt**catcatcatcatcatc**accatcatc
atcat**AGCATCTGGCAAATACTCTGTGGCAAGCAGGGTGCTAGG**

Base Numbers: 84,327,374 - 84,327,391; 84,327,395 - 84,327,409
Number Repeats: 6, 5
Nearest Gene: Between FLJ14993 & PHRETI

caatatagcaagaccccatctcaagaaaaataaaaa**aTCATCATCATCATC**TTTTTAAATGTCTAAAAC
TAAAAAATATGGCCACATAC

Base Numbers: 84,475,703 - 84,475,717
Number Repeats: 5
Nearest Gene: Between FLJ14993 & PHRETI

Repeat (GAT)

caacagagcgagactctgtctcaaagtaataataataataat**gatgatgatgatgat**aTAGGTGTATCAGAGAA
TATCAAGTCCCTAGGAATAAATCTACCAGGTCTTCAAGATCT

Base Numbers: 76,147,185 - 76,147,199
Number Repeats: 5
Nearest Gene: Between SKD3 & PDE2A

TAAGTTATATGTTCTAGCATataatgt**gatgatgatgatgatgat**aatataatgaGTAGAATCAATGA
AATAGGAAAAGATATAAGGCAAATGTTATATACGTTATATGtaa

Base Numbers: 77,952,993 - 77,953,010
Number Repeats: 6
Nearest Gene: Between RAB30 & PCF11

TTCTTATAGTAAATACTGATTTTt**gatgatgatgatgat**ggtgatgatgacaatgaCAATAGTTAAA
TAGGACAATGAAAACAATTTCAAGTTTTTCATTGTTCACCAAACTTTGCACAGCCCCTGA
TGCATAATACACT

Base Numbers: 79,781,848 - 79,781,862
Number Repeats: 5
Nearest Gene: Between DLG2 & HT007

GCCAGAACAACCAGCAGGTATGAAGTTTTTTCCTAGAAAATTTTTT**GATGATGATGATG**
ATAAAGACAATATTAAaatagcaacattagttgagtgtgtttaatgctctaa

Base Numbers: 80,827,894 - 80,827,908
Number Repeats: 5
Nearest Gene: CLTH

CAGTGCCACCAGTTTCAGAAgatgag**gatgatgatgatgatgat**gCTACCCCACCACCAGTGATTGC
TCCACGCCCAGAGCACACAAAATCTGTGAGTCTTTGGGGACCTGGCCAGTTTTGTGATTTA
TTAGAGTAAG

Base Numbers: 81,170,564 - 81,170,581
Number Repeats: 6
Nearest Gene: PAK1

ctggagtccactgtccagtgggcaagacaaacACGTAAAatgacgacgacaat**gatgatgatgatgat**gaGGCTT
GCCTTAgagcacttaccacccgccattgttctgagtgatttacacagtattgactca

Base Numbers: 81,905,102 - 81,905,116
Number Repeats: 5
Nearest Gene: Between GARP & GL002

gtgggtatgaatcctggctctgctgggtgtcctttggtaagtgggtcaacctctctgtgcttcagagttctcatctataaa
atggg**gatgatgatgatGAT**ACTTtggggatgtaacatggtgcagtcacattggaaaagtttggcagttcctcaaaag
gttaaacataaaagttaccat

Base Numbers: 83,048,912 - 83,048,926
Number Repeats: 5
Nearest Gene: Between SERPINH2 & PP1665

GGGTTGAGGAATGAAAT**GATGATGATGATGAT**GACAGCCAACATTAAGCATTTACTGTA
TACTAGACTCCAGTATAAAGGCTTTATCTACATAAGAACCACttatttctactttacagataaggaa
acccagattaagtaatttgcctaatatcacatggttagtcaggac

Base Numbers: 83,606,605 - 83,606,619
Number Repeats: 5
Nearest Gene: KIAA0102

TCCTGTAATAGCTACAGTCAAATGGGCAGCTGAA**GATGATGATGATGATGAT**CTTGACAC
CGAGAAGCAGAAGACCAATGAAGATGACCAGACAGCAAAAAAGGAT

Base Numbers: 85,549,763 - 85,549,780
Number Repeats: 6
Nearest Gene: MGC2840

Repeat (AAT)

caaaa**ataataataataataataat**aaaataaagatttgatgtcataaagaaattagtgctccccaaacaatctaaaaa
tccaaatgcaattctaatacaaatcttaatatctggtggtgagaac

Base Numbers: 72,548,177 - 72,548,197
Number Repeats: 7
Nearest Gene: FLJ23392

ccagcctgggtaacacagtcagaccccgtctctaca**aataataataataataAAT**ATTATATAAAGTATggttg
ggcacagtggctcacgcctgtaatcccagcacttt

Base Numbers: 72,548,535 - 72,548,552
Number Repeats: 6
Nearest Gene: FLJ23392

aaa**aataataataataataat**aaaataaagatttgatgtcataaagaaattagtgctccccaaacaatctaaaaatccaa
atgcaattctaatacaaatcttaatatctggtggtgagaactttgtaagatgac

Base Numbers: 72,632,932 - 72,632,949
Number Repeats: 6
Nearest Gene: Between FLJ23392 & FLJ20556

cctgggtaacacagtcagaccccgtctctaca**aataataataataataat**attatataaGTATggttgggcacagtgg
ctcacgcctgtaatcccagcacttt

Base Numbers: 72,633,287 - 72,633,304
Number Repeats: 6
Nearest Gene: Between FLJ23392 & FLJ20556

atcgcgccagtgcactccagcctgggcaacagagcaagactccatctca**aaataataataataataataataataataat aataataat**TATTATTATTATTATTAcctatcatctcggcattctgggaagctgagggaagaggatcacctgag gccaagagttcgagaccagtctgggcaacacagtgagaccccaactctacaaaaaaaaaaaaaaa**aaTAATAATAAT AATAAT**Gtataagtaacaatcatagtatgctaaagacacagggatggg

Base Numbers: 72,633,964 - 72,634,002; 72,634,131 - 72,634,148
Number Repeats: 13, 6
Nearest Gene: Between FLJ23392 & FLJ20556

cctcagaaaaacacccaac*AAGCC*TAT*CAACCAA***aataataataataataataataataataa**ATAAATAa ataagtaataattttaaaaagcacccaggaccaaacagaatcacagctgaattctactagaag

Base Numbers: 72,956,772 - 72,956,798
Number Repeats: 9
Nearest Gene: Between HRLP5 & SLC22A8

cagtatcgcaccattgcactccagcctgggggataagagtgaaacttcgtctcaa**aataataataataataataat**Gag ctgggtgtggtggcacatatcccagctacttg

Base Numbers: 73,426,246 - 73,426,266
Number Repeats: 7
Nearest Gene: Between SLC22A6 & HREV107

acctgcctctaaaa**aataataataataat**aacaataataataataATggccgggcgaggtggctcatgcctgtaatcc cagcactttgg

Base Numbers: 73,670,696 - 73,670,710
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

gggactctgcttcaaaaagaaaaaaagtgagaaaaaca**aaTAATAATAATAA**tgttatatataaaatgtttattta gaaacagaatacttgttcgtcggtact

Base Numbers: 73,696,199 - 73,696,213
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

attgcactccagcctgggggataagagtgaaacttcgtctcaa**aataataataataataataat**Gagctgggtgtggtg gcacatatcccagctacttgacttgggaggctgaggcaggagaattgctt

Base Numbers: 73,770,597 - 73,770,617
Number Repeats: 7
Nearest Gene: Between HREV107 & MRPL48

ccagcctgggcaaagaagcaagacctgcctctaaaa**aataataataataat**aacaataataataataATggccgggcg aggtggctcatgcctgtaa

Base Numbers: 73,817,489 - 73,817,503
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

tgacaggggggagactttatctcaa**aaTAATAATAATAAT**ATGCCAGGAGCAATCGGCGcgcctgtaat
cccagcactttgggaggctgaggtg

Base Numbers: 73,980,949 - 73,980,963
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

taaaaaaacaaa**aataataataataataataat**tctaaaatgtcagccaccttagtattggcatctacatattgtcattt
tt

Base Numbers: 74,262,573 - 74,262,593
Number Repeats: 7
Nearest Gene: Between SERPINH1 & P2RY2

cggcgaaacccttctctaaaaaaaa**aataataataataat**acagaaaaactagctgggcatggtagcacacacctgtag
tccctgctactcaggaagctgaagtgggaggat

Base Numbers: 74,596,007 - 74,596,021
Number Repeats: 5
Nearest Gene: KIAA0769

tccagcctgggcgacagagcgagactctgtctcaaaaataataataa**aataataataataataataataataataa
taataataat**ggctggacacagtggc

Base Numbers: 74,773,736 - 74,773,777
Number Repeats: 14
Nearest Gene: Between KIAA0769 & UCP2

aac*AGCAGAGA*Ttagtagtagtaataataataatgat**aataataataataataataat**aacaat*GGATTATTA
GTAC*

Base Numbers: 74,969,017 - 74,969,037
Number Repeats: 7
Nearest Gene: Between UCP3 & PME-1

aaaacttaaagtat**aataataataataataataataat**aa*AAGTTTGGGAC*aaaaaaaataaaaaataaataaaat*G
CAGCTT*

Base Numbers: 75,141,994 - 75,142,017
Number Repeats: 8
Nearest Gene: Between PME-1 & FLJ10661

cattaagttaca**ATTTAATAATAATAATAATGC**TTTAACAATAATGAAGTAATGTTATAATC
TTAATAACAAAAGTA

Base Numbers: 75,295,962 - 75,295,976
Number Repeats: 5
Nearest Gene: Between PME-1 & FLJ10661

agcctgggctacaagagcgaaactccgtctcaa**aataataataataataataa**taaACCACATCACACCCACCA
CAAACCAGCTG

Base Numbers: 75,402,505 - 75,402,525
Number Repeats: 7
Nearest Gene: Between FLJ10661 & FLJ11099

ggcaacagagcgagactctgtctcaaagt**aataataataataat**gatgatgatgatgataTAGGTGTATCAGAG
AA

Base Numbers: 76,147,170 - 76,147,184
Number Repeats: 5
Nearest Gene: Between SKD3 & PDE2A

tctcTATtaataataatgatgat**aataataataataata**GCAGGGGGTGCAGACTCTGGGGGTGATGTA
GAGTC

Base Numbers: 76,343,055 - 76,343,069
Number Repeats: 5
Nearest Gene: Between CENTD2 & FOLR1

tctacaaaaaatacaa**aataataataataataataat**aacaggcatggtgtcgcatcctgtagtctcagctactca

Base Numbers: 76,452,028 - 76,452,048
Number Repeats: 7
Nearest Gene: Between FOLR1 & PRCP

GTTAATCCATAGGAAGTCAGGCt**aaaaataataataataat**aaCCATGCATAGACATCAGGAGCC
ACAGCAGTTTCTGCAGTTTATGTCTGGACAA

Base Numbers: 76,927,820 - 76,927,834
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

aaacttctaggcttaagcaatcggtctgcctcagccacccaatgtgctgggattacaggcatgagccaccacacccagcc
aggttcagtttttaatctattaaataggg**aataataataataATAAAGAATAATAATCTATTAAATAGA**
AATA

Base Numbers: 77,123,392 - 77,123,406
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

aggcggagcttacagtgagccaaaatcgcgccactgcactccagcctgggcgacagagcaagattccgtctctaac**aata
ataataataataat**aagcatttgattaaattcattatatatattcatgataaaatctcctagcaaaccaggaatagagcca
aatttccaaattt

Base Numbers: 77,156,213 - 77,156,230
Number Repeats: 6
Nearest Gene: Between FOLR1 & PRCP

ACAGGAAAGTtaataataataataataataataataataataatagtaataGCAACTTGCTTCATTT
TGGGAGACTAAAAGTAAGCTG

Base Numbers: 77,171,742 - 77,171,777
Number Repeats: 12
Nearest Gene: Between FOLR1 & PRCP

CTGTAAAACCTGCTCTTGTACACataataataataataataataataataacaataataaATGGAGAAGT
GGGTAGAATACACCATGTACTTTGTCCCTAACTAGGAA

Base Numbers: 77,830,564 - 77,830,584
Number Repeats: 7
Nearest Gene: RAB30

catctcatttaatcctcacaactatgcaatgaggtaataataataataataataataaataatgactaccatttattgagcatc
ttttacatatagactgaacactat

Base Numbers: 78,338,569 - 78,338,586
Number Repeats: 6
Nearest Gene: DLG2

ACAGGACCCTCACaataataataataataataaACAAAGTCACAAATGTTACTTTCAGGTTTGAA
TTCCACTTGAATTCATTGAGTTA

Base Numbers: 78,466,832 - 78,466,849
Number Repeats: 6
Nearest Gene: DLG2

gaaatcaagtaattaataataataataataatgtaggctaatttctataactttggaaggatggaggaggtcccctcacccat
gctgagaac

Base Numbers: 78,689,106 - 78,689,120
Number Repeats: 5
Nearest Gene: DLG2

ATATGGTGTCCTTCATTtaataataataataataataataaAAGATTTGCTCTCTGGTAGACAGGA
GAGCGGTAATTTAGCTACAATTT

Base Numbers: 78,787,761 - 78,787,781
Number Repeats: 7
Nearest Gene: DLG2

aacaaatacaataataataataataataataataataataataaccacatgcactagtgcatgcctgtagtcccagcta
tttgagaggctgagatgggaggatcacttgagccca

Base Numbers: 78,798,119 - 78,798,151
Number Repeats: 11
Nearest Gene: DLG2

cgagactccatctcaaaaaaagaa**aataataataataat**aaattaaataaGTCTGTCAGCTTTCTGGCAAAT
TATCTTCCAGTGTTTCCCCACCCTC

Base Numbers: 78,799,934 - 78,799,948
Number Repeats: 5
Nearest Gene: DLG2

gcaacagagtgagattccatctca**aataataataataataat**attaaGAAGCAGCATTCTCAATTAATGCT
ATCTATGTAAAGTGCTAAGTAGCAAAGTTTCTTTGTTTCCTCCAG

Base Numbers: 79,042,516 - 79,042,533
Number Repeats: 6
Nearest Gene: DLG2

ttcatctcaaaataac**aataataataataat**aaaataaagtggtaatacatactttgaagtacacatatgaaaattaaata
atacatgATAATCCTTAACATGTAGTAGTAGTTATCATTTTACTAGTCATAAATGTGGATT
TGCTC

Base Numbers: 79,159,865 - 79,159,879
Number Repeats: 5
Nearest Gene: DLG2

gtaacaaacctgcatgttgtgcacatgtaccctagaacttgaagtattat**aataataataataataataataat**aaAGA
AACTTTGTAAACTTTGTAAACAGATTATATTTCCTGTTTCAGCTCTAA

Base Numbers: 79,362,242 - 79,362,265
Number Repeats: 8
Nearest Gene: DLG2

tctgtctc**aataataataataataat**gaaggaaataaagtcaaggaaataagatgctccaagagagcacaagaa

Base Numbers: 79,375,798 - 79,375,815
Number Repeats: 6
Nearest Gene: DLG2

AAGTTAAAAGAGTCAATGTTTTCCaa**aataataataataat**actataattaaGGAAGTATAACCAT
AATTACATGTTGACAATAAATTATGGGACAGTTATAGACATAAAATTG

Base Numbers: 79,403,607 - 79,403,621
Number Repeats: 5
Nearest Gene: DLG2

ttgatcagcatagctgacctctgcaaaaaaaga**aataataataataat**aaaaattagctgggtgcagtagcatgcactca
tag

Base Numbers: 79,498,775 - 79,498,789

Number Repeats: 5
Nearest Gene: DLG2

TCTGTGATCACAGTATataataac**aataataataataataataat**TATTATTATTATTTTTGCCCTT
AGACTGTATGGGGGCTCTTCCAGG

Base Numbers: 80,504,378 - 80,504,398
Number Repeats: 7
Nearest Gene: Between DKFZp586C1924 & ZF

catctctaaaaaaacaaa**aataataataataataataat**tctaaaatgtcagccaccttagtattggcatctacatattg
tcatttttcatttagtttacaatcttcctggttcttggtatgatgagtgatttattgaaacctggaca

Base Numbers: 80,703,186 - 80,703,206
Number Repeats: 7
Nearest Gene: Between SYTL2 & CLTH

TCACTTAGTCTCTTGTACCaa**aataataataataat**ttttttaaaGTAAATGATACATATCTGAAGT
TAAATGTACAAATAACAAAATAAAAAACAAAAGGACATGAAAAAGACATTAACTGTGT
A

Base Numbers: 80,793,801 - 80,793,815
Number Repeats: 5
Nearest Gene: Between SYTL2 & CLTH

gctatgattgtgccactgcactccagcttgattgacagaatgagaccatgtctct**aataataataataataat**ttttaaaa
taattaacaagatcataaagggcttcatctaataacataaagaa

Base Numbers: 81,048,456 - 81,048,473
Number Repeats: 6

Nearest Gene: Between CLTH & PAK1
gactctgtctct**aataataataataataat**acaaagaatacaaaaaagcatatagaataaagctataaagaaaatattgt
atagctgtac

Base Numbers: 81,113,532 - 81,113,549
Number Repeats: 6
Nearest Gene: Between CLTH & PAK1

tgtatcaaaa**TAATAATAATAAT**GTAGCAGCTGCTGAGCTTTGTTTGCTTAGAGGCAGCTAT
AGGGGCCCA

Base Numbers: 77,745,240 - 77,745,254 ; 81,117,084 - 81,117,098
Number Repeats: 5
Nearest Gene: PRCP; Between CLTH & PAK1

ATCAGATACCAAAAACACATCAAGTATG**AATAATAATAATAATAAT**tttttttttgagac

Base Numbers: 81,237,475 - 81,237,492
Number Repeats: 6

Nearest Gene: Between PAK1 & MYO7A

CTAGGGAAA**aataataataataataataataa**Ggcaaataattcaaaagcacattaaaaggatcattcaccatgat caagtaagat

Base Numbers: 81,450,353 - 81,450,373
Number Repeats: 7
Nearest Gene: CAPN5

tgggcacaagagcaagactctgtctcaaaa**aataataataataataataataataataataat**ggaaagccaatgaagg tttcaccaatagataacatggtcagactcatgtatt

Base Numbers: 81,533,096 - 81,533,128
Number Repeats: 11
Nearest Gene: Between CAPN5 & PHCA

tctgcaaaatggagctgctgct**aataataataataataataataataat**atctaccttcttgggtgtag

Base Numbers: 81,853,705 - 81,853,728
Number Repeats: 8
Nearest Gene: Between E2IG4 & GARP

ccactgtactccagcctgggcaacaggagtgagaccctgtctcaaaa**aataataataataataa**aaataataataataat **aata**GCAGCACAGAAATGTtaaaacgatgggaaaatatactaggcaaatgctaagcaaaataatgcaatattgac agcttttaag

Base Numbers: 82,133,197 - 82,133,211; 82,133,214 - 82,133,231
Number Repeats: 5, 6
Nearest Gene: Between GL002 & PRKRIR

ggtgacagagcaagactcctctcaaa**aataataataataataataataataat**aactaagagtataactggattgtttgtaac acaaaggatgaatgcttgaggtgatggatacccatttaccctgatgtgattgctatacattgtatgcc

Base Numbers: 82,135,610 - 82,135,633
Number Repeats: 8
Nearest Gene: Between GL002 & PRKRIR

CTCACTGAAAACAACTAAAATGTCAGGtaaaatatttttttt**aaataataataataat**CACCATTAAC ATCAAAGAATTTCCAAACCAAGGAAGAAAGAACCCAGAGAGGTAAGTGGAGATCT

Base Numbers: 82,653,348 - 82,653,362
Number Repeats: 5
Nearest Gene: UVRAG

aacacccaagaatgatcaataaaaaaataaataaataaaT**AAATAATAATAATAATAAA**taaaataaaaata aaaattagtcagagatg

Base Numbers: 83,120,336 - 83,120,350
Number Repeats: 5
Nearest Gene: PP1665

gtgagatcgtatatctaaa**aataataataataat**aaaaattagtcaggcatggtggcccatg

Base Numbers: 83,141,306 - 83,141,320
Number Repeats: 5
Nearest Gene: Between PP1665 & RPS3

gccactccactccagcctgggtgacagagcgagactccatctcca**aataataataataataat**gataataataatCTG
GGGGAAAAACATCATAAATTGTGTTTTCTACCCAAGACTGTAGGGAGACAGACAGTTTA
AA

Base Numbers: 83,226,502 - 83,226,519
Number Repeats: 6
Nearest Gene: Between RPS3 & ARRB1

ccactgtgcccagcccattgtcttattttaagaaattatctatgtaagtaagaaatgcttctagaaa**aataataataataa
taataataataataat**acactaaatcttcagcaaccaccaccctgataagtcagcagtcaccaac

Base Numbers: 83,533,699 - 83,533,728
Number Repeats: 10
Nearest Gene: Between SLC21A9 & NEU3

tgggtgacagagtgagactccgtctcaaaaaaaaaaaa**ataataataataataataat**aaaaaaaaGATTCCATCA
GTCCCTGGGCAGGTTTTGGCCACATGTTTACTCTGGGCCTCTCTCTCTCCGCTGCCCCCAAC
TAAGTGCCG

Base Numbers: 83,828,924 - 83,828,944
Number Repeats: 7
Nearest Gene: Between KIAA0102 & FLJ22596

gagattgtgccactgcactccagcctgggtgacagagtaagactccatctca**aataataataataataataataat**aaa
ataa**AGACACCATCAATTATTTACCCTGCCCACATGTACAAATCCTTATAATTTCAAAGAT
GTTTTAATGTCAAT**

Base Numbers: 83,834,039 - 83,834,062
Number Repeats: 8
Nearest Gene: Between KIAA0102 & FLJ22596

ccagcctgggcaacaagagcttaatccgtctcaaaaaa**TAATAATAATAAT**AGGCCTCTCCAAACCTAG
GGTTTGTTTTATTGATGGTTATTCACTTCATAGGGGCAAGAGGTACAATAAAA

Base Numbers: 84,425,966 - 84,425,980
Number Repeats: 5
Nearest Gene: RAB6

ACCCATTCCCTCCATCCCACCACTAGCaaatt**aataataataataataat**ttaaataattaaataaaataaaa
agacaaatcct

Base Numbers: 84,445,679 - 84,445,696
Number Repeats: 6
Nearest Gene: RAB6

tttGTTTAaataataataataataataataataataataataaatgaaataaaaataatagtgtatatctttttt
aac

Base Numbers: 84,620,594 - 84,620,629
Number Repeats: 12
Nearest Gene: RAB6

gggcaacaaaagcgaaaactccatctcaaaaaTAATAATAATAATTTTACCATACTGCATACTTTAC
AGTCATATAAAGAACTATAATCCTCTACCCTTAA

Base Numbers: 85,032,972 - 85,032,986
Number Repeats: 5
Nearest Gene: Between RAB6 & THRSP

tgggcaacagagagagaccccatttctataaaataataATAATAATAATCCCTAAACTAGGATTTAGAC
AGCTTCTGAGTTGG

Base Numbers: 85,210,150 - 85,210,167
Number Repeats: 6
Nearest Gene: Between RAB6 & THRSP

tgggcaagagcaagactccgtctcaaaataataataataataataataGCCTTATAATTATGTGTTTAA
AAACACAGAGTGAAAGAAAATTCCAAAAGGCATTCCCTATCTCTAAAATAAACCAAA

Base Numbers: 85,291,221 - 85,291,244
Number Repeats: 8
Nearest Gene: Between RAB6 & THRSP

ggcaacagagcgagactccatctcaaaaaataataataataataataataataataataataataataataataa
ACAAATAAATACATAAGTAAGGCTGCAGGCTCAATAAACTGAGAA

Base Numbers: 85,364,856 - 85,364,903
Number Repeats: 16
Nearest Gene: Between RAB6 & THRSP
Repeat (ATT)

taataataataataataataataataTATTATTATTATTATTAcctatcatctcggcattctgggaagctga
gggaagaggatcacctgaggccaagagttcgagaccag

Base Numbers: 72,634,001 - 72,634,018
Number Repeats: 6
Nearest Gene: Between FLJ23392 & FLJ20556

CTCATCTTTCAAGACCGAAACAAAAATCTCAGGTGCTCAGGTGCTCTCTTTGACATCACct
ccagttcatcttctcccattctctctctttctttattattattattattttttttaaattttagagatgggtctcac

Base Numbers: 72,642,459 - 72,642,476
Number Repeats: 6
Nearest Gene: Between FLJ23392 & FLJ20556

GGGACAAACAGAATTCAGAAGGTTG**attattattattattattatt**ttgagatggagtcttgctctgtcac
ccaggctggagtgcaatggcacgatcttg

Base Numbers: 72,769,308 - 72,769,328
Number Repeats: 7
Nearest Gene: HRLP5

tataaactgctagttgccatgtattcctttt**attattattattattattatt**gttattatttcaatagcttttggggaacag
gtggtgtttggttacatagataagtcctgtagtgggtgatttctgaga

Base Numbers: 72,845,628 - 72,845,648
Number Repeats: 7
Nearest Gene: Between HRLP5 & SLC22A8

TACTTAAAG**attattattattattattattatt**atcattattttatttttatGctttaagttttagggtacatgtgca
caatgtgcaggttagttacatatgtatacatgtg

Base Numbers: 72,911,789 - 72,911,812
Number Repeats: 8
Nearest Gene: Between HRLP5 & SLC22A8

acttaaaaaataataa**attaTTATTATTATT**TTTAAaaagaaggtcattatgtaattatacagggatccattcagc
aaggggatataatacttccaaatatatatgca

Base Numbers: 73,194,274 - 73,194,288
Number Repeats: 5
Nearest Gene: Between HRLP5 & SLC22A8

TACTTAAAG**attattattattattattattatt**atcattattttatttttatGctttaagttttaggg

Base Numbers: 73,303,451 - 73,303,474
Number Repeats: 8
Nearest Gene: Between SLC22A6 & HREV107

accacacccagctCAttttt**attattattattatt**ttgagacgaagtttcactcttatcccccaggctggagtgcaatggt
gcgatactggctcactgcaacctctgcctcc

Base Numbers: 73,476,694 - 73,476,708
Number Repeats: 5
Nearest Gene: Between SLC22A6 & HREV107

ttttctttt**attattattattattatt**ttttaaattttagagatggggtctcacatgtttcccaggctggccttgaactcttg
gctcaag

Base Numbers: 73,558,031 - 73,558,048
Number Repeats: 6
Nearest Gene: HREV107

tatatataac**aTTATTATTAttatt**tgttttctcacttttttctttttgaagcagagtccccctttgccgcccagg
ctggagtgcagtggtgggatcttggctcactgcaacctccacct

Base Numbers: 73,668,825 - 73,668,839
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

tttctaaataaacattttatatataaca**TTATTATTATTAtt**tgtttttctcacttt

Base Numbers: 73,668,825 - 73668,839
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

attattattattgtt**attattattattatt**ttttagaggcaggtcttgcttctttgcccaggctggagtgcagtgtggcaat catagctc

Base Numbers: 73,694,328 - 73,694,342
Number Repeats: 5
Nearest Gene: Between HREV107 & MRPL48

TGCTCAAAAAGTAGCAGCTGta**attattattattatt**aataCTGCTACTATTCCAGAACTGCCTTG TATGTC

Base Numbers: 74,018,141 - 74,018,155
Number Repeats: 5
Nearest Gene: Between MRPL48 & LOC51287

GATTCCAACTGGCAattttttgttttttatttttacttattt**attattattattattatt**tttattttgagacagggtct cactctgtcactggggctggagggcagtagcgtgatctcagctcactacag

Base Numbers: 74,593,503 - 74,593,520
Number Repeats: 6
Nearest Gene: KIAA0769

AGCatttaa**attattattattatt**atctttagagacagggtcttgctctgttgcccaggct

Base Numbers: 75,140,272 - 75,140,286
Number Repeats: 5
Nearest Gene: PME-1

TCACTTAAGCaaataaaaat**attattattattatt**ataGCCTACTAGG

Base Numbers: 75,172,445 - 75,172,459
Number Repeats: 5
Nearest Gene: Between PME-1 & FLJ10661

CAtattattgtt**attattattattattatt**gagacagagtttcactcttgtcacccaggctgg

Base Numbers: 75,515,915 - 75,515,932
Number Repeats: 6
Nearest Gene: Between FLJ10661 & FLJ11099

TTTTAattattatc**attattattattattatt**tgagacggagtctcgctctgtcacccaggctggagtgcagtggcgcg
atctcagc

Base Numbers: 75,751,538 - 75,751,555
Number Repeats: 6
Nearest Gene: Between FOLR3 & FOLR2

tatctggtggaagaaatttgttt**attattattattattattattattattattatt**ttgagatagagtctcactctg

Base Numbers: 75,755,654 - 75,755,689; 76,376,183 - 76,376,218
Number Repeats: 12
Nearest Gene: Between FOLR3 & FOLR2

ccactacacctagctaatttttgtt**attattattattattattatt**tgggtagagactaggtctcgctatgatgcccaag
ctggtcgaactcctgggctcaagcagtcctcctgcctcagcctcctaaagtgctgggatta

Base Numbers: 76,090,890 - 76,090,910
Number Repeats: 7
Nearest Gene: Between SKD3 & PDE2A

GTTATGACACTATt**attattattattattatt**aCTTTTTTTTTTT

Base Numbers: 76,104,755 - 76,104,772
Number Repeats: 6
Nearest Gene: Between SKD3 & PDE2A

GCTTCATTTTT**ATTATTAttattatt**atactttaagttctggggcacatgtgcagatcgtgcaggtttattacat
aggtatacatgtgccat

Base Numbers: 76,553,399 - 76,553,413; 76,697,516 - 76,697,530
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

CATTTTCAGGTCA**tattattattattattatt**aATTTTCTTATATTCTTTCTGG

Base Numbers: 76,591,861 - 76,591,878; 76,735,978 - 76,735,995
Number Repeats: 6

Nearest Gene: Between FOLR1 & PRCP

ttgcaaagttttcagggctttt**attattattattattattattattattattatt**aaataggctttctatgacatctccc
atctgtaatccttttagcacaatccc

Base Numbers: 76,761,862 - 76,761,897
Number Repeats: 12
Nearest Gene: Between FOLR1 & PRCP

ttacagatgaggaaactgagCCtt**attattattattatt**aatattattattaGCTTGCAAATAAAAATACTGT
GACTTGAATCCATGTCAGCTTGATAACAAATCTGAATTAA

Base Numbers: 76,766,838 - 76,766,852
Number Repeats: 5
Nearest Gene: Between FOLR1 & PRCP

TGTGTGTTTt**tattattattattattatt**ataagttctgggttacatgtgcagaacgtgcaggtttgttacacaggta
tacacttgcca

Base Numbers: 76,953,222 - 76,953,239
Number Repeats: 6
Nearest Gene: Between FOLR1 & PRCP

ttttaa**ttattattattattattattattattattattatt**ttgagacagagtttcgctcttgcccaggctggagtgc
aatggcgtgc

Base Numbers: 77,326,612 - 77,326,650
Number Repeats: 13
Nearest Gene: Between FOLR1 & PRCP

tttgcttaaatattttt**attattattattattattatt**ttgagacagggtctcgctctgttgcccaggctggagtgcagtgg
tgtc

Base Numbers: 77,757,298 - 77,757,318
Number Repeats: 7
Nearest Gene: Between PRCP & RAB30

TGGTCTGattatcattattattattgtt**attattattattatt**tgagctggagtctcgctctgtcgcccaggctggagg
acattggcctaatcttggctcactgccaccttcacctcctgggttcaagcaattcttgtgcttcagcctcc

Base Numbers: 77,851,794 - 77,851,808
Number Repeats: 5
Nearest Gene: Between RAB30 & PCF11

TAATTGt**attattattattattattattatt**tttgagaccggatctaactgctggagtgcagtggcacaatcatagc

Base Numbers: 78,862,218 - 78,862,241
Number Repeats: 8
Nearest Gene: DLG2

GAATTTATttattttatcattttt**attattattattatt**tgagaccaagtctctgtcaaccaggctgcagtgtagggg
tgtgatctcggctcactgcaacctcaaccttctgggttcaagtgattctcatgcatcagcctccggag

Base Numbers: 79,060,238 - 79,060,252
Number Repeats: 5
Nearest Gene: DLG2

TGATGTTAGTAA**attattattattattattattatt**actattttgagacaagaccttgttctgtcacccaggatgga
gtgcagtgt

Base Numbers: 79,262,184 - 79,262,207
Number Repeats: 8

Nearest Gene: DLG2

TAAGAGTTCACt**tattattattattattattattatt**atGTTAATCAGAAATAATTAAAAACAG
TAAATTAGAA

Base Numbers: 79,331,148 - 79,331,174
Number Repeats: 9
Nearest Gene: DLG2

ccgggcggcggagcttgcagtgagctgagatcgcgccactgcactccagcctgggcaacaaagcgaggttccgtctcaaa
aTAATAATAattcttattaataatatattattatatattaataattatatttattaatattttattattatatttattaa
tatattat**attattattattatt**gttat*GGAAGCATGTAACTATCATGGCCATTGAAGGTTTATTTT*
ATTAGGTC

Base Numbers: 79,689,217 - 79,689,231
Number Repeats: 5
Nearest Gene: DLG2

cccagctt**attattattattatt**tgttgtagagacagggtctcactatgttgcccacactggttttgaacttctgTTCA
AAGAACTTTTAAGAG

Base Numbers: 79,822,310 - 79,822,324
Number Repeats: 5
Nearest Gene: Between DLG2 & HT007

GTTTTGtactattatttttt**attattattattattatt**atactttaagttttagggtacatgtgcacaatgtgctggttag
ttacatatgtatacatgtgacat

Base Numbers: 80,229,444 - 80,229,461
Number Repeats: 6
Nearest Gene: Between DLG2 & HT007

agctcttatgacaacttagtgaggagaagga**attattattattatt**tcaattttacaaagaaaattgagtcattgagaaat
cgaatggattgctccaagttgcatcagcagttattggttaagacagCTTGTGTAGATGTCAT

Base Numbers: 80,230,060 - 80,230,074
Number Repeats: 5
Nearest Gene: Between DLG2 & HT007

TCCATTTAGTTCACATATTATATGGCTTC**CATTATTATTATTAt**tttttttttt

Base Numbers: 80,380,193 - 80,380,207
Number Repeats: 5
Nearest Gene: Between DLG2 & HT007

taataataataata**atTATTATTATTATT**TTTGCCCTTAGACTGTATGGGGGCTCTTCCAGGGC
AGTAA

Base Numbers: 80,504,397 - 80,504,411
Number Repeats: 5

Nearest Gene: Between DKFZp586C1924 & ZF

gggtgtgggggcaggcgcctataatcccagttactcgggaggctgaggcaggacaatcgcttgaacccaggaggcagagg
ttgcagtcagccgagaacatgcaactgcactccagcctgggcgacagagcgagattccatctcaa**aacaacaacaacaa**
caacaacaacaaaaac*CCA*gatccacctgccacggcctcccagagtgctgggattacaggtgtgagccaccatgcctgg
c*ACAT*

Base Numbers: 80,693,536 - 80,693,559
Number Repeats: 8
Nearest Gene: Between SYTL2 & CLTH

ATTTAtatc**attattattattattattatt**tgagatggagtcttgctctgtcacccaggctggagtacagtggcatgat
ctcagctcac

Base Numbers: 80,629,112 - 80,629,132
Number Repeats: 7
Nearest Gene: SYTL2

taggtactactatctttttttaca**ATTATTATTATTATT**GATTAATCATCATAATAAATACTTTTC
ATTTCTACCACAAATGACTATTTCAAAATCAT

Base Numbers: 81,444,010 - 81,444,024
Number Repeats: 5
Nearest Gene: CAPN5

ccaaactagtt**attattattattattattattattattatt**tctggtagagacagggtctcac*A*AGAATCATTT
CTttacaaat

Base Numbers: 81,519,331 - 81,519,363
Number Repeats: 11
Nearest Gene: Between CAPN5 & PHCA

aaatttattatg**attattattattatt**tgagacagagtcttgctctgtcgccaggctggagtgcagtggcgtgatcttggc
ttactgcgatctccgcc

Base Numbers: 81,569,104 - 81,569,118
Number Repeats: 5
Nearest Gene: PHCA
*AATTACAG*ttattattattttt**attattattattatt**ttgagatggagtctcactctgtctccaggctggagtaca

Base Numbers: 82,017,809 - 82,017,823
Number Repeats: 5
Nearest Gene: Between GL002 & PRKRIR

*ACTCCCCCA*TTAAAACA**Attattattattatt**ttatttttttatagagatggggtcttgctaccttgaccagtctg
gtcttgaacccccctggcctcaggagagcttc

Base Numbers: 82,985,014 - 82,985,028
Number Repeats: 5
Nearest Gene: SERPINH2

AGCAAAATGGGA**ttattattattattatt**ttgagacggagtctcgctcgttgcccagg

Base Numbers: 83,212,431 - 83,212,448
Number Repeats: 6
Nearest Gene: Between RPS3 & ARRB1

acccttggctaagtttctttttt**attattattattatt**ttttggagagatggggcctcactgtgttgcccaggctggtctc
aac

Base Numbers: 83,318,873 - 83,318,887
Number Repeats: 5
Nearest Gene: Between ARRB1 & SLC21A9

GGAAATTGCAATTTTT**CATTATTATTATTATTATT**Agagacagggtctcacttatcacacaggcttga
gtacaggggtgtgattatagctcatt

Base Numbers: 83,401,056 - 83,401,073
Number Repeats: 6
Nearest Gene: SLC21A9

ttgaggatttctgcatcaatattcatcagatatattggcctgt**attattattattatt**tttatttatttattttattgattttt

Base Numbers: 83,629,311 - 83,629,325
Number Repeats: 5
Nearest Gene: Between KIAA0102 & FLJ22596

GGTCAAAGGTtt**attattattattattattattattattatt**TTAAATTTTATTTCTTTTT

Base Numbers: 84,290,357 - 84,290,386
Number Repeats: 10
Nearest Gene: KIAA0337

ttcaaatttctctacatcctcaacaaccttgtttt**attattattattattattatt**atagccatcctat

Base Numbers: 84,321,716 - 84,321,736
Number Repeats: 7
Nearest Gene: Between FLJ14993 & PHRETI

CACCTCTTAAAattataatttt**attattattattattatt**TTGgaggcagagtctcactctgtcaccaggctggagt
gcagtggcacaatctcggctcactgcaacctctgcc

Base Numbers: 84,409,318 - 84,409,335
Number Repeats: 6
Nearest Gene: Between FLJ14993 & PHRETI

gttttctccttaagtat**attattattattattattattatt**actattattattatgactcgtggatttaaacatgttttata
gtaattactgattttcaagtttcctacagtttgaccagtag

Base Numbers: 84,601,425 - 84,601,448
Number Repeats: 8

Nearest Gene: RAB6

tgcctgactaattttt**attattattattatt**tttagatataggatctcactatgttgcccaggctagtcttgaactcctggc
ctcaagccatcctcctgccttgacctcccaaagtt

Base Numbers: 84,711,338 – 84,711,352
Number Repeats: 5
Nearest Gene: Between RAB6 & THRSP

ttcttcttcttcttcttcttcttcttcttt**attattattattattattattatt**agagacagtctcactctgttgcccaggct
ggagtgcaa

Base Numbers: 85,054,928 – 85,054,951
Number Repeats: 8
Nearest Gene: CLNS1A

AATCAGCTGAAACAAACCTCCAAGTT**ATTATTATTATTAtt**tttttttt

Base Numbers: 85,305,641 – 85,305,655
Number Repeats: 5
Nearest Gene: PTD015

caggctaatttttatttt**attattattattattatt**tgagacaagtcttgctctgtcgcccaggctggagtgcagtagtgc
aatctcagctc

Base Numbers: 85,460,008 – 85,460,025
Number Repeats: 6
Nearest Gene: Between MGC2376 & PRO2849

Repeat (ACC)

CAATTAGAATTCATGATACCAGTTTCc**caccaccaccaccaccaccc**GTCGCCCTGCCATTGTTAG
CAAAACCATCTCTTGAGTGGAG

Base Numbers: 73,959,445 – 73,959,462
Number Repeats: 6
Nearest Gene: Between HREV107 & MRPL48

aacaacctaataaatccatctctggcaaagtcac**accaccaccaccaccaccaccc**acaaactctctcagcctagaccact
gagatact

Base Numbers: 79,562,544 – 79,562,564
Number Repeats: 7
Nearest Gene: DLG2

cattgtcaacaccaccatcactacctttaccatcaccactatcatcactacccacaacactgtcatcaccaccaacaccat
cactacctccaccatc**caccaccaccaccaccc**atcactaacatcatc*CCTACCTCCAATATTACCATAGTCAT
CA*

Base Numbers: 81,336,257 – 81,336,271

Number Repeats: 5
Nearest Gene: Between PAK1 & MYO7A

cacacagccgtgcgccctcacatacaccac**accaccaccaccaccaccac**acacacacacacacacac

Base Numbers: 81,336,257 - 81,336,274
Number Repeats: 6
Nearest Gene: Between PAK1 & MYO7A

atagcaaaaactcaatacaaattgaccattTact**accaccaccaccaccaccaccaccaccaccac**actactacTATTT
ATGAAG

Base Numbers: 81,683,381 - 81,683,410
Number Repeats: 10
Nearest Gene: PHCA

AGCACAGAGCCTGGCA**CACCACCACCACCACC**ACAAGGTGCCCGGCGAAAACCAGCCCCTTC
TTCCTGAATGTGTCAGGCTTCTCCCCTCGAGGGTCGTAACTC

Base Numbers: 83,102,606 - 83,102,620
Number Repeats: 5
Nearest Gene: Between SERPINH2 & PP1665

AACTTCAAAGACCACAGTGTTGGGAAGGCAA**CACCACCACCACCACC**AACTGGATTCTTG
ATTCAGTTTCCCACTGGAAAGGCAGATCCAAAAAGAGGTGAGAG

Base Numbers: 83,146,897 - 83,146,911
Number Repeats: 5
Nearest Gene: Between PP1665 & RPS3

CGTGcgcacacacgcacaactgcacaaagacgtgcacactcctccatccacaccaatacactccctaatagggcccgca
cacagccgtgcgccctcacatacaccac**accaccaccaccaccaccac**acacacacacacacacactctcacacggactcg
cacgc

Base Numbers: 83,177,504 - 83,177,521
Number Repeats: 6
Nearest Gene: Between RPS3 & ARRB1

Tcaccactaccactgtcactgccatt**accaccaccaccaccaccac**gttactgccacctctaccaccaggaccacccctcac
acttgcagg

Base Numbers: 83,839,223 - 83,839,240
Number Repeats: 6
Nearest Gene: Between KIAA0102 & FLJ22596

Repeat (GGT)

GCGGGTGATCTCCCCCATCCCGCAGCGTGG**ggtggtggtggtggtggt**ggAAGCAGCACTAACTGG
GACCCAGGCGTCCCGAGTGCAGGCCCTGCTCTCTCAATGGAAAGAGCAGCTGTCATCTACC
A

Base Numbers: 83,214,448 – 83,214,465
Number Repeats: 6
Nearest Gene: Between RPS3 & ARRB1

TTTCCAGTGGGAAACTGAATCAAGAATCCAGTT**GGTGGTGGTGGTGGT**GTTGCCTTCCCA
ACACTGTGGTCTTTGAAGTTGAGTTGTCTAAACTCTCCAACCT

Base Numbers: 84,842,518 – 84,842,532
Number Repeats: 5
Nearest Gene: Between RAB6 & THRSP

Repeat (ACG)

Repeat (CGT)

Repeat (ACT)

Repeat (AGT)

ATAAATAAATATAAGGAAGATAGg**tagtagtagtagtag**agaggtagtcttagcagcagcagcagCAA
GAGTAAAAACATTGGTGTAATTTTTTCAACTCTACATGATG

Base Numbers: 79,086,747 – 79,086,761
Number Repeats: 5
Nearest Gene: DLG2

As derived from GenBank, BLAST by this author (2002).

Repeats in bold

Coding sequence in upper-case

Non-coding, intronic sequences in lower case

APPENDIX 11

NCBI Map Viewer

NCBI Map Viewer Chr11: 64,533,373-75,971,382

***Homo sapiens* Map View** **BLAST The Human Genome**

Build 34 Version 1

Chromosome: 1 2 3 4 5 6 7 8 9 10 [**11**] 12 13 14 15 16 17 18 19 20 21 22 X Y

Master Map: Genes On Sequence Maps & Options

Total Genes On Chromosome: **1501**

Region Displayed: **0-134M bp** Download/View Sequence/Evidence

Genes Labeled: **19** Total Genes in Region: **1501**

Symbol	O	Cyto	Description
KCNQ1	↓	11p15.5	potassium voltage-gated channel, KQT-like subfamily, member 1
CMT4B2	↑	11p15.3	Charcot-Marie-Tooth neuropathy 4B2 (autosomal recessive, with myelin outfolding)
SOX6	↑	11p15.3	SRY (sex determining region Y)-box 6
NAV2	↓	11p15.1	neuron navigator 2
NELL1	↓	11p15.2-p15.1	NEL-like 1 (chicken)
C11orf25	↓	11p14.3	chromosome 11 open reading frame 25
LOC375992	↑	11p12	similar to Trophinin
UVRAG	↓	**11q13.5**	**UV radiation resistance associated gene**
DLG2	↑	11q21	discs, large homolog 2, chapsyn-110 (Drosophila)
GRM5	↑	11q14.2	glutamate receptor, metabotropic 5
MAML2	↑	11q21	mastermind-like 2 (Drosophila)
CNTN5	↓	11q21-q22.2	contactin 5
GRIA4	↓	11q22	glutamate receptor, ionotrophic, AMPA 4
GUCY1A2	↑	11q21-q22	guanylate cyclase 1, soluble, alpha 2
KIRF4	↑	11q23.2	immunoglobulin superfamily, member 4
DSCAML1	↑	11q23	Down syndrome cell adhesion molecule like 1
NEPH2	↑	11q24	NEPH2
HNT	↓	11q25	neurotrimin
OPCML	↑	11q25	opioid binding protein/cell adhesion molecule-like

NCBI's Map Viewer program integrates physical and genetic map information for specific sequences, proteins and genes. This figure demonstrates the positions of *UVRAG, WNT11* and *CLNS1A*, three of the five prime candidate genes, on chromosome 11, at the locus displaying linkage with HFM in the West Australian family.

APPENDIX 12

Bioinformatics derived from location of Markers

D11S1881-D11S911

APPENDIX 12

Bioinformatics derived from location of Markers D11S1881-D11S911, April 2002 Freeze (64,533,373-75,971,382)

HRASLS3 (64,501,887-64,514,534)

HRAS-like suppressor 3, HREV107, HREV107-3, H-REC107-1

Similar to rat HREV107

Homologous to mouse chromosome 19

ESRRA (64,620,871-64,628,909)

Estrogen-related receptor alpha, ERR1, ERRa, ESRL1, NR3B1, ERRALPHA, ERRalpha, Estrogen receptor-like 1

Homologous to mouse chromosome 19

Gene ontology

Nucleus

DNA binding

Ligand-dependent nuclear receptor

Pol II transcription

HSPC152 (64,628,862-64,629,732)

Gene ontology

Unknown

PRDX5 (64,630,351-64,976,461)

Peroxiredoxin 5, ACR1, B166, PRXV, PMP20, AEOB166, Antioxidant enzyme B166

Homologous to mouse chromosome 18

Gene references into function

Crystal structure at 1.5 angstrom resolution

Glutaredoxin-dependent peroxiredocin from poplar: protein-protein interaction and catalytic mechanism

Gene ontology

Peroxisome

Respiration

Mitochondrion

Electron transporter

Inflammatory response

Oxidative stress response

Cell stress

Oxidoreductase

Anti-pathogen response

PLCB3 (64,631,222-64,637,236)

Homologous to mouse chromosome 19

Gene ontology

Phospholipase C

Tumor suppressor

BAD (64,639,514-64,665,041)

BCL2-antagonist of cell death, BCL2L8, BCL20-binding protein, BCL2-binding component 6, BCL-X/BCL-2 binding protein

Gene references into function

Expression in normal, hyperplastic and carcinomatous human prostate

Direct interaction with Bad with pro-survival members of the Bcl-2 family

Contributes to the progress of Sindbis virus-induced apoptosis

Protein kinase A Rialpha antisense inhibition of PC3M prostate cancer cell growth

Bcl-2 hyperphosphorylation

Bax up-regulation

Bad-hypophosphorylation

Gene ontology

Apoptotic program

LOC 56834 (64,665,908-64,667,812)

Homologous to mouse chromosome 19

Gene ontology

Unknown

KCNK4 (64,671,207-64,679,037)

TRAAK, DKFZP566E164, Two pore K+ channel, 2P domain potassium channel, tandem pore domain potassium channel TRAAK, potassium inwardly-rectifying channel subfamily K member 4, TWIK-related arachidonic acid-stimulated potassium channel protein

Homologous to mouse chromosome 19

Gene ontology

Potassium channel

Potassium transport

Channel (passive transporter)

LRP16 (64,695,195-64,865,774)

Homologous to mouse chromosome 19

Gene ontology

Unknown

FLRT1 (64,742,227-64,744,902)

Expressed in kidney and brain

STIP1 (64,757,292-64,769,066)

HOP, P60, STI1L, IEF-SSP-3521

Transformation-sensitive, similar to *Saccharomyces cerevisiae* STI1

Homologous to mouse chromosome 19

Gene ontology

Nucleus

Stress response

Golgi apparatus

HSPF2 (64,770,494-64,771,922)

MCG18, DANJC4, DNAJC4-PENDING, Heat shock 40kD protein 2

Homologous to mouse chromosome 19

Gene ontology

Membrane fraction

Heat shock protein

Unspecified membrane

MGC13045 (64,773,317-64,777,043)

Homologous to mouse chromosome 19

Gene ontology

Unknown

MGC11134 (64,777,182-64,779,533)

Gene ontology

Unknown

COX8 (64,790,817-64,792,719)

COX VIII

Homologous to mouse chromosome 19

Gene ontology

Mitochondrion

Energy pathways

Cytochrome-c oxidase

General cellular role

FLJ13848 (64,810,383-64,849,174)

Gene ontology

Unknown

FLJ20113 (64,865,917-64,875,998)

Homologous to mouse chromosome 19

Gene ontology

Unknown

HRLP5 (63,675,029-63,702,058)

Gene ontology

Unknown

LAGALS12 (64,906,969-64,917,653)

GALECTIN-12, Galactin-12

Homologous to mouse chromosome 19

RARRES3 (64,937,694-64,947,344)

RIG1, TIG3

Gene ontology

Tumor suppressor

Negative control of cell proliferation

HRASLS2 (64,953,656-64,964,268)

FLJ20556

Gene ontology

Unknown

DKFZp434G0920 (65,008369-65,012,188)

Gene ontology

Unknown

RPS6KA4 (65,013,825-65,026,869)

MSK2, RSKB, RSK-B

Ribosomal protein kinase B; mitogen- & stress-activated protein kinase 2

Homologous to mouse chromosome 19

Gene ontology

Nucleus

Protein kinase

Protein kinase cascade

Transferase

SLC22A11 (65,210,571-65,226,150)

HOAT4

Gene ontology

Small molecule transport

Integral plasma membrane protein

Sodium-independent organic anion transporter

Active transporter, secondary

Major facilitator superfamily

NRXN2 (65,260,795-65,368,621)

Neurexin 2, KIAA0921, Neurexin II

Homologous to mouse chromosome 19

RASGRP2 (65,381,710-65,400,252)

CDC25L, CALDAG-GEFI, Calcium and diacylglycerol-regulated guanine,

Nucleotide exchange factor I

Homologous to mouse chromosome 19

Gene ontology

Lipid binding

Calcium binding

Signal transduction

Guanyl-nucleotide exchange factor

Activator

Inhibitor/repressor

PYGM (65,401,354-65,415,587)

Phosphorylase, glycogen, muscle

Homologous to mouse chromosome 19

Phenotype

McArdle disease

Gene references into function

McArdle disease may be caused by R269X nonsense mutation in this gene

Gene ontology

Glycogen metabolism

Glycogen phosphoylase

Muscle action

Energy storage

Transferase

SF1 (65,419,478-65,433,636)

Splicing factor 1, ZFM1, ZNF162, D11S636, Zinc finger protein 162

Homologous to mouse chromosome 19

Gene references into function

The KH-QUA2 region of SF1 defines an enlarged KH (hn RNP K) fold which is necessary and sufficient for intron branched point sequence (BPS) binding

Gene ontology

Nucleus

Transcription co-repressor

RNA polymerase II transcription factor

RNA splicing

Inhibitor/repressor

MAP4K2 (65,444,013-65,458,059)

GCK, BL44, RAB8IP, Rab8 interacting protein (GC kinase)

Homologous to mouse chromosome 19

Gene ontology

JNK cascade

Golgi membrane

Immune response

Stress response

Soluble fraction

Hemocyte development

Protein serine/threonine kinase

Peripheral plasma membrane protein

Differentiation

Transferase

MEN1 (65,458,396-65,465,594)

Multiple endocrine neoplasia I, MEAL, SCG2, Menin, Wermer syndrome

Endocrine adenomatosis, multiple

Zollinger-Ellison syndrome, included

Homologous to mouse chromosome 19

Phenotype

Adrenal adenoma, sporadic

Angiofibroma, sporadic

Carcinoid tumor of lung

Hyperparathyroidism, AD

Lipoma, sporadic

Multiple endocrine neoplasia I

Parathyroid adenoma, sporadic

Prolactinoma, hyperparathyroidism, carcinoid syndrome

Gene references into function

Multiple endocrine neoplasia type 1Burin from Mauritius: a novel MEN1 mutation

Loss of heterozygosity of the MEN1 gene in a large series of TSH-secreting pituitary adenomas

Gene ontology

Nucleus

Protein binding

Tumor suppressor

Transcription regulation

Repression of transcription from Pol II promoter

Inhibitor/repressor

EHD1 (65,507,603-65,533,576)

PAST, HPAST, H-PAST

Testilin

EH domain containing 1

Homolog of drosophila past

Homologous to mouse chromosome 19

PPP2R5B(65,579,835-65,589,340)

Protein phosphatase 2, regulatory subunit B

Gene Ontology

Cytoplasm

Protein phosphatase type 2A regulator

CNS- specific functions

Regulatory subunit

GPHA2 (65,589,640-65,590,327)

Glycoprotein alpha 2

Cysteine knot protein

Gene References into Function

Capable of mediating the actin of relaxin through an adenosine 3',5'-monophosphate (cAMP)-dependant pathway

NAALADASEL (65,642,795-65,955,527)

I100

CAPN1(65,668,883-65,699,003)

CANP, muCL, CANPL1, muCANP, Calpain, large polypeptide L1, Calcium-activated neutral proteinase

Homologous to mouse chromosome 19

Gene References into Function

Calpain (mu-calpain) is a signal transducer and activator of transcription (STAT) 3 and STAT5 protease

Gene Ontology

Positive control of cell proliferation

MRPL49 (65,761,870-65,766,888)

Mitochondrial ribosomal protein L49

Gene Ontology

Unknown

FAU(65,767,036-65,768,592)

FAU1, RPS30, 40S ribosomal protein S30, ubiquitin-like protein fubi, ubiquitin-like-S30 fusion protein, FAU- encoded ubiquitin-like protein, FBR-MuSV-associated ubiquitously expressed gene

Homologous to mouse chromosome 19

Gene Ontology

Ribosome

Cytoplasm

Ubiquitin

RNA binding

Protein biosynthesis

Structural protein of ribosome

Protein conjugation factor

C11orf5 (65,711,524-65,772,819)

FON, Chromosome 11 open reading frame 5

Homologous to mouse chromosome 19

Gene Ontology

Unknown

TM7SF2 (65,772,992-65,777,320)

ANG1

Homologous to mouse chromosome 19

Gene Ontology

Endoplasmic reticulum

Integral plasma membrane protein

C11orf2(65,777,514-65,793,011)

ANG2, Chromosome 11 open reading frame2

Homologous to mouse chromosome 19

Gene Ontology

Unknown

ZFPL1 (95,800,822-65,804,985)

MCG4, D11S750, Zinc finger protein-like 1, Zinc-finger protein in MEN1 region

Homologous to mouse chromosome 19

MGC16386 (65,805,170-65,811,765)

Homologous to mouse chromosome 2

Gene Ontology

Unknown

MGC10966 (64,779,446-64,784,078)

Gene Ontology

Unknown

HSU79266 (65,955,935-65,959,794)

Homologous to mouse chromosome 19

Gene Ontology

Unknown

SNX15 (65,960,192-65,973,274)

Sorting nexin 15, HSAFOO1435, Clone iota unknown protein

ARL2 (65,978,586-65,986,574)

ARLF2, ADP ribosylation factor-like 2

Homologous to mouse chromosome 19

Gene References into Function

C.elegans evl-20 gene encodes a functional homolog of human ARL2. Elimination of evl-20 function results in abnormal vulval, gonad, and male tail development and disrupts embryonic proliferation, hypodermal enclosure and elongation

Gene Ontology

GTP binding

Tubulin folding

GTPase inhibitor

Hydrolase

Inhibitor/Repressor

GTP-binding protein/GTPase

LOC116071 (66,010,517-66,012,826)

Gene Ontology

Unknown

EMK1 (66,311,572-66,332,557)

MARK2 ELKL motif kinase 1, ELK motif kinase 1

Homologous to mouse chromosome 19

Gene Ontology

Protein phosphorylation

Microtubule cytoskeleton

Protein serine/threonine kinase

Cell polarity

Transferase

Cytosketal

Tubulin-cytoskeleton associated

POLA2 (66,448,552-66,484,208)

Homologous to mouse chromosome 19

CEP2 (66,501,451-66,509,016)

BORG1

REQ (66,520,465-66,539,575)

UBID4, MGC10180, Ubi-d4

Apoptosis response zinc finger protein

Homologous to mouse chromosome 19

Gene Ontology

Induction of apoptosis by extracellular signals

FKSG44 (66,573,293-66,598,258)

Gene Ontology

Unknown

SPRY4 (66,627,002-66,629,637)

Gene Ontology

Known/Inferred

NTKL (66,711,704-66,725,296)

GKLP, NKTL, P105, TAPK, TRAP, HT019, N-terminal kinase-like protein, Teratoma-associated tyrosine kinase, Telomerase regulation-associated protein, HT019 protein; telomerase regulation-association protein

Likely ortholog of mouse N-terminal kinase-like protein

Hypothetical gene supported by AB047077; AB051427; AB051428; AF297709; BC009967; NM_020680

Homologous to mouse chromosome 19

LTBP3 (66,725,395-66,734,641)

LTBP2, DKFZP586M2123

Homologous to mouse chromosome 19

SF3B2 (66,793,585-66,810,082)

SF3B, SAP145, SF3B145

Spliceosome associated protein 145, SF3b subunit

Homologous to mouse chromosome 3

Gene ontology

Spliceosome

mRNA splicing

mRNA processing

pre-mRNA splicing factor

nuclear

RNA-binding protein

GSTP1 (67,064,303-67,067,136)

PO, DFN7, GST3, FAEES3

Deafness, X-linked

Fatty acid ethyl ester synthase III

Homologous to mouse chromosome 19

Gene references into function

Expression in nasopharyngeal carcinoma

GSTP1 gene encodes the pi class glutathione S-transferase

Genetic determinants of lung cancer short-term survival

Gene expression level of beta-TUB, Bcl-XL and GSTpi was closely related with the IC50 for docetaxel

Data suggest that GST-pi expression in tumor cells are related to drug resistance in epithelial ovarian cancer

Expression in malignant tissue and plasma levels in the human colorectal and gastric tumors increased depending on tumor stage

Glutathione S-transferase P1 and NADPH quinone exidoreductase polymorphisms are associated with aberrant promoter methylation of P16 (INK4a) and O(6)-methylguanin-DNA methytransferase in sputum

The GSTP1 gene encodes for an enzyme, glutathione S-transferase pi (GSTpi), involved in detocification of carcinogens. An amino-acid substitution (I105)V in GSTP 1 produces a variant enzyme with lower activity and less capability of effective detoxification

Gene ontology

CNS development

Transferase

CNS-specific functions

NDUFV1 (67,087,479-67,092,932)

UQOR1

Homologous to mouse chromosomes 16 and 14

Phenotype

Alexander disease

Leigh syndrome

Gene ontology

Energy pathways

Membrane fraction

Mitochondrial inner membrane

NADH dehydrogenase (ubiquinone)

General cellular role

MGC9740 (67,109,306-67,117,431)

HCBP1, MGC9740, HCV core-binding protein

Homologous to mouse chromosome 19

CABP2 (67,195,637-67,200,084)

CaBP2, Calcium binding protein

Homologous to mouse chromosome 19

Gene ontology

Calcium binding

Signal transduction

CNS-specific function

Small molecule-binding protein

DOC-1R (67,210,401-67,212,535)

Homologous to mouse chromosome 19

Gene ontology

Tumor suppressor

PITPNM (67,223,447-67,238,839)

NIR2, DRES9

Drosophila retinal degenerationB

PYK2 N-terminal domain-interacting receptor 2

Homologous to mouse chromosome 19

Gene ontology

Brain development

Lipid metabolism

Phototransduction

Membrane fraction

Phosphatidylinositol transporter

AIP (67,239,504-67,243,610)

ARA9, XAP2, FKBP37

HBV-X associated protein

Aryl hydrocarbon receptor interacting protein

Homologous to mouse chromosome 19

Gene references into function

Two distince regions of the immunophilin-like protein XAP2 regulated dioxin receptor function and interaction with hsp90

Gene ontology

Cytoplasm

Signal transduction

Transcription co-activator

Transcription factor binding

Activator

Pol II transcription

Inhibitor/repressor

ALDH3B2 (67,256,288-67,267,245)

ALDH8, ALDH3B2-PENDING, Aldehyde dehydrogenase 8

Gene ontology

Lipid metabolism

Alcohol metabolism

Aldehyde dehydrogenase

UNC93B1 (67,518,981-67,531,960)

Unc93, UNC93B, Unc-93 related protein, Unc93 (*C.elegans*) homolog B

Homologous to mouse chromosome 19

Gene ontology

Exact function not known

ALDH3B1 (67,538,223,-67,557,147)

ADLH4, ALDH7, ALDH3B1-PENDING, Aldehyde dehydrogenase 7

Homologous to mouse chromosome 19

Gene ontology

Lipid metabolism

Alcohol metabolism

Aldehyde dehydrogenase

NDUFS8 (67,558,515-67,564,520)

NADH dehydrogenase (ubiquinon) Fe-S protein 8 (23kD)

NADH-coenzyme Q

Phenotype

Leigh syndrome

Gene ontology

Membrane fraction

Complex I (NADH to ubiquinone)

NADH dehydrogenase (ubiquinone)

Iron-sulfur electron transfer carrier

Energy generation

Oxidoreductase

TCIRG1 (67,566,889-67,578,772)

A3, Stv1, Vph1, ATP6I, OC116, OPTB1, TIRC7, ATP6N1C, ATP6N1D, OC-116kDa, T-cell, immune regulator, V-ATPase 116-kDa isoform a3, ATPase H+ transporting 116kD

Infantile malignant osteopetrosis

T cell immune response cDNA 7 protein

Osteoclastic proton pump 116 kDa subunit

Vacuolar proton translocating ATPase 116kDa

Homologous to mouse chromosome 19

Phenotype

Osteopetrosis

Gene ontology

Transporter

Plasma membrane

Proton transport

Cellular defence response

Integral plasma membrane protein

Positive control of cell proliferation

Anti-pathogen response

CHK (67,580,852-67,609,331)

Choline kinase, CK1

Homologous to mouse chromosome 19

Gene references into function

Activity regulated by Ras proteins through Ral-GDS and PI-3 kinase; has role in malignant transformation

Gene ontology

Choline kinase

Lipid transport

Lipid metabolism

Signal transduction

FLJ20039 (67,685,737-67,687,668)

Hypothetical protein FLJ20039

Gene ontology

Unknown

CGI-85 (67,694,363-67,741,898)

CGI-85 protein

Gene ontology

Unknown

C11orf24 (67,789,962-67,800,576)

DM4E3

Homologous to mouse chromosome 19

Gene ontology

Unknown

LRP5 (67,841,286-67,995,925)

HBM, LR3, LRP7, OPPG, BMND1

Low density lipoprotein receptor-related protein 7

Homologous to mouse chromosome 19

Phenotype

Bone mineral density variability 1

Osteoporosis-pseudoglioma syndrome

Gene references into function

Seven novel sequence variants/polymorphisms

Mutation results in an autosomal dominant high-bone-mass trait

LRP5V171 mutation causes high bone density by impairing the action of a normal antagonist of the *Wnt* pathway and thus increasing *Wnt* signalling

Gene ontology

LDL receptor

Lipid metabolism

Signal transduction

Integral membrane protein

Positive control of cell proliferation

Protein translocation

Intercellular transport

C11orf23 (68,038,431-68,153,845)

SAPL, FLJ11058

Sporulation-induced transcript 3-associated protein

Homologous to mouse chromosome 19

Gene ontology

Unknown

LOC51083 (68,229,061-68,229,782)

Galanin-related peptide

MTL5 (68,246,060-68,289,040)

MTLT, TESMIN, tesmin

Homologous to mouse chromosome 19

Gene ontology

Anti-apoptosis

Spermatogenesis

Heavy metal binding

Metal ion homeostasis

Oxidative stress response

Small molecule-binding protein

CPT1A (68,296,227-68,356,254)

CPT1, CPT1-L, L-CPT1

Homologous to mouse chromosome 19

Phenotype

CPT deficiency, hepatic, type I

Gene references into function

Human CPT1A, CPT1B, CPT2, CROT and CRAT are knownto encode

activate carnitine acyltransferases

Gene ontology

Mitochondrion

Fatty-acid beta-oxidation

Carnitine O-palmityltransferase

Energy generation

Energy storage

IGHMBP2 (68,381,164-68,478,692)

HCSA, CATF1, SMARD1, SMUBP2

Homologous to mouse chromosome 19

Gene ontology

DNA repair

DNA helicase

DNA replication

DNA recombination

Single-strand DNA binding

CCND1 (69,061,641-69,075,010)

BCL1, PRAD1, U21B31, D11S287E

B-cell CLL/lymphoma 1

G1/S-specific cyclin D1

Homologous to mouse chromosome 7

Phenotype

Centrocytic lymphoma

Leukemia/lymphoma, B-cell, 1

Multiple myeloma

Parathyroid adenomatosis 1

Gene references into function

Expression is related to apoptosis in thymus

Cyclin D1 play important roles in oesophageal carcinogenesis

A/G polymorphism of CCND1 was associated with the susceptibility to NPC

Endostatin causes G1 arrest of endothelial cells through inhibition of cyclin D1

TGF-beta and PTHrP control chondrocyte proliferation by activating cyclin D1 expression

Analysis of expression improves differentiation of mantle cells from other lymphoma cells

Cyclopentenone causes cell cycle arrest and represses cyclin D1 promoter activity in MCF-7 breast cancer cells

Over-expression of cyclin D1 is found to be significantly correlated with increased chromosomal instability in patients with breast cancer

Activation of cyclin D1 and D2 promoters by human T-cell leukemia virus type I tax protein is associated with IL-2 independent growth of T-cells

FGF19 (69,118,774-67,124,874)

Gene references into function

FGF19 transgenic mice had a significant and specific reduction in fat mass that resulted from an increase in energy expenditure. FGF19 transgenic mice did not become obese or diabetic on a high fat diet

Gene ontology

Neurogenesis

Neuronal development

Embryonic expression

FGF4 (69,193,565-69,195,939)

HST, KFGF, HST-1, HSTF 1, K-FGF, HBGF-4

Oncogene HST

Kaposi sarcoma oncogene

Transforming protein KS3

Heparin secretory transforming protein 1

Human stomach cancer, transforming factor from FGF-related oncogene

Gene references into function

Activation of human HST-1 gene in transgenic mice induces spermatogenesis and prevents Adriamycin-induced toxicity

Gene ontology

Oncogenesis

Extracellular

Growth factor

Signal transduction

Cell-cell signalling

Positive control of cell proliferation

Embryonic expression

FGF3 (69,230,492-69,231,225)

INT2, HBGF-3

INT-2 proto-oncogene protein precursor

Murine mammary tumor virus integration site 2, mouse, included

V-INT2 murine mammary tumor virus integration site oncogene homolog

Homologous to mouse chromosome 7

Gene ontology

Oncogenesis

Extracellular

Growth factor

Signal transduction

Cell-cell signalling

Histogenesis and organogenesis

Embryogenesis and morphogenesis

Embryonic expression

FLJ10261 (69,420,615-69,477,950)

Gene ontology

Unknown

FADD (69,492,036-69,496,058)

MORT1

Fas-associating protein with death domain

Homologous to mouse chromosome 7

Gene references into function

Apoptosis and NF-kappa B: the FADD connection

Binding of FADD and caspase-8 to *molluscum contagiosum* virus MC159 v-FLIP is not sufficient for its anti-apoptotic function

Significant decrease in the percentage of FADD-immunoreactive dopaminergic (DA) neurons in the substantia nigra pars compacta of patients with Parkinson's disease

Gene ontology

Cytoplasm

Oncogenesis

Antimicrobial humoral response

Death receptor associated factor

Cell surface receptor linked signal transduction

Induction of apoptosis via death domain receptors

Anti-pathogen response

PPFIA1 (69,559,414-69,672,161)

LIP.1

Homologous to mouse chromosome 7

Gene references into function

Physical and functional interactions between tyrosine phosphatase alpha, PI 3-kinase and PKCdelta

Gene ontology

Cytoplasm

Cell adhesion

Focal adhesion

Signal transduction

Cell-cell matrix adhesion

Cell shape and cell size control

EMS1 (69,687,248-69,725,279)

Cttn, oncogene EMS1

Homologous to mouse chromosome 13

Gene references into function

Substrate for caspase cleavage during apoptosis

Primary arrest of circulating platelets of collagen phosphorylation of

Syk, cortactin and focal adhesion kinase

Gene ontology

Cytoskeleton

Soluble fraction

Peripheral plasma membrane protein

Actin-cytoskeleton associated

Cytoplasmic

SHANK2 (69,756,543-67,950,362)

SHANK, CORTBP1, ProSAP1, SPANK-3, KIAA1022

Cortactin binding protein 1

GKAP/SAPAP interacting protein

Cortactin SH3 domain-binding protein

Gene ontology

Expressed in developing and adult brain

FLJ10661 (70,372,551-70,386,250)

Homology to chromosome 8 (2278350-2290754)

Gene ontology

Unknown

FLJ11099 (70,575,611-70,582,609)

Gene ontology

Unknown

IL18BP (70,584,077-70,588,767)

Gene References into Function

Transcriptional activation & release of IL-18 binding protein in response to IFN-gamma

Gene ontology

Immune Response

Soluble fraction

Signal transduction

NUMA1 (70,587,887-70,665,529)

NUMA

Phenotype

Leukaemia, acute promueliocytic, NUMA/RARA type

Gene ontology

Spindle

Nucleus

Oncogenesis

Mitotic anaphase

Structural protein

Spindle microtubule

Nuclear organisation & biogenesis

FLJ20625 (70,682,310-70,688,287)

Gene ontology

Unknown

DKFZP564M082 (70,694,600-70,697,791)

FOLR

Phenotype

Congenital abnormalities

Gene Ontology

Folate binding

Membrane fraction

Receptor Mediated endocytosis

Tumor antigen

Integral plasma membrane protein

Folate transport

Peripheral Membrane

GPI-anchored membrane-bound receptor

FOLR3 (70,720,736-70,724,899)

FR-G, FR-gamma, gamma-HFR

Gene ontology

Folate binding

Folate transport

Membrane binding

GPI-anchored membrane-bound receptor

Small molecule-binding protein

FOLR1 (70,774,567-70,781,306)

FOLR

Gene ontology

Folate binding

Tumor antigen

Folate transport

Membrane fraction

Receptor mediated endocytosis

Integral plasma membrane protein

GPI-anchored membrane-bound receptor

FOLR2 (70,801,808-70,806,954)

FR-P3, FR-beta, FBP/PL-1, beta hFR

Gene ontology

Folate binding

Folate transport

Membrane fraction

Unspecified membrane

Post-translational membrane targeting

Small molecule-binding protein

INNPL1 (70,809,790-70,824,114)

SHIP2

Gene ontology

Phosphate metabolism

SKD3 (70,877,434-71,019,533)

Gene ontology

Known / inferred

SDCCAG28 (71,167,031-71,167,405)

CGI-52, NY-CO-28, Similar to phosphatidtcholine transfer protein 2

Gene ontology

Tumor antigen

KIAA0769 (71,210,905-71,516,193)

Gene ontology

Unknown

P2RY6 (71,613,248-71,615,381)

Pyrimidinergic receptor P2Y, G-protein coupled, 6

P2Y6

Gene references into function

Expression profile in human peripheral tissues and brain regions using PCR

Gene ontology

G-protein signalling, linked to IP3 second messenger (phopholipase C activating)

G-protein coupled receptor

G-protein coupled receptor protein signalling pathway

G-protein linked receptor

G-protein linked receptor protein signalling pathway

.....

Integral plasma membrane protein

Integral plasma membrane protein

KIAA0337 (71,626,609-71,685,818)

Gene ontology

Unknown

TNFRSF19L (71,693,096-71,713,375)

RELT, FLJ14993

Gene ontology

Unknown

PHRET1 (71,962,911-71,977,586)

KPL1, PHR1

Gene ontology

Phototransduction

Integral plasma membrane protein

Membrane fraction

Signal transduction

PME-1 (72,140,165-72,223,544)

Gene ontology

Protein demethylation

Protein phosphatase inhibitor

Hydrolase

DKFZP586N2124 (72,930,425-72,944,986)

Gene ontology

Unknown

FLJ22596 (72,932,227-72,933,652)

Gene ontology

Unknown

KIAA0102 (73,182,947-73,211,393)

Gene ontology

Unknown

NEU3 (73,221,793-73,241,353)

SIAL3, neuraminidase 3 (membrane sialidase), neuraminidase 3 (ganglioside sialidase)

Gene ontology

Enzyme

Ganglioside

Integral plasma membrane protein

Hydrolase

SLX21A9 (73,384,775-73,439,911)

OATPB, OATP-B, KIAA0880

ARRB1 (73,499,777-73,521,035)

ARR1

Gene ontology

Cytoplasm

Plasma membrane

Soluble fraction

Enzyme inhibitor

Signal transduction

Heterotrimeric G-protein complex

Peripheral plasma membrane protein

Inhibitor/repressor

Protein translocation

Complete assembly protein

RPS3 (73,633,188-73,639,347)

405 Ribosomal protein S3

Gene ontology

Oncogenesis

RNA binding

Protein biosynthesis

Structural protein of ribosome

Cystolic small ribosomal (40S) subunit

PP1665 (73,668,299-73,675,453)

Gene ontology

Unknown

SERPINH2 (73,795,926-73,806,445)

CBP2, HSP47, colligin-2, collagen-binding protein 2 (collagen 2)

Gene ontology

Collagen binding

Heat shock protein

Endoplasmic reticulum

FLJ22644 (73,951,475-73,962,985)

Gene ontology

Unknown

DGAT (74,021,778-74,035,187)

GS1999 full

Gene ontology

Expressed in liver, white adipose tissue (PMID 11481335)

UVRAG (74,047,885-74,376,860)

Gene ontology

Cytoplasm

DNA repair

Embryonic expression

Axial asymmetry

WNT11 (74,419,980-74,440,184)

HWNT11

Gene ontology

Embryogenesis and morphogenesis

Signal transduction

Cell-cell signalling

Embryonic and foetal expression

Expressed during development

Embryonic expression of WNT14B in mice

PRKRIR (74,583,614-74,614,490)

Inhibitor of protein kinase PKR

Gene ontology

Stress response

Protein binding

Signal transduction

Translational regulation

Negative control of cell proliferation

Embryonic expression

C11ORF30 (74,680,593-74,783,885)

GL002, EMSY protein, EMSY

Gene ontology

Unknown

GARP (74,891,183-74,903,592)

D11S833E, garpin

Gene ontology

Integral plasma membrane protein

Associated with UVRAG

E2IG4 (75,029,256-75,031,796)

DKFZP586E011, DKFZP586011 Protein

Gene ontology

Unknown

PHCA (75,094,572-75,256,521)

APHC, FLIJ11238

Alkaline phytoceramidase

CAPN5 (75,300,663-75,357,854)

HTRA3, nCL-3

Gene ontology

Calpain

Protease (other than proteasomal)

Hydrolase

Signal transduction

Embryonic expression

OMP (75,336,495-75,336,986)

Olfactory marker protein

Gene ontology

Olfaction

Cytoplasm

Signal transduction

Synaptic transmission

Chemosensation and response

Foetal expression in rats

Expressed during development in humas

Foetal and embryonic expression in humans

MYO7A (75,361,919-75,448,892)

DFNB2, NSRD2, VSH1B, DFNA 11, Myosin VIIA

Deafness, autosomal dominant 11

Deafness, autosomal recessive 2

Gene ontology

Vision

Hearing

Myosin ATPase

Cell structure

Hydrolase

Motor protein

Cytoskeletal

Contains an internal repeat

PAK1 (75,556,638-75,707,585)

PAKalpha, P21/Cdc42/Rac1-activated kinase 1 (years Ste20-related)

Gene ontology

JNK cascade

Protein kinase

Protein phosphorylation

Transferase

Embryonic expression in mice

HBXAP (75,849,807-75,871,455)

XAP8

HBV pX associated protein-8

Hepatitus B virus-associated protein

ARIX (75,883,451-75,887,841)

Gene ontology

Regulate gene expression, morphogenesis & differentiation

Transciption factor

Neurotansmitter synthesis & storage

Transciption from Pol II promoter

Neuronal development & Transmission

DNA binding protein

DNA associated (indirect/direct)

Left-right asymmetry determination

Associated with craniofacial abnormalities

CLNS1A (75,889,882-75,937,186

CLC1, ICLN, ICLn, CLNS1B

Gene ontology

Vision

Circulation

Plasma membrane

Small molecule transport

Auxiliary transport protein

Photoreception

Integral membrane

Similar properties to UVRAG

All information obtained on each gene is adapted from LocusLink (NCBI website) - April 2001

APPENDIX 13

Patient consent form for DNA Banking and/or Analysis

APPENDIX 13

Patient Consent form for DNA banking and/or analysis

(Printed on Edith Cowan University Letterhead)

I, ...

of .....

hereby consent to .....

being sampled from me/my .....

for the purpose of DNA banking and/or analysis.

In addition I consent to my genetic material (DNA) being used for the development of genetic tests for the disorder carried by myself or my family and consent to my DNA being used for research purposes, provided this research is approved by the appropriate Hospital Ethics Committee.

I understand that not all attempts to establish immortal cell lines are successful, and that it may be necessary to give a further sample. If such attempts are successful, then I understand that;

1) My genetic material will be available for repeated analyses for an indefinite period of time.

2) My genetic material will be used for diagnostic purposes for myself and for such members of my family in whom it may be of use and who desire such a diagnostic service. My genetic material may also be used for research.

3) Traditional principles of medical confidentiality shall apply.

4) If the genetic test I require is not currently available, my genetic material may be supplied to another collaborating agency, after discussion with you.

The purpose and nature of the above-mentioned procedure has been explained to me.

Signature.....

Dated .....

Witness declaration:

I, ...

Have read over and explained to the consenting party the purpose and nature of the above-mentioned procedure, who stated they understood and affixed their signature in my presence.

APPENDIX 14

UVRAG exon 15 coding sequence

APPENDIX 14

UVRAG Exon 15 coding sequence

ttagctcagacctaagtatgctgttccctgtctctggatccagaaggcttacactgagttctgaagt
ccaaagtaacttcttgttttgtttctcttctagTGACAGACATCACACCTCCAGTGCAA
TCCCTGTTCCTAAGAGACAAAGCTCCATATTTGGGGGTGCAGATGTAGGC
TTCTCTGGGGGGATCCCTTCACCAGACAAAGGACATCGAAAACGGGCCAGC
TCTGAGAATGAGAGACTTCAGTACAAAACCCCTCCTCCCAGTTACAACTC
AGCATTAGCCCAGCCTGTGACCACCGTCCCCTCCATGGGAGAGACCGAGAG
AAAGATAACATCTCTATCCTCCTCCTTGGATACCTCCTTGGACTTCTCCAA
AGAAAACAAGAAAAAAGGAGAGGATCTAGTTGGCAGCTTAAACGGAGGC
CACGCGAATGTGCACCCTAGCCAAGAACAAGGAGAAGCCCTCTCCGGGCAC
CGGGCCACAGTCAATGGCACTCTCCTACCCAGCGAGCAGGCCGGGTCCGCCA
GTGTCCAGCTTCCAGGCGAGTTCCACCCAGTCTCAGAAGCTGAGCTCTGCT
GTACTGTGGAGCAAGCAGAAGAAATCATCGGGCTGGAAGCCACAGGTTTC
GCCTCAGGTGATCAGCTAGAAGCATTTAACTGCATCCCAGTGGACAGTGC
TGTGGCAGTAGAGTGTGACGAACAAGTTCTGGGAGAATTTGAAGAGTTC
TCCCGAAGGATCTATGCACTGAATGAAAACGTATCCAGCTTCCGCCGGCCG
CGCAGGAGTTCCGATAAGTGAAGTGAGCAGGTCAACAGTAGGACTGGGGC
AGAAGCTCTGCCTAAAATGAAGTGAAAGCTGCACTTAACCCTTTGTGATA
ATGATGACACAAAATGAATATTAATGGAGGATATTCCTCGGAAAAACAG
ACTTTGGGAATGAAGGAGGGACTCAGGATCATTGTTATCAGTGGGCCAAA
GTTAGATTTTGCTTTCAAGATTTGCTTTTCGGGCCTGATGATTTTAAAGC
AAAAATCACCCTCTAGTTGAAAGAGCTTACAGCTCGAGTCACCTTTTAGC
TATTTGTCTGCTTTTTATTTACCCTTGTATGTTATCCTCAGAGGGAAGAT
GATAATATATAAATAATATAATGAACACACCCTTAGTTTCTCATAAGCA
TTTGCCCTCACCATGGTTTATAAAACTTTGGGAAAACGGAATATTCAGAA
ATAGGTTTCCGCCATGTACTGAAAGGTCTGTGGCCATCTGTGAGGTAGAT
GAAGAAGCAGCATAGTGGTCTCCTTACATCTAGGCCTAACTGTCCCTCTTC
CTGCCCCCGGGTACCACAGTCCACCTTTAGACCCTACTGTCGCCCCATCTTC
TCCGTGGATGGGCCATGCGTCCTGAAAACAGGACATCAGATTCACTGGTT
CTGTAACCCAGTAGCTGTGACGTTCCATCTCTTCTAACCAGCCATGGCCTT
CCCCTCCTCTGCCATACCCTTAATGCGGCCCTCAGATTAGATGAAAAACTT
GCTCCTGGTGGATCCCAAGGGACCCTCAAGGACCTCGAGGTTACTGCAGTC
AGATGCCATCTCATCCCCTGTGGGGGCCAAAGTTTTTATGTGGGCAGATG
CTGTGGTCAGGAACTAGGCATGCTTTCTGGCAATGCACTCACCAGACAAA
AATCCCTTGATGTAAATCCCATGTTAATTTATTAAATTTCAGTCAGAAGG
TCAGCATTTACATGACAGAATGTATGTAGAGAGTTGGGGTGTCTGGTAG
GCAAACTGCAAGGCAGTTGAGATAGTTGGATTAAGAGGCTAGACGAGAC
ATAGAATACTATTGGTATGTGTGCAATTTCATGAATATTAAATTATGTT
TCGAAGTCCAGTTGTCATTCCCGCATTCAGATTTCATTTGCTGTTGCTTTA

TACGTTACGTACCCAAGGACATTGCCTCAGGGTTGCAAACTCTTTAAAGG
AAAATTTATCCATATATCCATGTATTATATAGAAGAATAAAAATTGAGT
TTACTTC

Upper case depicts exonic sequence

Lower case depicts 100 intronic bases prior to start of exon

Adapted from the Draft Human Genome, UCSC, 2002

APPENDIX 15

Identification of mouse genes homologous to candidate HFM genes on chromosome 11

APPENDIX 15

Identification of mouse genes homologous to candidate HFM genes on chromosome 11

Candidate gene	Human Gene	Mouse chromosome	cM	Mouse gene
11q12-q13	COX8	19		Cox8a
11q12-q13	EMK1	19	3	Emk
11q13	STIP1	19		Stip1
11q13	KCNK4	19	4.5	Kcnk4
11q12	ESRRA	19	3	Esrra
11q11-q13	RPS6KA4	19		Rsp6ka4
11	MGC13045	19		Pygm
11q13	NRXN2	19		Nrxn2
11q13	RASGRP2	19		Rasgrp2
11q12-q13.2	PYGM	19	2	Pygm
11q13	MAP4K2	19		Map4k2
11q13	MEN1	19		Men1
11q13	EHD1	19	2	Ehd1
11q13	C11orf5	19		ORF6
11	POLA2	19		Pola2
11q13	REQ	19	2	Req
11q12	LTBP3	19	2	Ltbp3
11q13	DOC-1R	19		5830466O21Rik
11cen-q12.1	CHK	19	3	Chk
11q13.4	LRP5	19		Lrp5
11q13	C11orf23	19		9130026N02Rik
11q13.2-q13.3	MTL5	19		Mtl5
11q13.1-q13.2	CPT1A	19		Cpt1a
11q13.2-q13.4	IGHMBP2	19	0	Ighmbp2
11q13.3	FGF4	7	72.4	Fgf4
11q13	CCND1	7	72.3	CcndI
11q13.3	FADD	7	70	Fadd
11q13	EMS1	13	31	Cttn
11q13.2-q13.5	DHCR7	7		Dhcr
11q13.3-q14.1	FOLR1	7		Folr1
11q13.3-q13.5	FOLR2	7		Folr2
11q23	INPPL1			Inppl1
11q13	CGI-52			Pctpl
11	SKD3			Skd3

Candidate gene	Human Gene	Mouse chromosome	cM	Mouse gene
11q13.5-q14.1	P2RY2			P2ry2
11q13.5	NEU3	7		Neu 3
11q13	ARRB1	7	50	Arrb1
11	SERPINH1	7		Sepinh1
11q13.5	SERPINH2	7		Serpinh2
11q13.5	**WNT11**	7	48b	Wnt11
11q13.5	PRKRIR	7		2900052B10Rik
11q13.5-q14	**GARP**	7	46	Garp
11q14	CAPN5	7		Capn5
11q13.5	MYO7A	7	48.1	Myo7a
11q13-q14	PAK1	7	46.5	Pak1
11q13.5-q14	**CLNS1A**	7	50	Clcni

Bold case depicts candidate genes identified in this project
Chromosome 11 candidate genes not listed in this table have no homologous mouse gene

Adapted from USCS Human-Mouse homology Map

http://www.ncbi.nlm.nih.gov/Homology/view.cgi?map=ucsc_mgd&chr=11&tax_id=9606&mode=text

APPENDIX 16

Markers for Chromosome 14 Candidate Region

APPENDIX 16

Markers for chromosome 14 linkage region

D14S1142 (92,880,607-92,880,769)

SYNONYMNS -

PRIMERS - 5' TTGCAGGGAGTCAGGTGTATG

3' GCATCACACAGAGACACGGAT

D14S1143 (93,119,794-93,119,956)

SYNONYMS -

PRIMERS - 5' ATACATATATACGTATACACAT

3' GTAAGCAATGAGCAATGATT

GATA168F06 (93,507,497-93,507,718)

SYNONYMS - CHLC.GATA168F06.P66298, D14S1434

PRIMERS - 5' ACAATTCCAGAAACTTCCCC

3' ATCAGTGAGCCAATTCCTTG

D14S987 (95,080,876-95,081,183)

SYNONYMS - stCP4371, 28967, D14S987, AFM161yd12, Z51047, RH31328, 161yd12, HS161YD12, STS3601, SHGC-20885, DBSTS:28967, DBSTS:46779, AFM161YD12

PRIMERS - 5' ACCATATAGGGTGACGATGA

3' AGCTGAACTATTTTAATTCAATTGT

D14S65 (96,108,679-96,108,841 & 96,108,697-96,108,847)

SYNONYMS - D14S65, 538, RH73259, RH49006, SHGC-707, SU884, RH800, RH12952, Z16553, AFM093yg5

PRIMERS - 5' GCTCCACCCCCTAAAGATC

3' TCAATACACCCTGTGGAAAG

D14S267 (97,711,801-97,712,038 & 97,711,831-97,712,043)

SYNONYMS - 263wh9, RH84477, rh53507, RH3660, HS263WH9, STS29598, RH85839, gdb:199593, SHGC-1405, STS59558, Z23865, D14S267, 1797, DBSTS:1797, AFM263WH9, AFM263wh9, RH9623, RH15357

PRIMERS - 5' TTAATGCCCACTGAATGCT

3' AAGGCAGCCCTGGTTT

Above information was taken from ENSEMBL database (June 2002)

APPENDIX 17

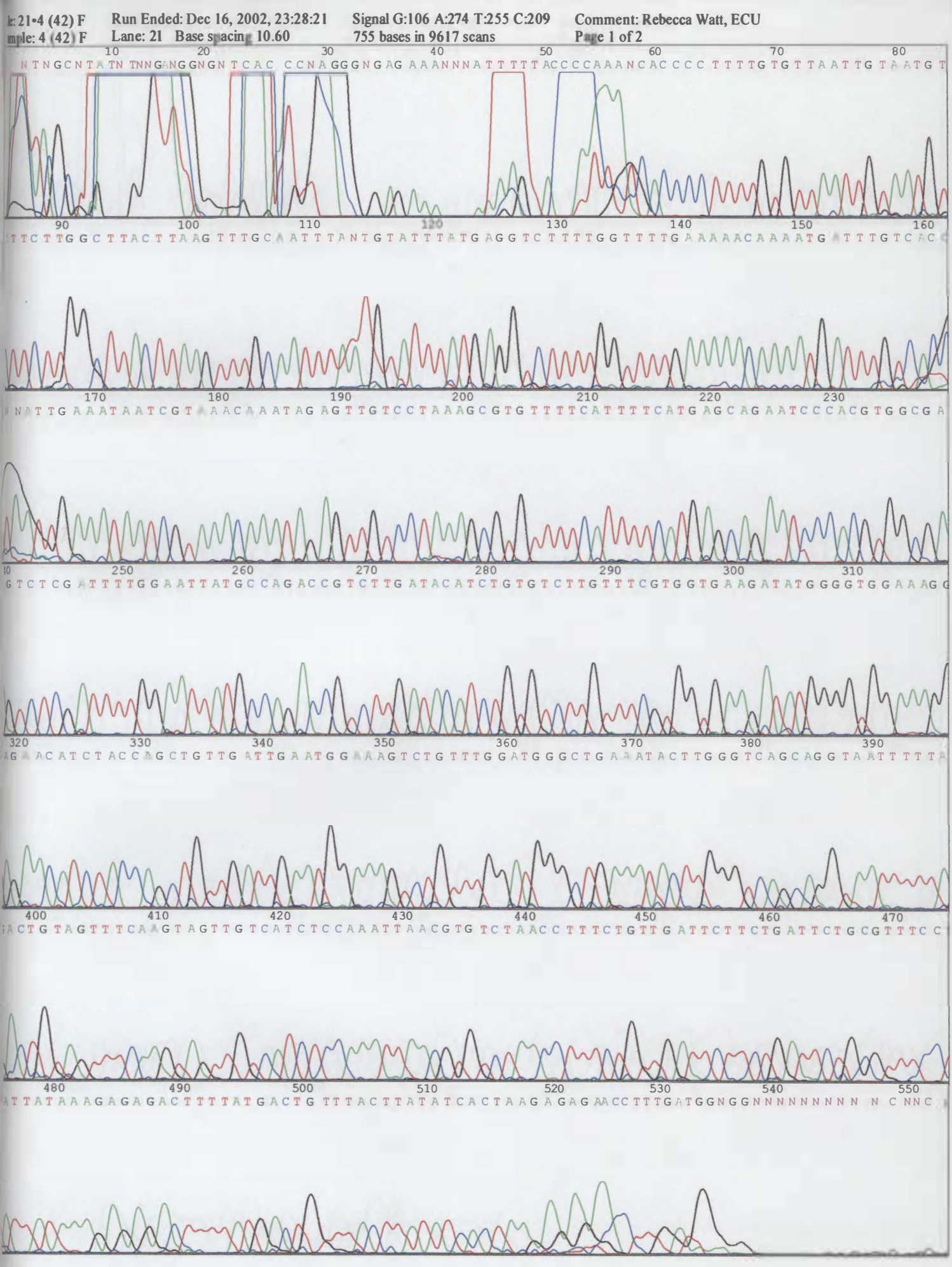
Chromatograms of genes sequenced on chromosome 11

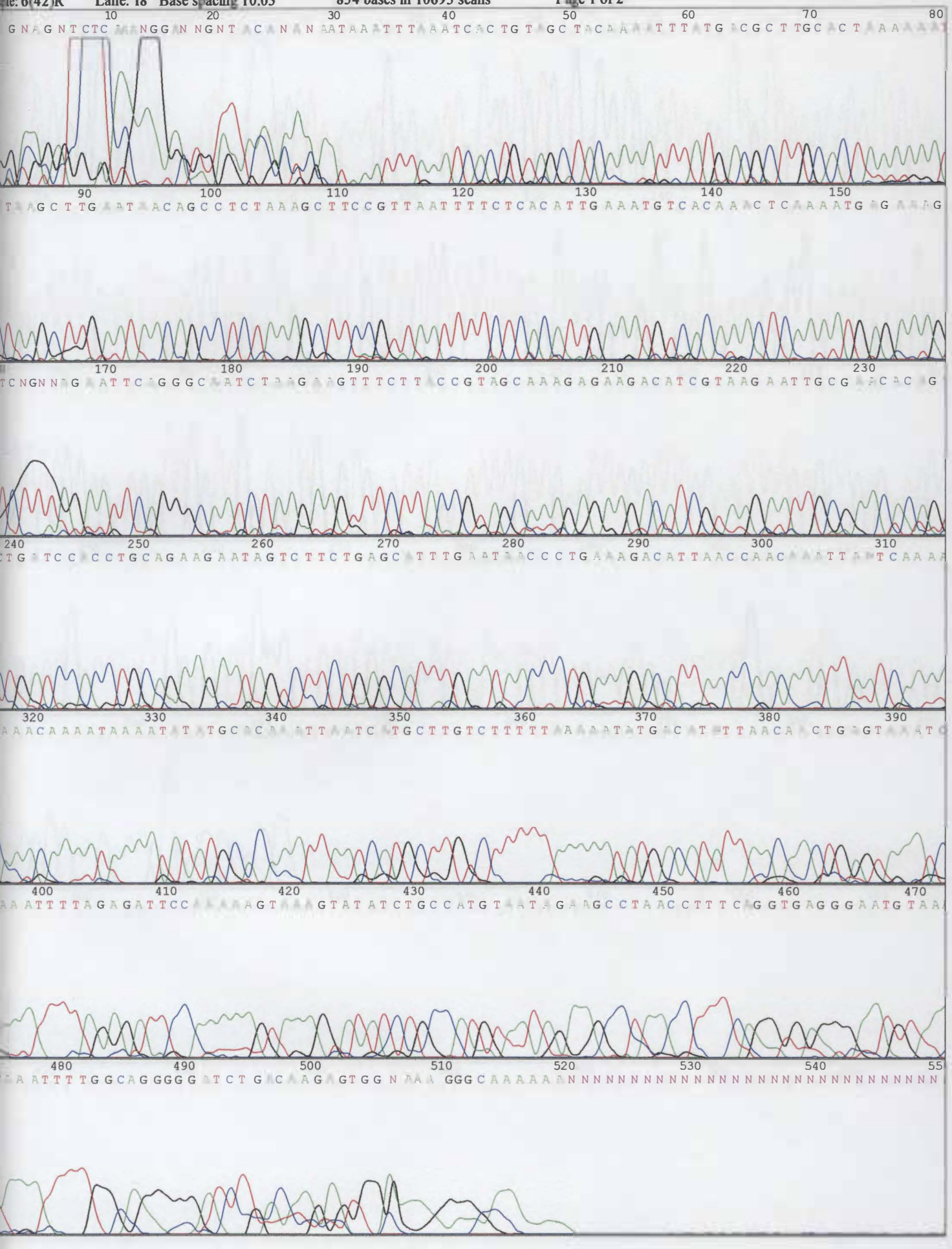
APPENDIX 17

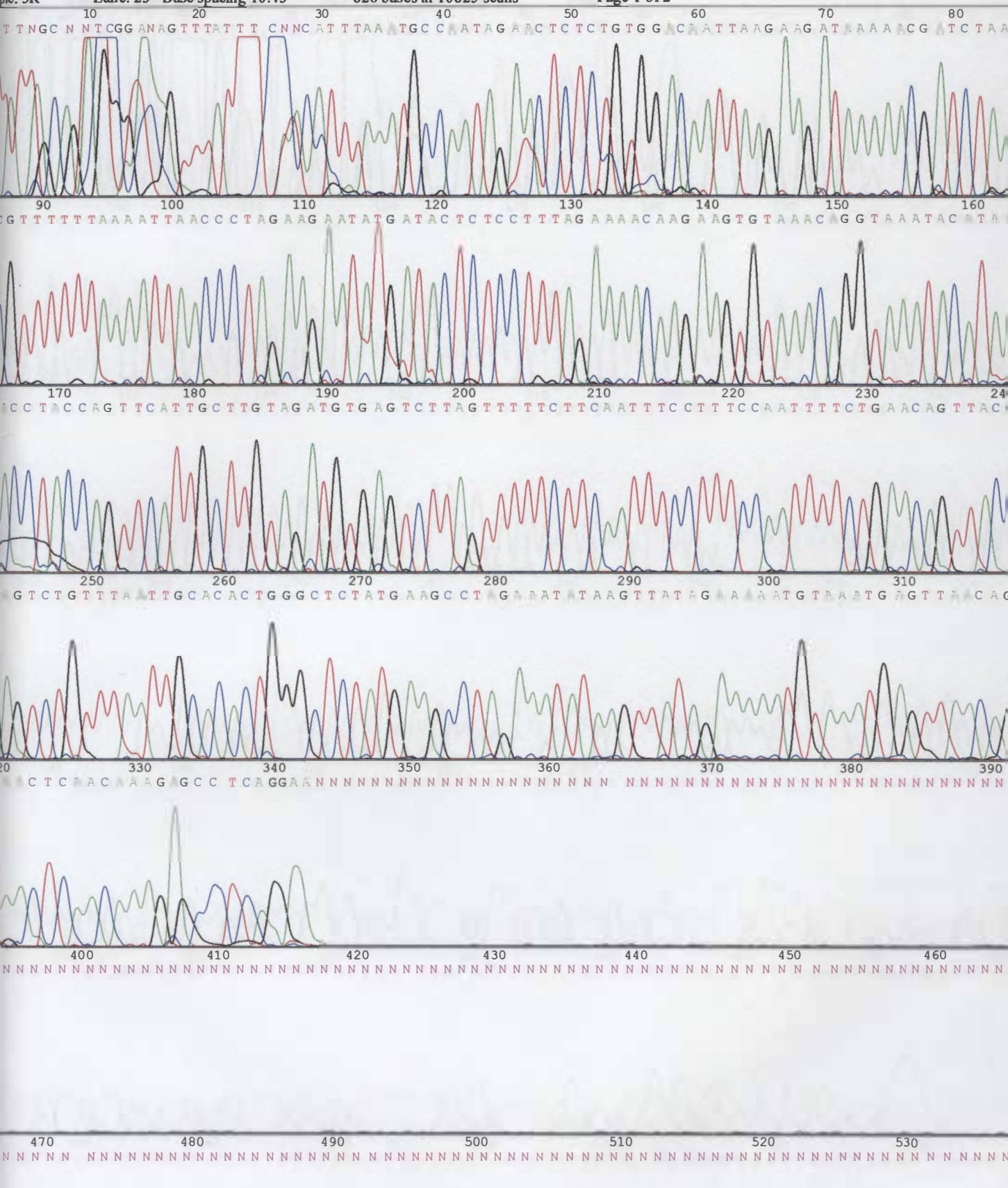
Chromatograms of genes sequenced on chromosome 11

In order of presentation:

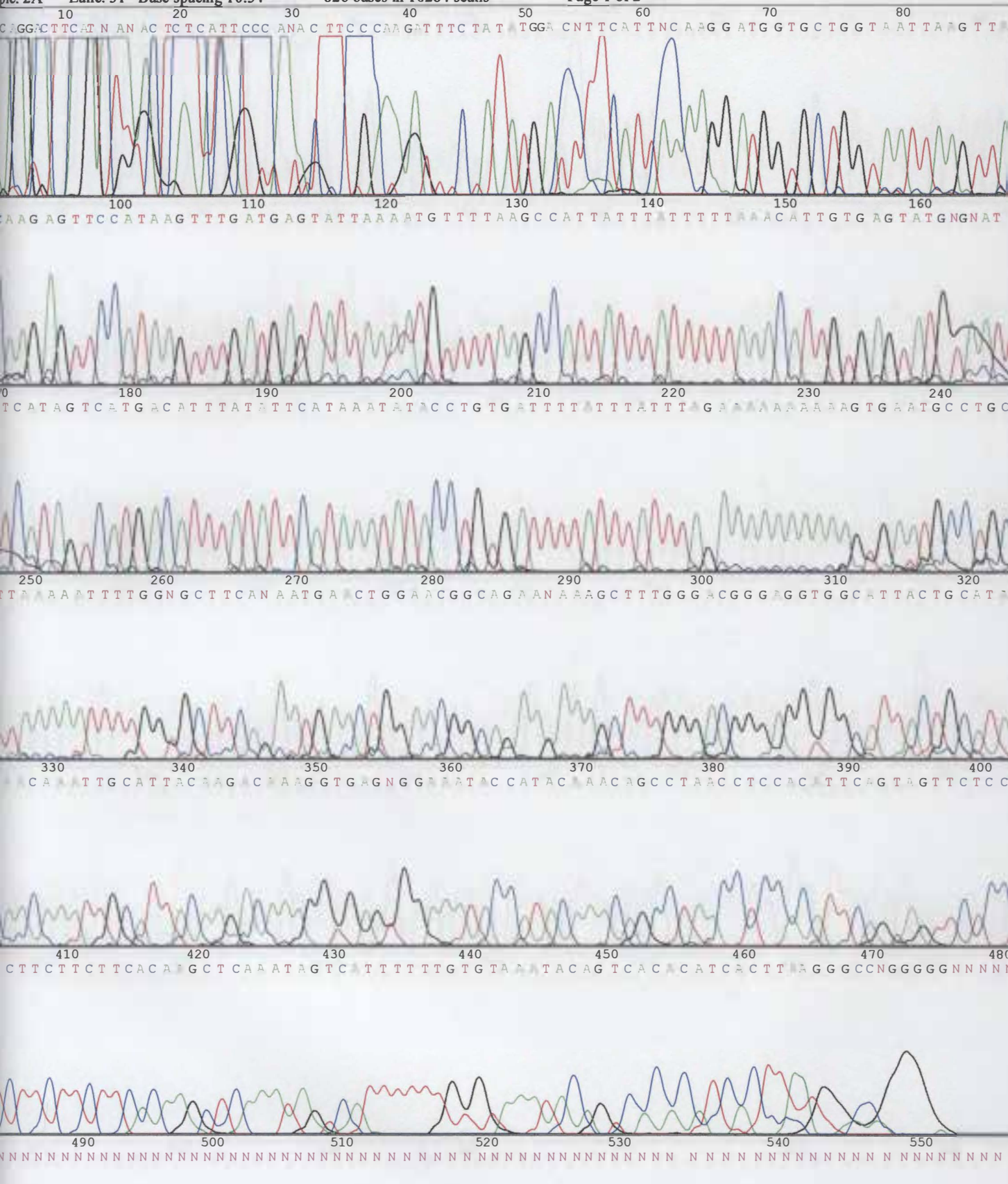
Gene	Exon	Sequence direction
UVRAG	exon 4	forward sequence
UVRAG	exon 6	reverse sequence
UVRAG	exon 7	reverse sequence
UVRAG	exon 8	forward sequence
UVRAG	exon 10	forward sequence
CLNS1A	exon 1	forward sequence
CLNS1A	exon 4	forward sequence
CLNS1A	exon 4	reverse sequence
CLNS1A	exon 5	forward sequence
CLNS1A	exon 6	forward sequence
GARP	exon 1	forward sequence
GARP	exon 2	forward sequence
GARP	exon 3(ii)	forward sequence
GARP	exon 3(v)	reverse sequence


le: 21•4 (42) F
mple: 4 (42) F
Run Ended: Dec 16, 2002, 23:28:21
Lane: 21 Base spacing 10.60
Signal G:106 A:274 T:255 C:209
755 bases in 9617 scans
Comment: Rebecca Watt, ECU
Page 1 of 2

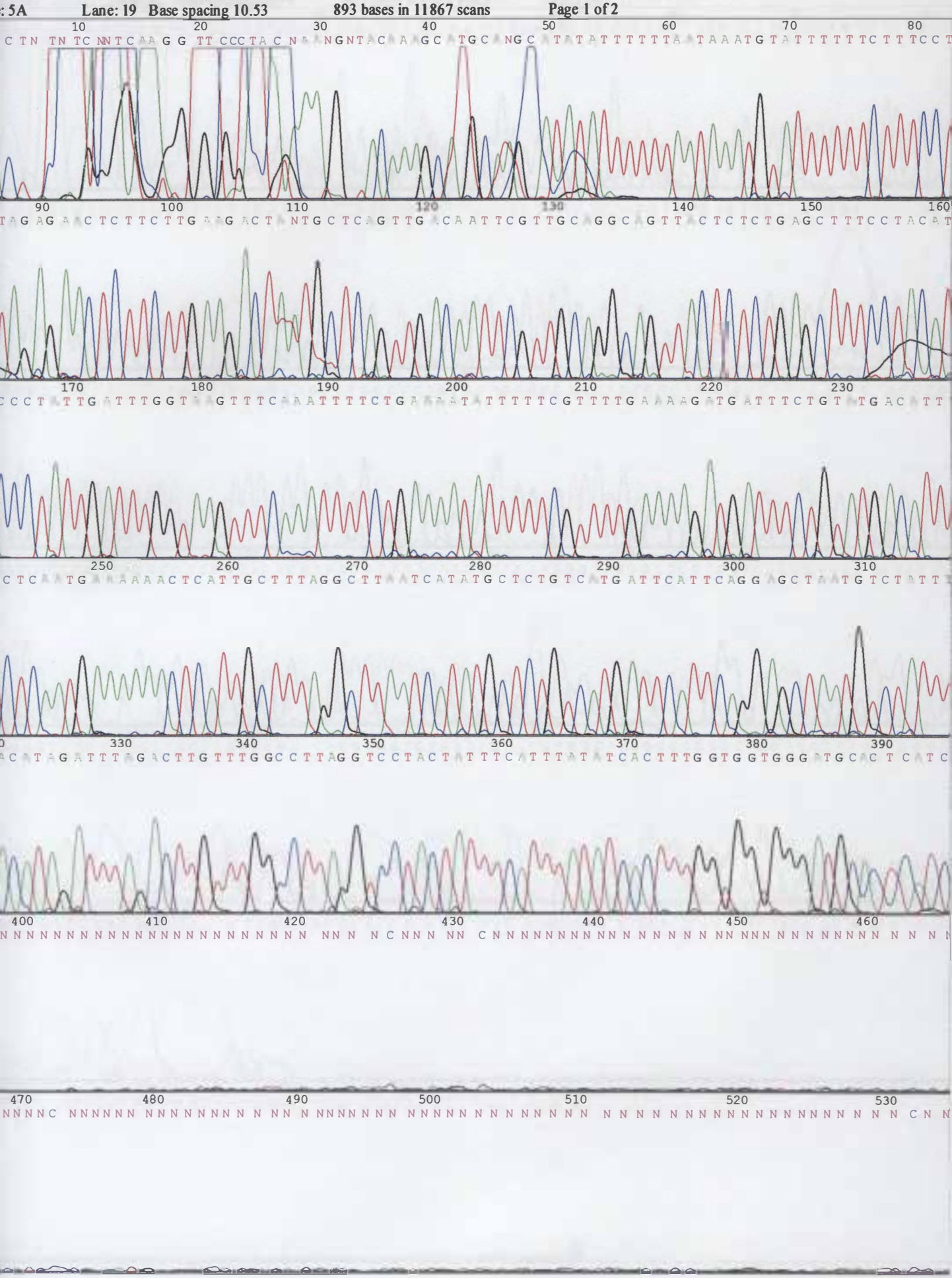




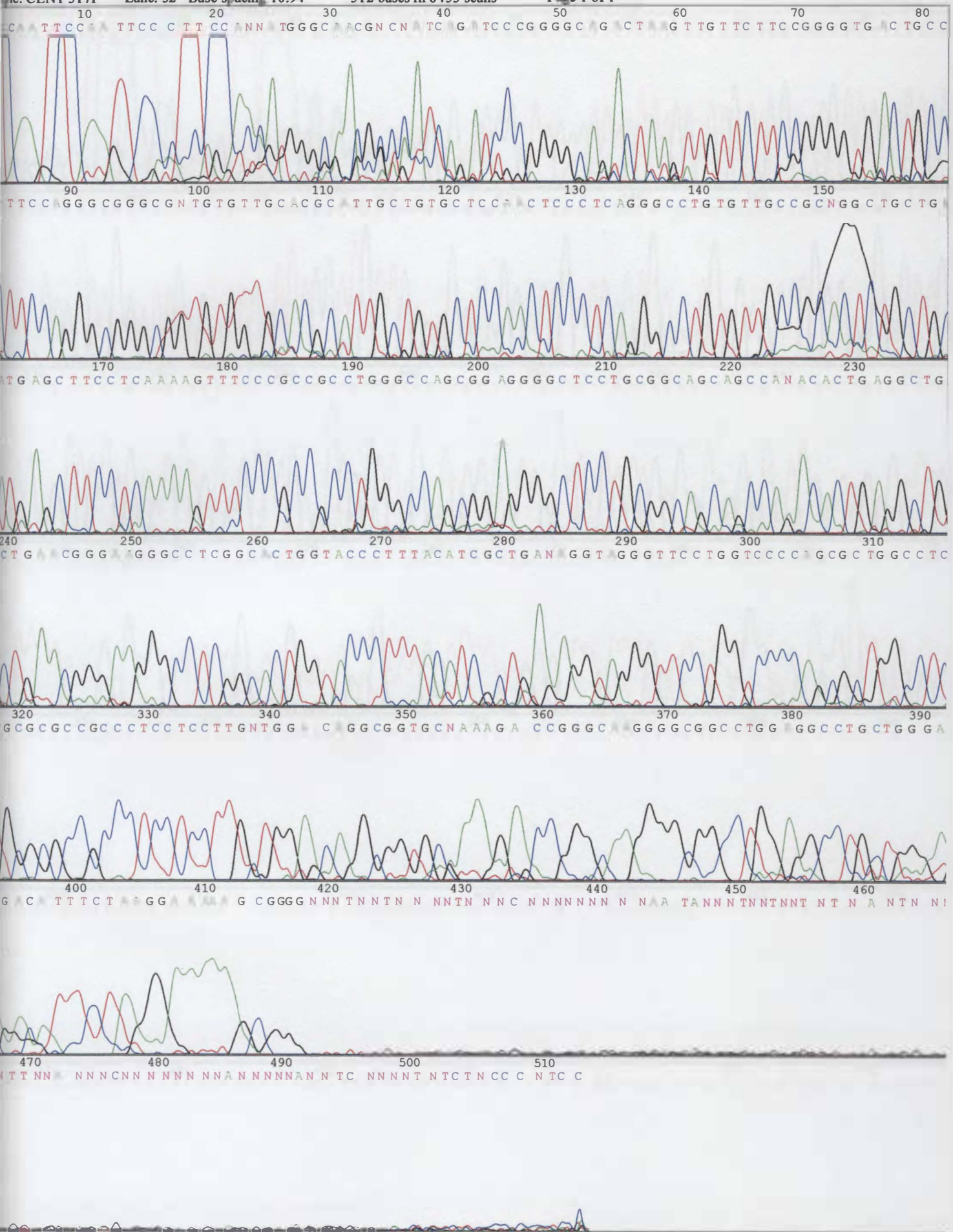
31•2A Run Ended: Dec 9, 2002, 23:23:42 Signal G:143 A:342 T:352 C:183 Comment: Rebecca Watt, ECU
ple: 2A Lane: 31 Base spacing 10.54 826 bases in 10284 scans Page 1 of 2

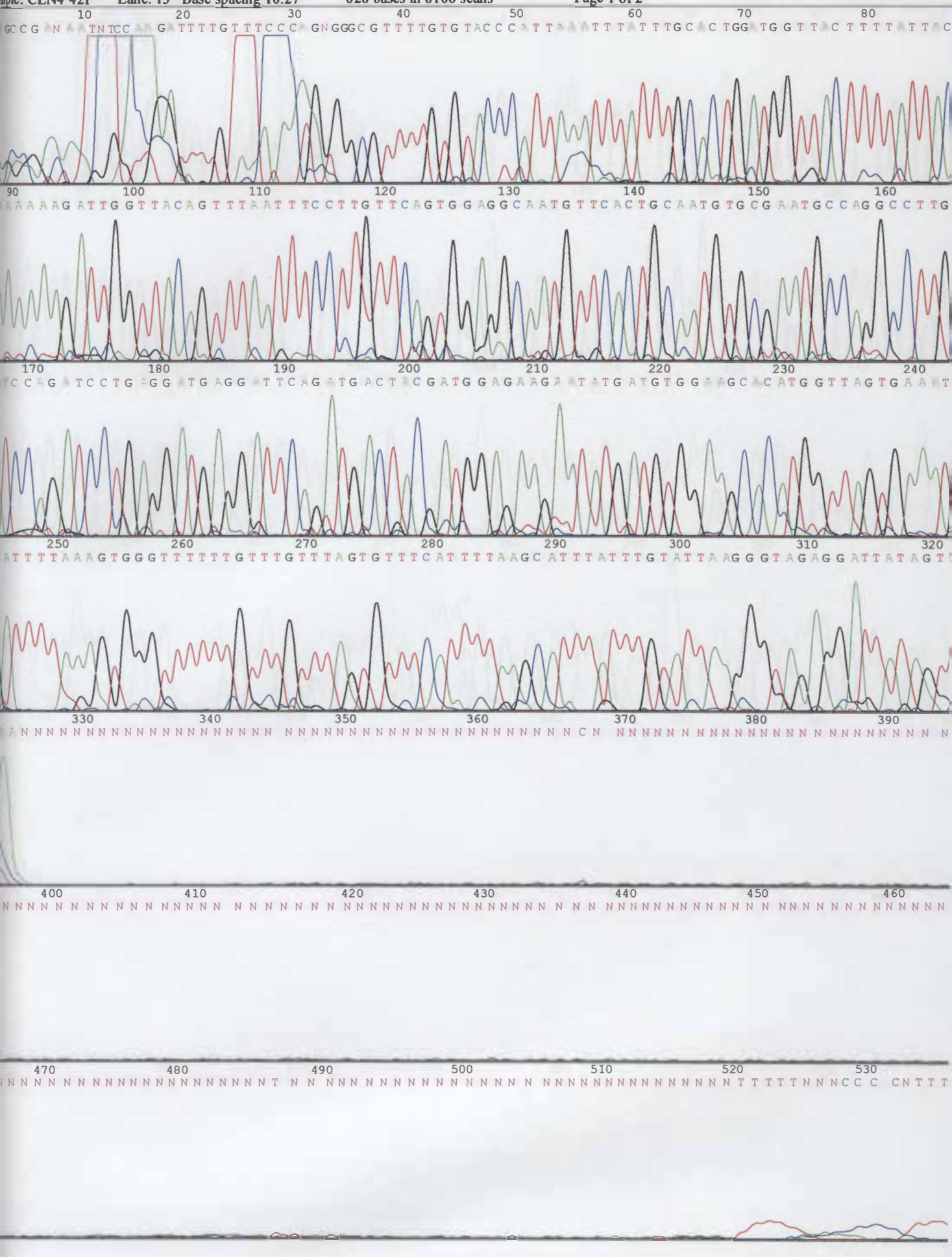


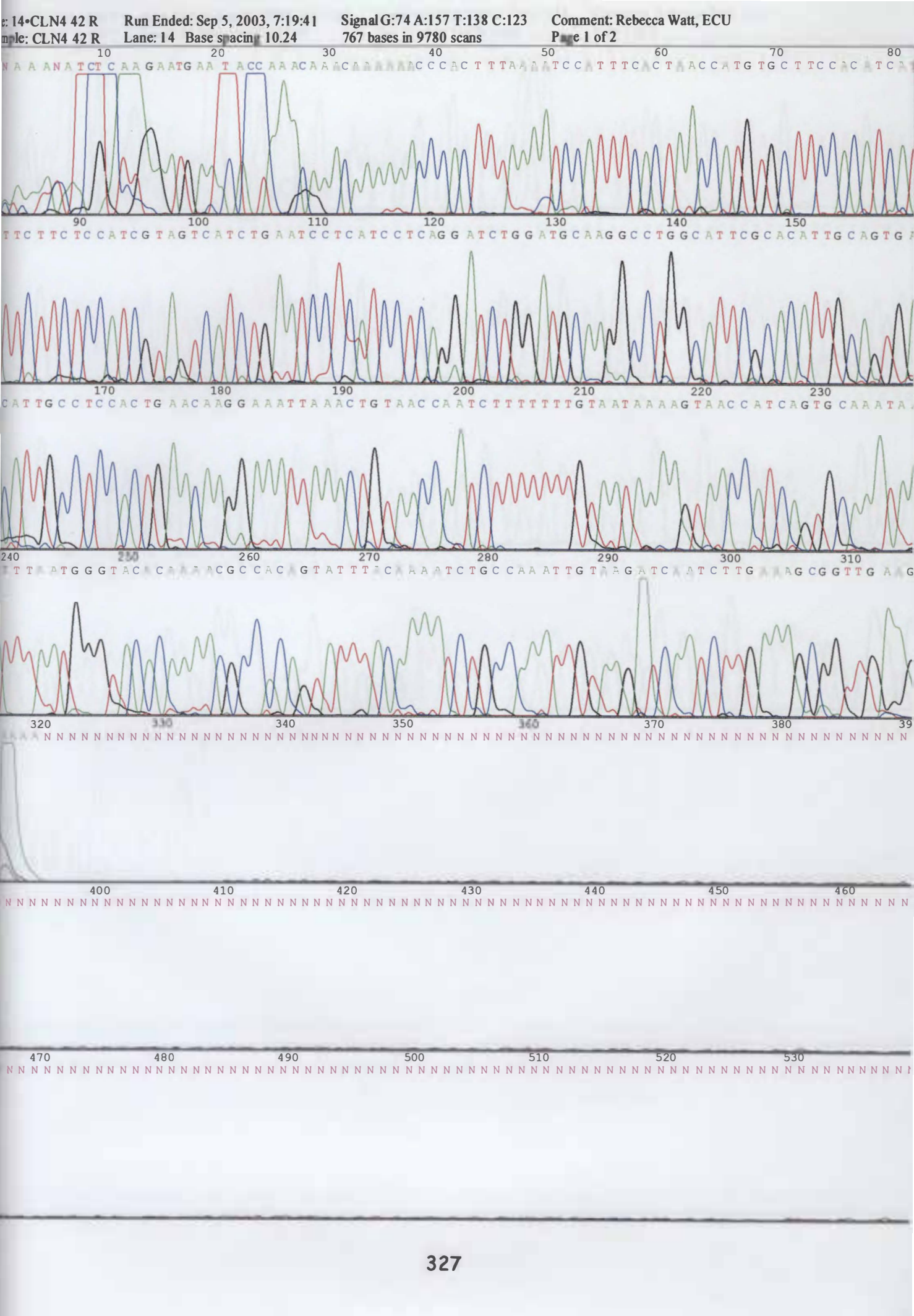
on10.ab1 Run Ended: Dec 11, 2002, 22:36:35 Signal G:70 A:141 T:206 C:91 Comment: Rebecca Watt, ECU
: 5A Lane: 19 Base spacing 10.53 893 bases in 11867 scans Page 1 of 2

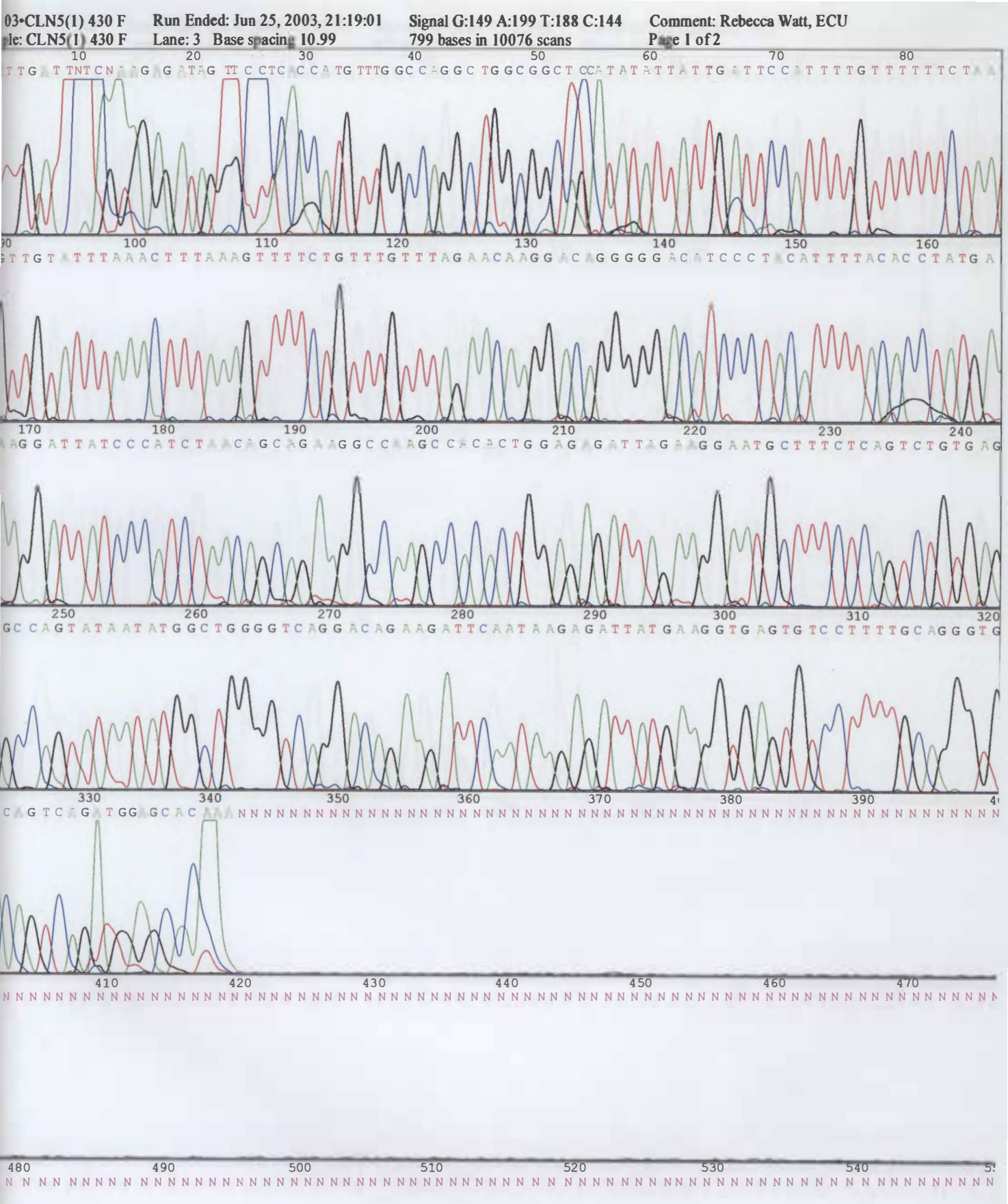


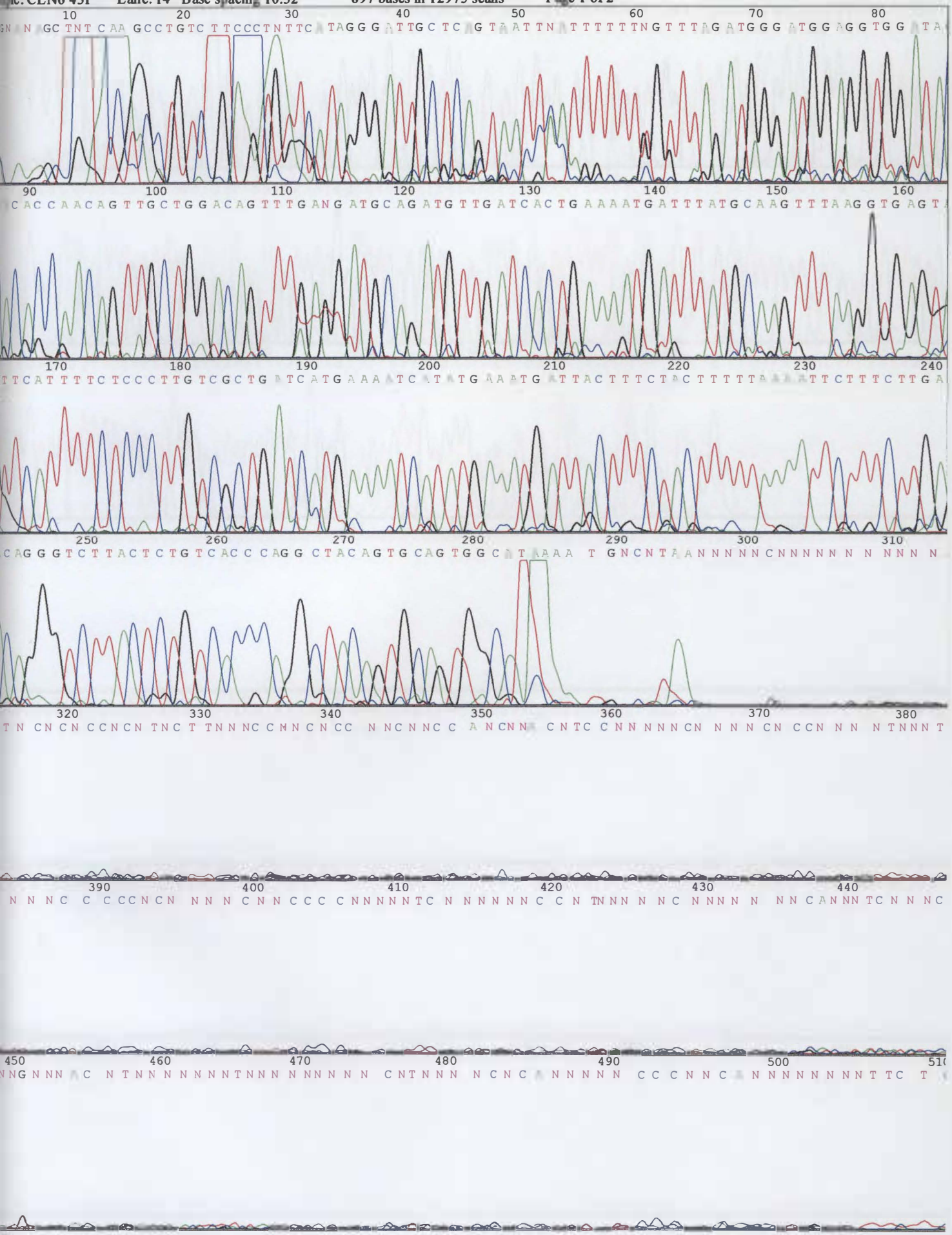
32•CLN1 517F Run Ended: Sep 2, 2003, 7:31:31 Signal G:65 A:59 T:61 C:82 Comment: Rebecca Watt, ECU
le: CLN1 517F Lane: 32 Base spacing 10.94 512 bases in 6453 scans Page 1 of 1

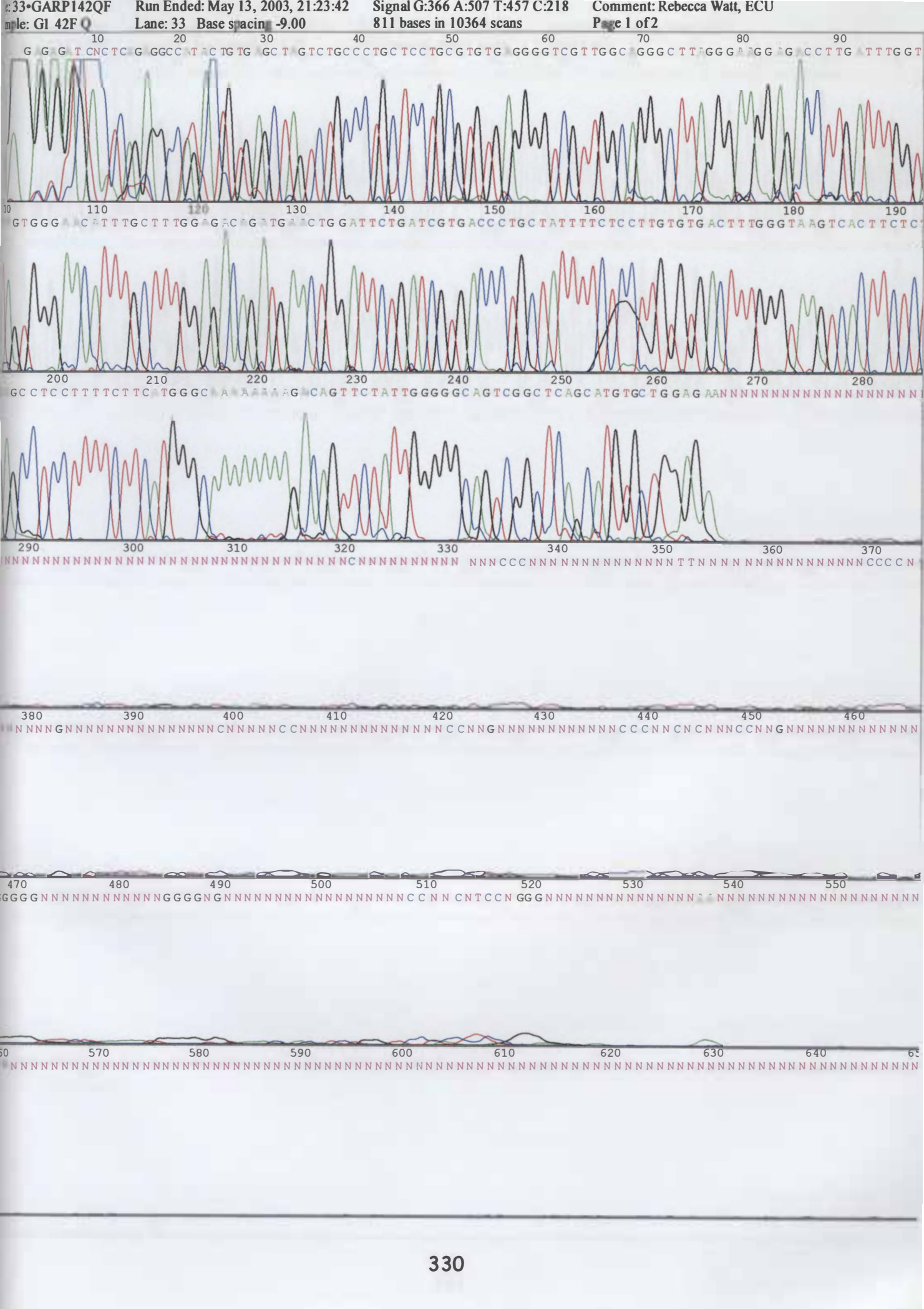


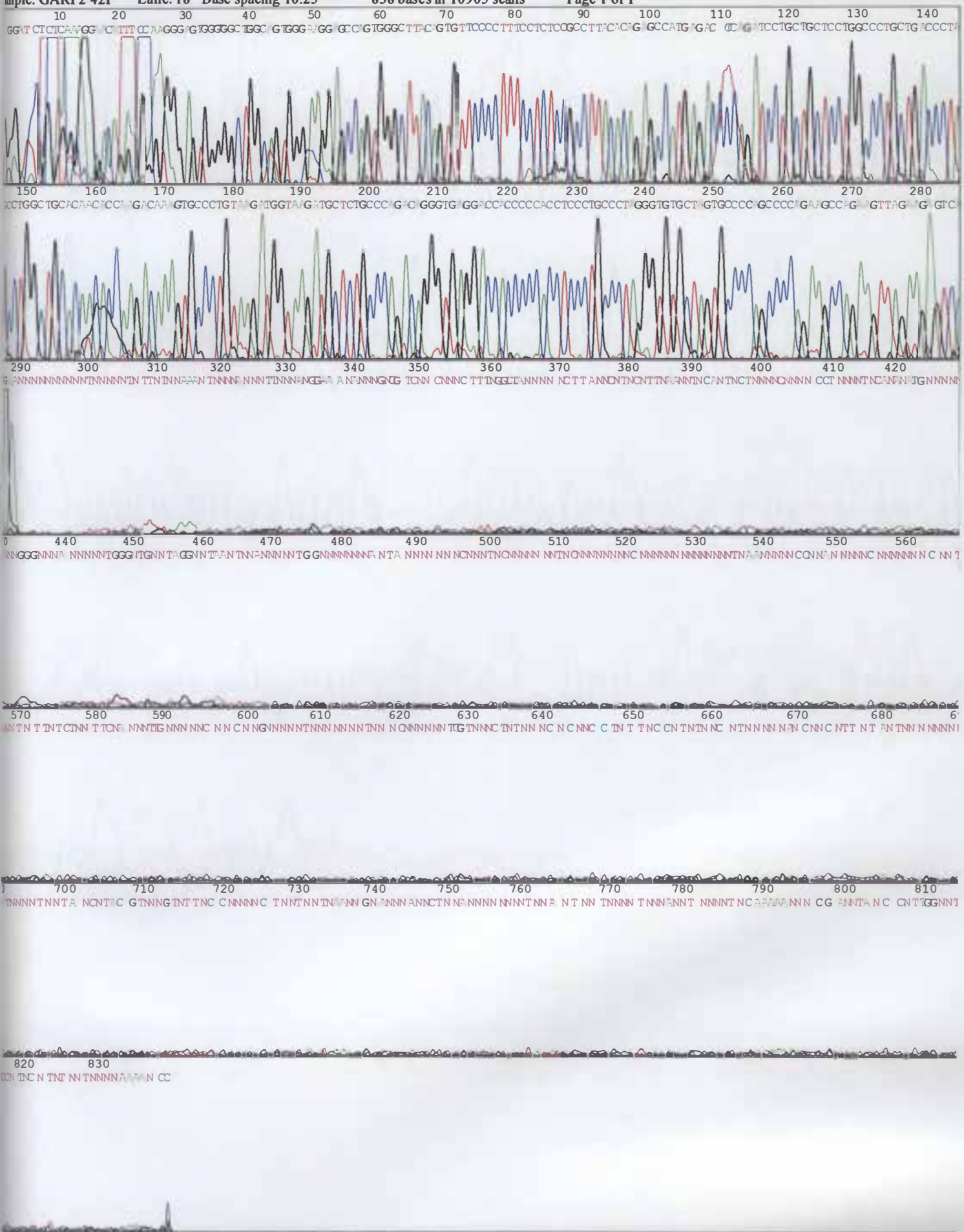


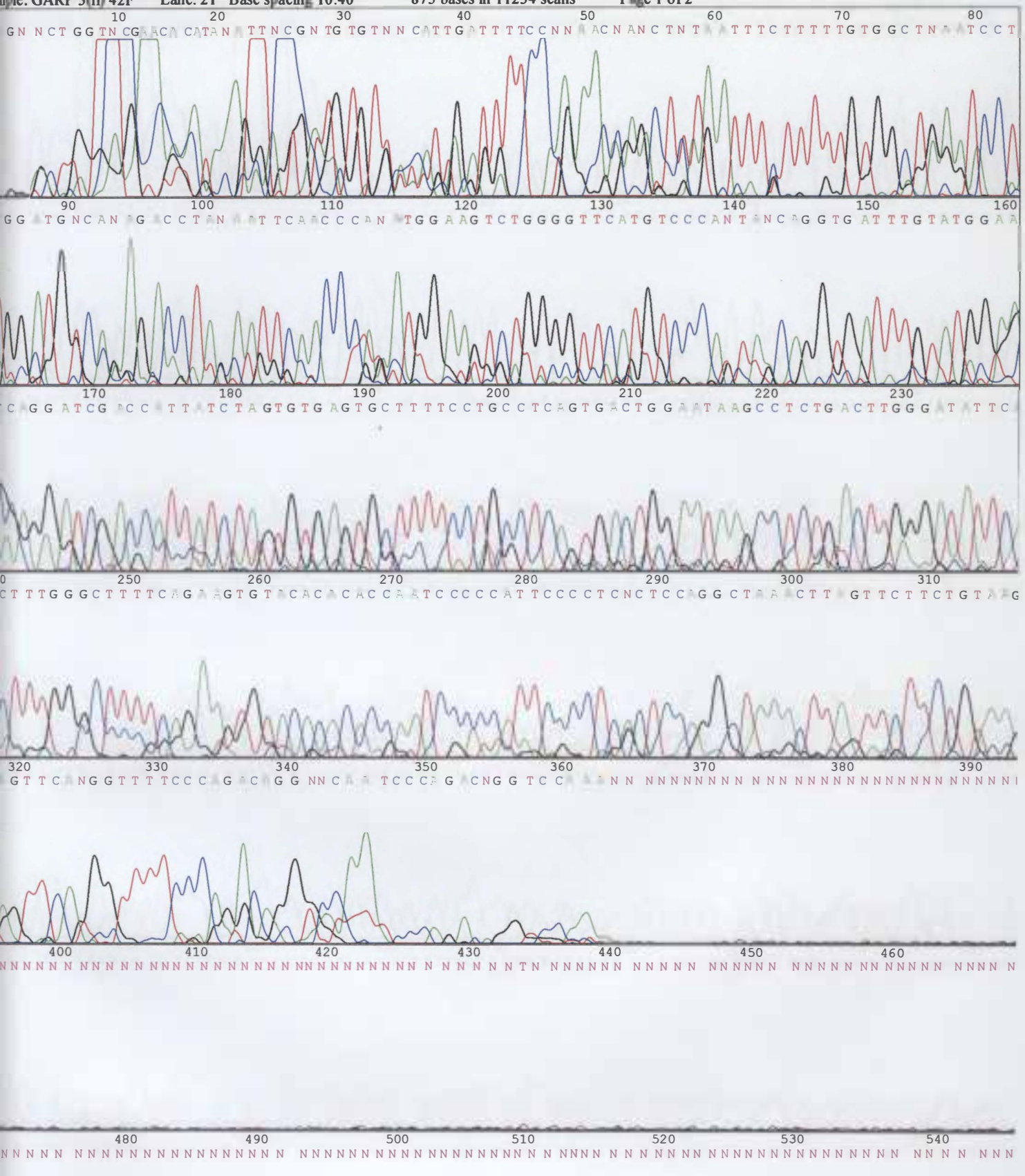

e: 14•CLN4 42 R
mple: CLN4 42 R
Run Ended: Sep 5, 2003, 7:19:41
Lane: 14 Base spacing 10.24
Signal G:74 A:157 T:138 C:123
767 bases in 9780 scans
Comment: Rebecca Watt, ECU
Page 1 of 2

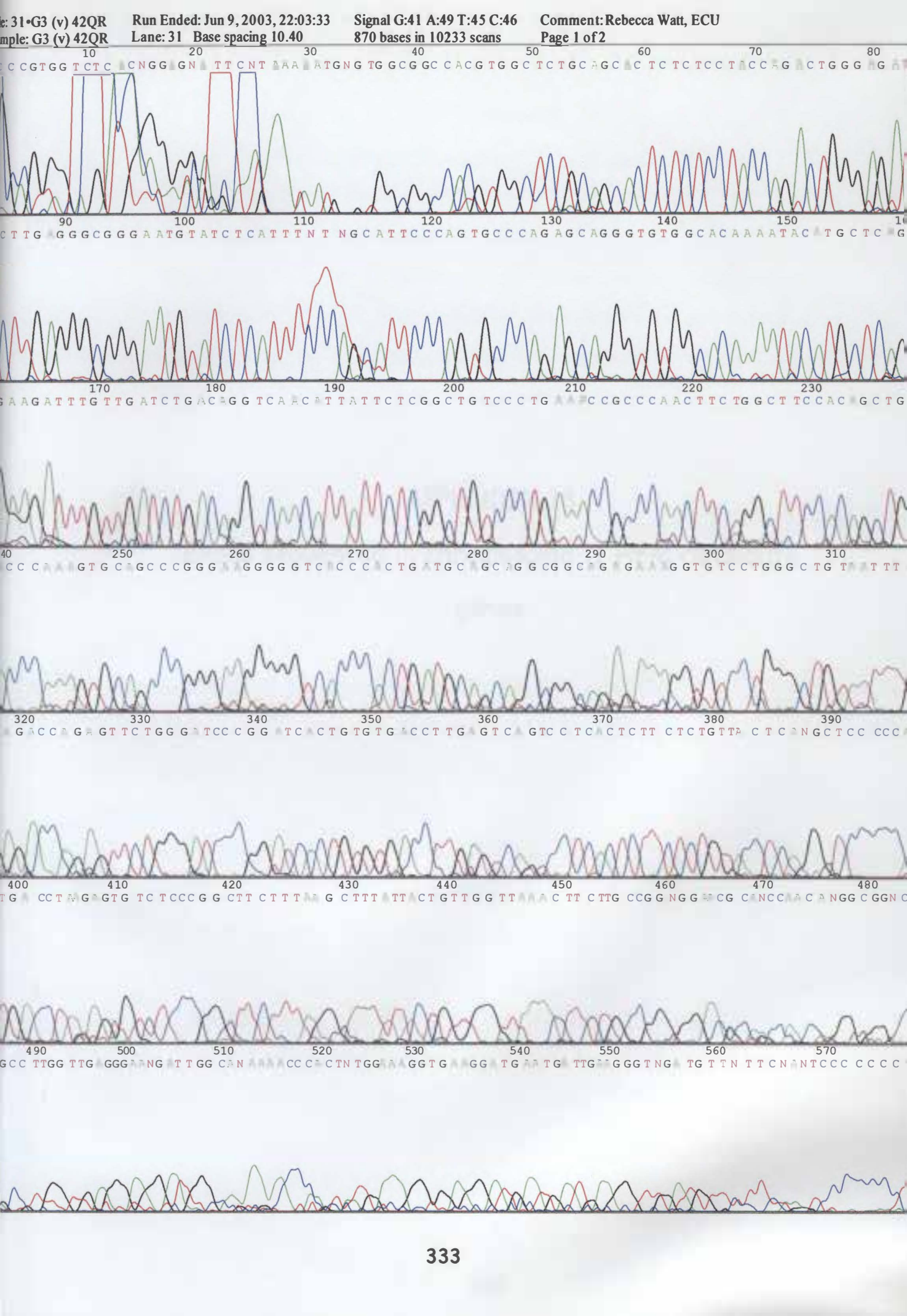

03•CLN5(1) 430 F
Run Ended: Jun 25, 2003, 21:19:01
Signal G:149 A:199 T:188 C:144
Comment: Rebecca Watt, ECU
le: CLN5(1) 430 F
Lane: 3 Base spacing 10.99
799 bases in 10076 scans
Page 1 of 2




33•GARP142QF
Run Ended: May 13, 2003, 21:23:42
Signal G:366 A:507 T:457 C:218
Comment: Rebecca Watt, ECU
Lane: 33 Base spacing -9.00
811 bases in 10364 scans
Page 1 of 2


e: 18•GARP2 42F
Run Ended: Jun 9, 2003, 22:03:33
Signal G:61 A:70 T:71 C:83
Comment: Rebecca Watt, ECU
mple: GARP2 42F
Lane: 18 Base spacing 10.25
838 bases in 10905 scans
Page 1 of 1




e: 31•G3 (v) 42QR
mple: G3 (v) 42QR
Run Ended: Jun 9, 2003, 22:03:33
Lane: 31 Base spacing 10.40
Signal G:41 A:49 T:45 C:46
870 bases in 10233 scans
Comment: Rebecca Watt, ECU
Page 1 of 2

APPENDIX 18

Exons sequenced in chromosome 11 HFM candidate genes

APPENDIX 18

Exons sequenced in chromosome 11 HFM candidate genes

Sequence Results for *UVRAG* and *ARIX*

Region	***UVRAG***	**Region**	***ARIX***
EXON 1	No variation	EXON 1	non definitive
EXON 2	No variation	EXON 2	non definitive
EXON 3	No variation	EXON 3	non definitive
EXON 4	No variation		
EXON 5	No variation		
EXON 6	No variation		
EXON 7	No variation		
EXON 8	No variation		
EXON 9	No variation		
EXON 10	No variation		
EXON 11	No variation		
EXON 12	No variation		
EXON 13	No variation		
EXON 14	No variation		
EXON 15(i)	non definitive		
EXON 15(ii)	non definitive		
EXON 15(iii)	non definitive		
EXON 15(iv)	non definitive		

Table Sequence Results for *CLNS1A* and *GARP*

Region	*CLNS1A*	Region	*GARP*
EXON 1	No variation	EXON 1	No variation
EXON 2	No variation	EXON 2	No variation
EXON 3	No variation	EXON 3(i)	No variation
EXON 4	No variation	EXON 3(ii)	non definitive
EXON 5	No variation	EXON 3(iii)	non definitive
EXON 6	No variation	EXON 3(iv)	No variation
EXON 7(i)	non definitive	EXON 3(v)	No variation
EXON 7(ii)	non definitive	EXON 3(vi)	No variation
		EXON 3(vii)	No variation
		EXON 3(viii)	No variation