Molecular Genetics of Epilepsy and Mental Retardation Limited to Females (EFMR)

A thesis submitted for the degree of Doctor of Philosophy to the University of Adelaide

By Kim Hynes

School of Molecular and Biomedical Science, Division of Genetics, University of Adelaide

December 2009
# Chapter 1 – Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Intellectual disability</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>X-linked intellectual disability</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Strategies for identifying causative genes for XLID</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Epilepsy in ID</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Idiopathic epilepsy</td>
<td>7</td>
</tr>
<tr>
<td>1.6</td>
<td>Epilepsy and mental retardation limited to females</td>
<td>9</td>
</tr>
<tr>
<td>1.6.1</td>
<td>Clinical description</td>
<td>9</td>
</tr>
<tr>
<td>1.6.2</td>
<td>Inheritance pattern</td>
<td>11</td>
</tr>
<tr>
<td>1.6.3</td>
<td>Molecular analysis and mode of inheritance of EFMR</td>
<td>15</td>
</tr>
<tr>
<td>1.7</td>
<td>Cadherins</td>
<td>16</td>
</tr>
<tr>
<td>1.8</td>
<td>Protocadherins</td>
<td>20</td>
</tr>
<tr>
<td>1.8.1</td>
<td>Protocadherin functions</td>
<td>27</td>
</tr>
<tr>
<td>1.9</td>
<td>PCDH19</td>
<td>30</td>
</tr>
<tr>
<td>1.10</td>
<td>Project aims</td>
<td>32</td>
</tr>
</tbody>
</table>
1.1 Intellectual disability

Intellectual disability (ID), also termed mental retardation (MR) (Salvador-Carulla and Bertelli, 2008), is a common disorder affecting an estimated 1.5-2% of the population in Westernised countries (reviewed in Leonard and Wen, 2002). ID is defined by the American Association on Mental Retardation (AAMR) as a disability characterized by significant limitations both in intellectual functioning and in adaptive behaviour that originates before age 18 (Luckasson et al., 2002). ID is subdivided into five categories based on a patient’s intelligence quotient (IQ).

<table>
<thead>
<tr>
<th>Class</th>
<th>Intelligence quotient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profound</td>
<td>&lt;20</td>
</tr>
<tr>
<td>Severe</td>
<td>20-34</td>
</tr>
<tr>
<td>Moderate</td>
<td>35-49</td>
</tr>
<tr>
<td>Mild</td>
<td>50-69</td>
</tr>
<tr>
<td>Borderline</td>
<td>70-85</td>
</tr>
</tbody>
</table>

Table 1.1 The IQ criterion used to classify the severity of ID based on WHO guidelines (WHO, 1996)

The most common forms of ID within the population are mild to moderate (Toniolo, 2000, Vasconcelos, 2004). The least common form is severe ID which only affects 0.3-0.5% of the population (Toniolo, 2000, Vasconcelos, 2004). ID can present on its own or more commonly as one of the comorbidities of a more complex syndrome such as Down syndrome (Toniolo, 2000). ID can present in infancy and early childhood but it is commonly diagnosed during school years, as IQ tests are more objective and only reliable in children over five (Reviewed in Toniolo, 2000, Vasconcelos, 2004, Shevell et al., 2003).

The underlying causes of ID are highly heterogeneous and include both genetic and non-genetic factors. Non-genetic factors occur pre-natal or during early infancy and result in brain injury. Examples of these include infectious diseases, very premature birth, perinatal anoxia.
and foetal alcohol syndrome (Chelly and Mandel, 2001). Developmental disorders that affect the precise pattern of organization and connections in the brain can result in ID (Toniolo, 2000). Metabolic and mitochondrial disorders have also been associated with ID through inducing defects in neurons that alter brain housekeeping functions (Toniolo, 2000). This introduction will focus exclusively on genetic causes of ID and in particular X-linked intellectual disability (XLID).

1.2 X-Linked Intellectual Disability

XLID is defined by the occurrence of ID (IQ<70) where the causative gene has been mapped to the X-chromosome (Gecz et al., 2009). Historically the categories of XLID have be further subdivided into syndromic (S-XLID) and non-syndromic (NS-XLID) forms, depending on whether further consistent abnormalities (in addition to ID) are found on physical examination, laboratory investigation or brain imaging (Mulley et al., 1992). However with increased understanding of the molecular basis of XLID a proportion of NS-XLID are turning out to be S-XLID. Once a causative gene has been identified in a multiple NS-XLID patients the specific phenotypes of these individuals with mutations in this gene can be compared as a group against normal individuals. Specific syndromal features can then be indentified for the group. For example, Fragile(X) (Fra(X)) syndrome which was initially characterised as NS-XLID, but is now the most widely recognised example of S-XLID (Ropers and Hamel, 2005). Although some syndromologists can pick some cases of Fra(X) by clinical observation alone, many Fra(X) cases are not so easily diagnosed and require laboratory testing for a definitive diagnosis. A further complexity to the S-XLID / NS-XLID subdivision is that mutations in a gene can result in patients with S-XLID and patients with NS-XLID. For example, mutations in ribosomal protein S6 kinase (RSK2) cause Coffin-Lowry syndrome (CLS) (Trivier et al., 1996) a disorder characterised by severe psychomotor retardation, facial and digital
dysmorphisms and progressive skeletal deformations (Young, 1988). Mutations in RSK2 have also been identified in patients with NS-XLID who do not exhibit any of the facial, digital or skeletal features typical of CLS (Merienne et al., 1999).

Genes on the X-chromosome have long been implicated in ID due to the ~25-30% excess of males with ID compared to females (Turner, 1996). More realistic recent estimates predict a contribution from the X-chromosome of between 8-15% (Ropers, 2008, Ropers and Hamel, 2005). This is 2-4 times higher than expected based on the ~4% of the protein coding genes in the human genome being coded for on the X-chromosome (Ropers, 2008). The X-chromosome is a rich source of causative genes for ID.

The human X-chromosome contains ~818 known protein-coding genes (Gecz et al., 2009). Roughly 40% of X-chromosome genes are expressed in the brain and are therefore candidate genes for XLID (Ropers and Hamel, 2005). Out of these a large number of genes have been screened for mutations in XLID patients and so far more than 90 genes have been implicated in XLID (Chiurazzi et al., 2008, Tarpey et al., 2009). However mutations in these genes only explain ~50% of families with suspected XLID (reviewed in Gecz et al., 2009).

The absence of a definitive diagnosis leaves affected families without accurate genetic counselling and reproductive options such as prenatal diagnosis. ID has a high burden on society, not just on those families directly affected, as it also ranks highly in terms of health care expenditure in developed countries. ID is therefore one of the most important health issues in medicine. Whilst considerable effort has been put into the discovery of genes responsible for ID the landscape is far from complete and requires much more investigation to complete the picture.
1.3 Strategies for identifying causative genes for XLID

The identification of causative genes for XLID is not only important at a clinical level to aid those families and individuals affected, but it will also increase our understanding of normal brain function. For example, investigations into the function of FMR1, the gene involved in Fra(X) syndrome, have added to the knowledge of fundamental aspects of normal neuronal function (Willemsen et al., 2004).

Gene discovery is considered easier in the case of S-XLID, compared to NS-XLID, as affected families and individuals can be classified and sub-grouped on the basis of physical, metabolic and/or behavioural characteristics, in addition to ID. Linkage data can therefore be combined for families and individuals in a particular sub-group, which greatly enhances the power of linkage analysis and has the potential to significantly reduce the linkage interval to be assessed for candidate genes. The task of gene discovery in monogenic disease has become easier with the availability of the human genome sequence and increased knowledge of gene function (Ropers and Hamel, 2005). However, on the other hand the unexpected degree of genetic heterogeneity and mutation pleiotropy observed for some disorders represents a considerable challenge, as seen for disorders caused by ARX (Aristaless-related homeobox) (Gecz et al., 2006) and ATRX (Gibbons and Higgs, 2000) mutations.

Gene discovery for NS-XLID cases remains extremely difficult. Families with NS-XLID are impossible to separate in terms of their phenotype and need to be dealt with family by family because as a group they are clinically homogeneous. Progress in NS-XLID has been slow with less that 50% of causative genes identified due to extreme genetic heterogeneity which prevents the pooling of linkage information from unrelated families (Ropers and Hamel, 2005).
In recent years the strategies used to identify causative genes in XLID have changed significantly. Previous success in gene identification has arisen through the use of breakpoint mapping in patients with rearrangements involving the X-chromosome, positional candidate screening of genes associated with fragile sites and linkage mapping with candidate gene screening (reviewed in Ropers and Hamel, 2005). These were primarily conducted on a patient/patient basis. More recent approaches to gene identification have seen a shift to semi-automated high throughput mutation-detection screening in large cohorts of families. Analysing linkage intervals from 125 families with NS-XLID revealed a potential mutation hotspot with clustering of the underlying mutations to the Xp11.2-11.3 which was predicted to carry 30% of causative genes (Ropers et al., 2003). Subsequent mutation screening of 50 genes expressed in the brain from the Xp11.2-11.3 region identified 5 novel XLMR genes (Ropers and Hamel, 2005, Freude et al., 2004, Jensen et al., 2005, Kalscheuer et al., 2003a, Shoichet et al., 2003, Tarpey et al., 2004). A further large-scale screen for causative point mutations in 90 known and candidate XLID genes in 600 suspected XLID families identified 73 mutations in 21 genes (de Brouwer et al., 2007). From this analysis they identified a causative mutation in 42% of families that contained an obligate female carrier (de Brouwer et al., 2007). It is likely that some of the 600 families analysed in this screen may not have ID due to X-linked genes, this is particularly relevant for the 191 families which only had affected individuals in a single generation with 2-5 affected brothers, there was still a significant proportion of XLID families with no apparent mutation in any of the previously identified and suspected XLID genes.

To identify additional causative genes for XLID our laboratory in collaboration with others from the USA and UK embarked on a large scale, comprehensive screen, via high-throughput re-sequencing, of 718 X-chromosome genes in probands from 208 XLID or suspected XLID families (Tarpey et al., 2009). This effort led to the identification of at least 12 novel
XLID genes and contributed to the identification of 4 additional XLID genes (Tarpey et al., 2009). Several other putative XLID genes are yet to be analysed from the dataset of unique variants generated.

Structural sequence variants including deletions, duplications, translocations and inversions are also emerging as significant contributors to XLID. Three major duplication regions have been associated with XLID: Xp28 (involving $MECP2$), Xp11.22 ($HSD17B10$, HECT, UBA, $HUWEI$) and Xq26.2-q27 ($SRY$ and $SOX3$). Four deletions have also been associated with XLID: Xp22.3 (involving Kallmann syndrome), Xp11.4 ($CASK$), Xp11.3 (believed to involved $ZNF674$) and deletions involving $MECP2$ (reviewed in Gecz et al., 2009). Large scale screening for structural variants in XLID will likely identify additional regions and genes implicated in XLID.

### 1.4 Epilepsy in ID

Epilepsy is a frequent feature of many disorders with ID. The prevalence of epilepsy in the general population is 0.7 to 1.0% (reviewed in van Blarikom et al., 2006). However the prevalence of epilepsy in ID is much more frequent, with estimates ranging from 15 to 50%. A prevalence of 25% is believed to be a reasonable and conservative estimate (reviewed in van Blarikom et al., 2006). A search for epilepsy and mental retardation in OMIM returns 200 hits. These include syndromes such as: Rett syndrome (caused by mutations in $MECP2$)(Jellinger, 2003), West syndrome ($ARX$)(Hirose and Mitsudome, 2003), or Borjeson-Forssman-Lehmann syndrome - BFLS ($PHF6$)(Lower et al., 2002). The comorbidity of epilepsy and ID is not limited to syndromic forms of XLID as mutations in known XLID genes have also been identified in patients diagnosed with non-syndromic XLID, for example,
mutations in ARX have been identified in over 30 patients with NS-XLID (Gecz et al., 2006, Stromme et al., 2002).

1.5 Idiopathic epilepsy

Approximately 30% of patients with epilepsy have a known cause of their seizures, known causes can include head trauma, stroke, tumor or congenital lesion (Annegers et al., 1996). The remaining 70% of patients have no known or obvious extraneous cause for their seizures, these patients are classified as having idiopathic epilepsy. Idiopathic epilepsy is believed to have a strong genetic aetiology with a concordance rate of 70% in monozygous twins (reviewed in Helbig et al., 2008). Multiple families have also been identified with several affected members (Marini et al., 2004).

Idiopathic epilepsies are sub-divided into monogenic or complex epilepsies according to their mode of inheritance. Monogenic idiopathic epilepsy syndromes are caused by large effect mutations in a single gene, typically an ion channel gene. Mutations in 10 genes have been identified in various types of monogenic idiopathic epilepsies these are KCNQ2, KCNQ3, SCN2A, SCN1A, SCN1B, GABRG2, GABRA1, CLCN2, CHRNA4, CHRNB2 (reviewed in Helbig et al., 2008, Mulley et al., 2005). Mutations in these genes have been identified in numerous families with monogenic idiopathic epilepsy syndromes. For example, mutations in SCN1B have been identified in 75-80% of Dravet syndrome patients (Claes et al., 2003, Depienne et al., 2009c, Harkin et al., 2007, Marini et al., 2007).

Significant progress has been achieved in the identification of genes involved in the monogenic idiopathic epilepsies; however, the large families used in these gene discovery studies are not representative of the majority of epilepsy patients. Only 1-2% of idiopathic
epilepsies appear to be monogenic, with the vast majority believed to be polygenic with complex inheritance (Weber and Lerche, 2008). The underlying genetic architecture of the complex epilepsies is believed to be quite different from that of the monogenic epilepsies (Mulley et al., 2005). Unlike the monogenic epilepsies where a rare mutation in a single gene results in epilepsy, complex epilepsies are predicted to arise when an individual has susceptibility alleles in multiple genes which when combined have sufficient effect to exceed an arbitrary seizure threshold (Mulley et al., 2005). In this scenario each susceptibility allele is insufficient to causes seizures on its own and requires additional susceptibility alleles and environmental factors to cause seizures (Mulley et al., 2005, Hunter, 2005).

Determining the genetic contributors to complex epilepsy is an enormously challenging task; however, at least two genes, CACNA1H and GABRD, have been identified as susceptibility genes for complex idiopathic epilepsies. Candidate gene screening followed by electrophysiological validation was used to identify both of these genes.

CACNA1H (Calcium channel, voltage-dependent, T type, alpha-1a subunit) was considered a good candidate gene for idiopathic epilepsies as it encodes a T-type calcium channel. In excess of 150 variants have since been identified in patients with childhood absence epilepsy (Chen et al., 2003) and idiopathic generalised epilepsy (IGE) (Heron et al., 2004, Heron et al., 2007). These variants are considered susceptibility alleles as they were often inherited from an unaffected parent and do not completely segregate with the disease (Chen et al., 2003, Heron et al., 2004, Heron et al., 2007). Electrophysiological studies revealed that 9 out of 11 variants tested increased channel activity; however these variants alone were not sufficient to induce epilepsy (Heron et al., 2007).
GABRD (gamma-aminobutyric acid receptor delta) was considered a further good candidate gene for IGE as mutations in two closely related family members GABRG2 and GABRA1 had been identified in monogenic forms of IGE (Dibbens et al., 2004). Two variants of GABRD, E177A and R220H, have been identified in IGE patients and electrophysiological studies revealed that these variants decreased GABA-A receptor current amplitudes (Dibbens et al., 2004). The R220H variant is a low-frequency polymorphism that is present in affected and control individuals, suggesting it is a susceptibility allele (Dibbens et al., 2004).

There is also evidence to suggest that variants in EFHC1 (Annesi et al., 2007, Stogmann et al., 2006, Suzuki et al., 2004), ME2 (Greenberg et al., 2005), BRD2 (Pal et al., 2003), NEDD4L (Dibbens et al., 2007) and SCN9A (Singh et al., 2009) may also be susceptibility genes for complex idiopathic epilepsy.

This thesis will focus on a specific disorder involving co-morbidity of epilepsy and ID known as Epilepsy and Mental Retardation Limited to Females (EFMR [OMIM 300088]).

1.6 Epilepsy and Mental Retardation Limited to Females

Epilepsy and mental retardation limited to females (EFMR) was first recognised in 1971 in a large family in which 15 females from four generations were affected by grand mal convulsive disorder and mental retardation (Juberg and Hellman, 1971). In this kindred none of the transmitting males were affected (Juberg and Hellman, 1971).

1.6.1 Clinical description

The original EFMR family was characterised by affected individuals that appeared normal until 4-18 months, when they began to suffer from partial and generalised tonic-clonic
seizures. These seizures gradually increased in frequency and were accompanied by developmental regression (Ryan et al., 1997, Fabisiak and Erickson, 1990, Juberg and Hellman, 1971). The seizures declined dramatically by the age of 2-3 years, however cognitive development remained markedly impaired in the vast majority of patients (Ryan et al., 1997, Fabisiak and Erickson, 1990, Juberg and Hellman, 1971). A wide range of severity of ID was present within the affected females. A number of these had profound ID and required institutional care, while two out of twenty affected females in this family have grossly normal IQ and emotional function (Ryan et al., 1997). Three females with only mild ID also suffered from disabling psychiatric illness (Ryan et al., 1997).

A subsequent investigation found four additional unrelated EFMR families: two Australian and two Israeli, suggesting that this disorder may be more common than previously thought (Scheffer et al., 2008). Clinical investigation into 27 affected females from these four families extended the EFMR phenotype beyond the original description (Fabisiak and Erickson, 1990, Juberg and Hellman, 1971, Ryan et al., 1997). Affected females had a mean age of seizure onset of 14 months (range 6-36 months) typically presenting with convulsions. Seizures ceased at a mean of 12 years. A broader range of seizure types was present in the additional families. These included tonic-clonic, absence, myoclonic, partial, tonic and atonic seizures. The developmental course often varied with, for example, 4/23 females who were never normal, 7/23 females had normal development and 12/23 had a period of normal development followed by developmental delay (Scheffer et al., 2008). A wider range of psychiatric features was also present in these families including autism spectrum disorder, obsessive features, depression, panic attacks and self-injury, aggressive behaviours and schizophreniform psychosis (Scheffer et al., 2008).
Detailed clinical assessment of 58 individuals from the additional 4 EFMR families raised the possibility of a carrier male phenotype. Previously, transmitting males have been seen as unaffected as they had no seizures and were of normal intellect. Detailed assessment of transmitting males suggested the presence of obsessive, rigid and inflexible personality traits (Scheffer et al., 2008). The presence of prominent obsessive behaviours and autistic features in affected female offspring lead the authors to hypothesise that these personality traits in obligate male carriers maybe subtle markers of carrier status (Scheffer et al., 2008).

1.6.2 Inheritance pattern

Interest in EFMR centres on the unusual pattern of inheritance of the disorder, whereby females are affected but males appear to be spared, with regard to epilepsy and ID. Initially it was believed that this disorder resulted from a mutation in “a simply inherited factor with sex-limited expression” (Juberg and Hellman, 1971). However, an update on the original kindred (Fabisiak and Erickson, 1990) identified that four brothers of affected females had produced five unaffected females (as illustrated in the teal dashed boxes of figure 1.1), while four affected females had produced four affected and one unaffected female and two unaffected sons (as illustrated in the purple dashed boxes of figure 1.1). The authors concluded that the best explanation for the unusual inheritance pattern was that the trait was X-linked and sex-limited (Fabisiak and Erickson, 1990).
Figure 1.1 Pedigree of the founder EFMR family

This pedigree demonstrates the unusual inheritance pattern of EFMR. The family branches shown in the dashed boxes are those that indicate an X-linked and sex limited inheritance pattern. The teal dashed boxes illustrate the brothers of affected females who gave rise to unaffected females. The purple dashed boxes illustrate the affected females who gave rise to affected daughters and one unaffected daughter. This figure was generated from information in (Fabisiak and Erickson, 1990).
A third study of this original family was conducted after additional births and molecular confirmation of paternity (Ryan et al., 1997). This work demonstrated an X-linked pattern of inheritance for this disorder based on two features of the EFMR pedigree. Firstly there was no evidence of male to male transmission, indicative of an X-linked disorder (Ryan et al., 1997). Secondly, there was a large proportion of affected females among the daughters of transmitting males. For an autosomal locus a segregation ratio of 1:1 is expected. Here, however, the observed ratio is 9:1. This observation strongly suggested an X-linked pattern of inheritance, in which the phenotype is limited to females, in whom the penetrance is high but incomplete (as demonstrated by one asymptomatic female)(Ryan et al., 1997). Linkage of the EFMR locus was established to a 25cM interval at Xq22 flanked proximally by DXS1222 and distally by DXS6804. Linkage analysis performed on the additional 4 families was consistent with linkage to the Xq22 locus but was unable to narrow the linkage interval (Scheffer et al., 2008).

**1.6.2.1 CFNS and Ephrins**

Interestingly a similar, but not identical, pattern of inheritance to EFMR is seen in a rare X-linked disorder, Craniofrontonasal Syndrome (CFNS [OMIM 304110]). CFNS is characterised by craniofacial defects including abnormal development of the cranial and nasal bones, and craniosynostosis (premature fusion of the coronal sutures), as well as extracranial anomalies (including polydactyly and syndactyly), affecting mainly females (Grutzner and Gorlin, 1988). CFNS is caused by mutations in EphrinB1 (EFNB1) (Twigg et al., 2004, Wieland et al., 2004).

EFNB1 belongs to the ephrin and Eph family, which comprises the largest subfamily of receptor protein-tyrosine kinases. The Eph family act as receptors and bind to cell–membrane associated ephrin ligands on neighbouring cells. The Ephs are subdivided in to an A-subclass
(consisting of EphA1-EphA8 and EphA10) and a B-subclass (of EphB1-EphB4 and EphB6). These two subclasses have similar overall structure but differ in their binding affinities to either the EphrinA or EphrinB members. EphA members predominantly bind to EphrinA members. Likewise EphB binds to EphrinBs; however, crossover does occur between the classes, for example EFNA5 interacts with EphB2 (Himanen et al., 2004). Ephrins are also divided into an A and B-subclass on the basis of the presence or absence of a cytoplasmic cell anchor. EphrinBs have a transmembrane domain and a cytoplasmic tail. EphrinAs are tethered to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor.

Interactions between Eph and ephrins instigate bi-directional signalling with a “forward signal” being transduced through the Eph expressing cell and a “reverse signal” occurring through the ephrin expressing cell. Bidirectional signals resulting from Eph-ephrin interactions have emerged as a prominent form of contact dependant cell-cell communication mechanism. Eph-ephrin signalling is involved in a wide range of developmental processes, normal physiology and homeostasis of many organs (reviewed in Pasquale, 2008) through widespread effects on the actin cytoskeleton, cell-substrate adhesion, intercellular junctions, cell shape, cell movement, cell proliferation, cell survival, differentiation, and secretion (Reviewed in Pasquale, 2008). Disruption to the Eph-ephrin signalling has been associated with a wide range of disease states, including a role in immune function, glucose homeostasis and diabetes, bone maintenance and remodelling, intestinal homeostasis, and a number of cancers (breast, colorectal, skin, melanoma and tumour angiogenesis) (reviewed in Pasquale, 2008). Eph-ephrin signalling is also highly involved in correct functioning of the nervous system in particular in dendritic spine development and synaptic plasticity (reviewed in Klein, 2009).
Ephrins are expressed in a complex, spatially and temporally dynamic pattern and disruption to this pattern is believed to result in CFNS (Compagni et al., 2003, Davy et al., 2004).

### 1.6.3 Molecular analysis and mode of inheritance of EFMR

Four hypothetical models for X-linked dominant inheritance with male sparing have been suggested (Ryan et al., 1997). The first model proposes that a functional homologue of the disease locus on the Y-chromosome is able to protect males. For this to occur the X-linked EFMR locus must be silenced in the usual fashion in females, and the mutant allele must be recessive to the normal Y-chromosome paralog. Male carriers would therefore be unaffected, as each cell will contain a functional allele (the Y-chromosome allele). Females on the other hand will have approximately half of their cells expressing the mutant allele (assuming random X-inactivation), leading to the disease phenotype (Ryan et al., 1997). This hypothesis is supported but the linkage of EFMR to Xq22 as part of the linkage interval contains an XY homology region (Yoshida and Sugano, 1999).

The second hypothetical model, known as “metabolic interference”, proposes that homozygosity for either alleles (mutant or wild type) confers no phenotypic effects however when the alleles are in a heterozygous state they produce a harmful effect due to metabolic interference between the protein products of the two alleles (Ryan et al., 1997).

The third model proposes a defect in X-inactivation resulting in females having “functional dichotomy” in the genes that fail to inactivate, causing the mutant phenotype. Ryan et al. have proposed that an escape from X-inactivation by specific genes requires regional-specific-signalling and that mutations interfering with this process would result in segmental failure of X-inactivation resulting in functional dichotomy (Ryan et al., 1997).
The final model proposes that the gene defective in EFMR is not required for the development of the male brain. Alternatively the male brain maybe protected from the adverse effects of the EFMR mutation through foetal androgens (Ryan et al., 1997).

The identification of the causative mutations in CFNS has helped to shed light onto the possible molecular mechanism responsible for the sex-limited expression of this disorder. The sex-dependent manifestations of CFNS has been hypothesised to result from promiscuity of ephrin receptors in combination with a mosaic pattern of cells expressing EFNB1 as a result of X-inactivation in females (Twigg et al., 2004).

1.7 Cadherins

The mutations identified in the EFMR patients were all in the protocadherin19 (PCDH19) gene (see chapter 3). Protocadherins belong to the cadherin superfamily of cell-cell adhesion molecules that function in tissue patterning during development, growth control during tumorigenesis and in the maintenance of adult tissue architecture (Takeichi, 1991, Geiger and Ayalon, 1992, Gumbiner, 1996, Marrs and Nelson, 1996, Takeichi, 1995, Yap et al., 1997, Shapiro and Colman, 1999).

Cadherins are simple type I membrane proteins which consist of variable numbers of cadherin ectodomain (EC) repeats in their extracellular domain, a single transmembrane domain and a C-terminal cytoplasmic domain consisting of conserved sequences (Takeichi, 1990, Uemura, 1998) (figure 1.2). The extracellular domains of cadherins have calcium dependent homophilic binding specificities and preferentially bind to cells which express the same cadherin types (Nose et al., 1988). Upon calcium binding the cadherin molecule becomes rigid, capable of dimerising and is stabilised. In this calcium bound strong adhesive state,
cadherins are resistant to proteolytic degradation for extended periods of time (Takeichi, 1990, Tanaka et al., 2000). At regions of cell-cell contact cadherins assemble into large macromolecular complexes such as adherens junctions and desmosomes (Green and Gaudry, 2000, Holthofer et al., 2007, Wheelock and Johnson, 2003). The orientation and dimensions of the extracellular cadherin repeats appear optimised for bridging the synaptic cleft between two neurons (Emery and Knoblich, 2006).

**Figure 1.2 The protein structure of cadherins**

This figure illustrates the structural domains of the cadherin family of proteins.

The cadherin super-family contains more than 100 members (Morishita and Yagi, 2007) which are divided into a number of sub-groups across species. The latest phylogenetic analysis of the growing cadherin super-family and cadherin-related proteins has recommended that the cadherin family should be divided into two major branches, a cadherin and a cadherin-related branch. The cadherin major branch (CMB) further subdivided in to two sub-branches, C-1 and C-2, which together consist of eight protein families and a few individual members (figure 1.3). The cadherin-related major branch (CrMB) consists of three sub-branches, Cr-1b, Cr-2 and Cr-3 which together consist of more than 10 families (figure 1.3) (Hulpiau and van Roy, 2009). The largest cadherin sub-group is the protocadherin family
(Sano et al., 1993, Wu and Maniatis, 1999) consisting of >70 members (reviewed in Morishita and Yagi, 2007).
Figure 1.3 The cadherin super family

Phylogenetic tree of the cadherin super family on the basis of EC1 sequence homologies demonstrating the two major branches and the subfamilies within each major branch. Figure was taken from (Hulpiau and van Roy, 2009).
1.8 Protocadherins

The discovery of >70 protocadherins (reviewed in Morishita and Yagi, 2007) has more than doubled the number of known cadherins expressed in the brain (Wu and Maniatis, 1999). Protocadherins have only been identified and characterised in the last decade. The extracellular domains of protocadherins are related to that of classical cadherins, however the cytoplasmic domains show no significant similarity (Wu and Maniatis, 1999, Frank and Kemler, 2002, Suzuki, 1996, Wolverton and Lalande, 2001, Kohmura et al., 1998, Angst et al., 2001, Alagramam et al., 2001, Nollet et al., 2000, Obata et al., 1995, Sano et al., 1993). Protocadherins, like classical cadherins, are involved in high specificity, homophilic, Ca\(^{2+}\) dependent, cell-cell adhesion (Obata et al., 1995, Hirano et al., 1999, Phillips et al., 2003, Bradley et al., 1998); however, the interactions are weaker between protocadherins than between the classical cadherins (Hirano et al., 1999, Obata et al., 1995, Redies et al., 2005, Sago et al., 1995, Sano et al., 1993, Yamagata et al., 1999, Yoshida, 2003).

Interests in protocadherin research have centred around the hypothesis that the gene expression patterns of protocadherins may regulate the outgrowth of neurons along specific pathways and ultimately specify partners they should interact with (Hamada and Yagi, 2001, Redies, 2000, Frank and Kemler, 2002, Shapiro and Colman, 1999, Yagi and Takeichi, 2000). Protocadherins are believed to function in the establishment of specific neuronal connections in the brain. This is achieved through their involvement in highly specialised cell-cell interaction processes required in the central nervous system, such as network and synapse formation (Sano et al., 1993, Suzuki, 2000).

The large diversity of protocadherins both within and between the various sub-groups (the phylogenetic tree of the protocadherin family is illustrated in figure 1.4) also supports a role
in specifying neuronal connections. Variation in the extra-cellular domains of different protocadherins can lead to diversity in cell-cell interactions and subsequent signalling (Wu and Maniatis, 1999). Protocadherins within the same sub-family have identical cytoplasmic domains; (Kohmura et al., 1998, Obata et al., 1998, Wu and Maniatis, 1999) however, different protocadherin families contain similar but non-identical extracellular domains (Wu and Maniatis, 1999, Kohmura et al., 1998). Different members of the same protocadherin gene family may therefore be capable of activating a common signalling pathway; however, different families of protocadherins may activate different signalling pathways (Wu and Maniatis, 1999). Different protocadherin families are likely to have different cytoplasmic targets due to the variability of their cytoplasmic domains (Wu and Maniatis, 1999). Protocadherins have therefore been proposed to provide the adhesive diversity and molecular code for specifying neuronal connections in the brain (Hagler and Goda, 1998, Uemura, 1998, Wu and Maniatis, 1999, Kohmura et al., 1998). The finding that protocadherins are expressed in distinct patterns in synaptic junctions in the brain further supports this hypothesis (Wu and Maniatis, 1999, Kohmura et al., 1998, Sugino et al., 2000).
This figure demonstrates the subdivisions of the protocadherin family. The details shown in grey are the classification criteria used in the division of the family and sub-families based on whether members do or do not contain that specific feature (these features are discussed throughout this section). This figure was taken from (Morishita and Yagi, 2007).

The protocadherin family consists of two subfamilies; clustered (C-Pcdh) and non-clustered (NC-Pcdh) protocadherins (Morishita and Yagi, 2007) these groups are named based on the positioning of their cognate genes in the genome. The clustered protocadherin family is the largest protocadherin subfamily, consisting of more than 50 protocadherins (reviewed in Morishita and Yagi, 2007). Clustered protocadherins consist of the α-, β-, γ- families. Each of these families has a specific genomic organisation clustered in a small genome locus, as illustrated in figure 1.5 (Hirayama and Yagi, 2006). These clusters are arranged in tandem in the 5q31 region of human chromosome 5 (Wu and Maniatis, 1999, Sugino et al., 2000). The genomic arrangement of the PCDH gene cluster is similar to that of immunoglobulin or T-cell receptor gene loci (Wu and Maniatis, 1999, Wu et al., 2001). It was initially proposed that clustered $PCDH$s would be regulated by a mechanism of gene rearrangement similar to that seen in the immunoglobulin and T-cell receptor gene clusters (Wu and Maniatis, 1999, Yagi, 2006).
2003). This would result in a highly diversified expression of clustered PCDHs that could function as molecular specifiers for synaptic formation (Frank and Kemler, 2002, Suzuki, 2000). However, it has been proven that such gene rearrangement does not exist (Tasic et al., 2002, Wang et al., 2002a). Instead transcripts of the clustered potocadherins are generated from cis alternative splicing with multiple promoters, leading to the formation of a large number of isoforms with alternative extracellular domain sequences (Tasic et al., 2002, Wang et al., 2002a).
Figure 1.5 Genomic organisation and post-transcriptional processing of the clustered protocadherins

Each cluster contains the N-termini of different protocadherins (illustrated here with ovals). The PCDH-α, -β and -γ contain a number of variable N-termini. The PCDH-α and γ clusters also contain a common C-termini (illustrated in purple rectangles) encoded by three small exons (illustrated by short black lines). The bottom panel illustrates the post-transcriptional processing of the PCDH-α and γ clusters (using the PCDH- γ cluster as the example). Transcription of the PCDH-α and γ protocadherins generates a long pre-mRNA, which encodes multiple N-terminal and C-terminal regions. Cis-alternative splicing generates a transcript, which encodes a single N-terminal and C-terminal region. The N-terminal region encodes the signal peptide, extracellular domain, transmembrane domain and part of the cytoplasmic domain. The C-terminal region encodes the remainder of the cytoplasmic domain. Figure adapted from (Tasic et al., 2002).
The clustered PCDHs are controlled by a highly conserved 22bp consensus sequence element positioned upstream in the promoter region of each variable exon (Tasic et al., 2002). This commonality in control results in the various members of clustered PCDHs showing similar expression patterns across broad brain regions (Frank et al., 2005). Furthermore, individual neurons within the same brain region/neuronal population express distinct combinations of PCDH-α and -γs (Esumi et al., 2005, Frank et al., 2005, Kaneko et al., 2006). These results indicate that the clustered PCDHs are not significant players in the specification of diverse synaptic connections among various brain regions (Kim et al., 2007). There is however contention in the literature about the function of neurons in a particular brain region expressing different combinations of PCDH-α and -γs.

Alternatively it could be the non-clustered protocadherins that function as specifiers of interactions of presynaptic and postsynaptic partners in a variety of brain regions (Kim et al., 2007). The non-clustered PCDHs are distributed throughout the genome therefore each non-clustered PCDH is regulated by different upstream regions. Comparative analysis of the expression pattern of 12 non-clustered protocadherins identified a distinct brain region-specific expression pattern for each PCDH (Kim et al., 2007). The non-clustered PCDHs were differentially expressed in specific nuclei, cortical subregions and layers in the postnatal brain indicating that each brain region/specific neuronal population expresses different combinations of non-clustered PCDHs. This is consistent with a function for non-clustered PCDHs as molecular tags or synaptic specifiers for recognition of specific neuronal circuits (Aoki et al., 2003, Vanhalst et al., 2005). In addition PCDH10 expression has been identified to correlate with specific neuronal circuits including the visual and limbic neuronal circuits (Aoki et al., 2003, Hirano et al., 1999, Muller et al., 2004). PCDH7, PCDH9 and PCDH11 genes are expressed in specific brain regions, although a definitive correlation with known neuronal circuits has not yet been fully established (Vanhalst et al., 2005). PCDH21
expression is restricted to the olfactory circuits (Nagai et al., 2005, Nakajima et al., 2001). This trend appears to be broadly applicable to all non-clustered protocadherins with each member differentially expressed in different brain regions/neuronal circuits (Kim et al., 2007).

The non-clustered protocadherins can be further subdivided (figure 1.4) into Pcdhδ and the solitary protocadherins (Pcdh-12, 15, 20, 21) (Redies et al., 2005, Wolverton and Lalande, 2001). The PCDHδ sub-family in humans consists of 9 protocadherins that contain two highly conserved motifs in their cytoplasmic domain (CM1 and CM2). The PCDHδ sub-family is further divided into two subgroups PCDHδ1 and 2. This division is based on overall homology, number of EC repeats (6 vs 7) and the presence of a third conserved cytoplasmic motif (CM3) (Redies et al., 2005). PCDHδ1 consists of PCDH-1, 7, 9, 11X and 11Y and is characterised on the basis of two criteria; 1) containing seven extracellular cadherin domains, 2) containing a small but highly conserved CM3 motif (RRVTF) in addition to the CM1 and CM2 motifs. The CM3 motif has been identified to be necessary for binding to protein phosphatase-1α (Reviewed in Redies et al., 2005). The PCDHδ2 subgroup contains PCDH-8, 10, 17, 18 and 19. These protocadherins contain six extracellular cadherin domains, but the CM3 domain is absent. The CM1 domain is also more highly conserved in the δ2- than the δ1- protocadherin subgroup (Redies et al., 2005).

The high degree of conservation of the CM2 motif in δ-protocadherins suggests that the members of this sub-family may have common interactions in their cytoplasmic domains (Wolverton and Lalande, 2001). The CM2 motif contains some similarity with the laminin-type EGF-like (LE) domain, as well as the C2HC-type zinc finger-binding domain. This motif could therefore represent a functional interaction domain which could potentially mediate intracellular signalling (Wolverton and Lalande, 2001).
δ-Protocadherins are also more widely expressed throughout the body than the α, β and γ protocadherins. The clustered protocadherins are mainly expressed in the central nervous system; however, δ-protocadherins are additionally expressed in a variety of tissues such as the primitive streak, paraxial mesoderm, kidney, heart and lung. Accordingly, functional studies reveal a role in somitogenesis, axial mesoderm patterning and gastrulation (Gaitan and Bouchard, 2006).

### 1.8.1 Protocadherin functions

A decade after their discovery, the cellular role of most protocadherins has yet to be elucidated. Unlike cadherins which exhibit strong homophilic adhesion activity, protocadherins either exhibit no homophilic adhesion activity, or have been implicated in mediating weak adhesion (Frank et al., 2005, Morishita et al., 2006, Mutoh et al., 2004, Obata et al., 1995, Reiss et al., 2006, Sago et al., 1995). However it is unclear if this weak adhesion seen in some protocadherins is in fact a true homophilic cell adhesion function or if it is actually involved in other functions such as signal transduction (Chen and Gumbiner, 2006b). It has also been proposed that some of the functions associated with protocadherins may be mediated by cytoplasmic cleavage (Bonn et al., 2007, Haas et al., 2005, Hambsch et al., 2005) whereby a cytoplasmic fragment localises to the nucleus and can transactivate all protocadherin-γ promoters (Hambsch et al., 2005).

A recent investigation into γ-protocadherins has shed new light on the situation. Clustered protocadherins are expressed during neural development where they are enriched at synapses (Blank et al., 2004, Morishita et al., 2004b, Phillips et al., 2003, Wang et al., 2002b, Zou et al., 2007, Kohmura et al., 1998, Frank et al., 2005). Expression of clustered protocadherins is dramatically reduced upon myelination as neurons mature (Kallenbach et al., 2003, Morishita...
The clustered protocadherins have been proposed to mediate highly specific recognition and adhesion between neurons in synaptogenesis (Frank and Kemler, 2002, Kohmura et al., 1998, Shapiro and Colman, 1999, Wu and Maniatis, 1999). This was proposed in light of the large number of different protocadherins, their differential expression in otherwise similar neuronal populations (Wang et al., 2002b), and their similarities to classical cadherins which exhibit strong homophilic binding (Kohmura et al., 1998, Wu and Maniatis, 1999). However the biological function of PCDHs in neural cells has remained elusive, with no detectable adhesive activity (Morishita et al., 2006, Obata et al., 1995). Cytoplasmic binding partners for the vast majority of protocadherins also remain elusive.

Recently Fernandez-Monreal et al have for the first time identified homophilic interactions between Pcdh-γs (Fernandez-Monreal et al., 2009). Pcdh-γs accumulate at axonal and dendritic cell contacts in neurons specifically during development where they are involved in cell-cell interactions (Fernandez-Monreal et al., 2009). Removal of the intracellular domain of Pcdh-γs resulted in a substantial accumulation of Pcdh-γs at sites of cell-cell contact (Fernandez-Monreal et al., 2009). Previous research has predominantly focussed on the role of the extracellular domains of protocadherins in cell-cell adhesion however in the case of Pcdh-γ it appears that the cytoplasmic domain is important for correct function of Pcdh-γs in neuronal and synaptic development (Fernandez-Monreal et al., 2009).

The non-clustered protocadherins appear to function through different means. Heterophilic cell adhesion has been reported between Pcdhα4 and β1-intergrin, where β1-intergrin activation is necessary for cells to bind to Pcdhα4 (Morishita et al., 2006, Mutoh et al., 2004). PAPC the *Xenopus laevis* homolog of Pcdh-8 can mediate cell-cell adhesion through regulation of C-cadherin (Chen and Gumbiner, 2006b, Chen and Gumbiner, 2006a). This
suggests that protocadherins are involved in cell-cell adhesion but require additional interacting genes.

As knowledge of protocadherins increases so does the number of functions associated with them. This information is predominantly being obtained from investigations into disease states. Table 1.2 outlines disorders and some of the functions that protocadherins have been associated with.

| Gene       | Disorder/function                                                                 | Reference                                                      |
|------------|-----------------------------------------------------------------------------------|                                                               |
| PCDH 1     | Susceptibility for Bronchial Hyperresponsiveness                                  | (Koppelman et al., 2009)                                      |
| PCDH 1 &   | Associated with regulation of angiogenesis in vertebrate brain development        | (Krishna and Redies, 2009)                                    |
| PCDH 17    |                                                                                   |                                                               |
| PCDH 8     | Breast Cancer                                                                     | (Yu et al., 2008)                                             |
| PCDH 9     | Associated with Autism                                                             | (Marshall et al., 2008)                                      |
| PCDH 10    | Cervical cancer, gastric cancer, associated with Autism                            | (Narayan et al., 2009, Yu et al., 2009, Morrow et al., 2008)  |
| PCDH PC    | Prostate cancer                                                                    | (Giannakopoulos et al., 2007, Terry et al., 2006, Yang et al., 2005) |
| (11X/11Y)  |                                                                                   |                                                               |
| PCDH 12    | Putative association with Schizophrenia, associated with morphogenesis and transcript profile in the placenta | (Gregorio et al., 2009, Rampon et al., 2008)                  |
| PCDH 20    | Candidate tumor suppressor gene in lung cancer                                     | (Imoto et al., 2006)                                         |
| PCDH 21    | Candidate gene for human Retinal Dystrophies                                       | (Bolz et al., 2005)                                          |
| PCDH-γ A11 | Association with astrocytomas                                                      | (Waha et al., 2005)                                          |

Table 1.2 Disorders and functions associated with protocadherins

It is interesting to note that whilst protocadherins are predominantly expressed in the nervous system they have been associated with a number of disorders that occur in a number of other tissue systems including lungs, breast, stomach, prostate, placenta, ear and eyes. There is also
increasing numbers of protocadherins associated with cancerous states. It is not surprising that protocadherins are involved in cancer given their function in cell-cell adhesion.

1.9 PCDH19

PCDH19 was discovered through its sequence similarity with protocadherin-8 and because it contained the two unique conserved motifs (CM1 and CM2) which define the δ-protocadherin sub-class (Wolverton and Lalande, 2001).

Pcdh19 expression has been investigated through in situ hybridisation of mouse embryos. Pcdh19 is expressed in a tissue-specific manner during embryogenesis and has been identified in a wide range of tissues including kidney, olfactory system, eye, brain, the late stomach and in the mesenchymal compartment of the early stomach, dermomyotome, mesenteries, dermal papilla and presomitic mesoderm (Gaitan and Bouchard, 2006). The expression pattern of Pcdh19 is unique among the protocadherins documented so far, but is most similar to Pcdh8 and 10 (Gaitan and Bouchard, 2006, Kim et al., 2007).

Pcdh19 is also expressed throughout the central nervous system (Kim et al., 2007, Gaitan and Bouchard, 2006). In-situ hybridization of postnatal day 3 rat brains identified strong expression of Pcdh19 in layer 4 of the cerebral cortex, CA1-CA3 field of the hippocampus and subiculum. In the diencephalon, expression was seen in the anteroventral thalamic nucleus, suprachiasmatic nucleus and ventromedial hypothalamic nucleus. In the brain stem Pcdh19 is expressed in the anterior pretectal nucleus and the superior/inferior colliculi (Kim et al., 2007).

Two recent publications have also described Pcdh19 expression in the nervous system of zebrafish (Emond et al., 2009, Liu et al., 2009). Expression was first detected 12 hours post
fertilization (hpf) with signal detected in a segmental pattern in the anterior third of the neural keel. Strong expression seen in the presumptive forebrain and hindbrain was separated by regions of little or no Pcdh19 expression (Liu et al., 2009). By 24 hpf Pcdh19 had widespread expression throughout the brain and spinal cord with higher expression in the ventral telencephalon, dorsal and central thalamus, optic tectum, central tegmentum, cerebellum and dorsolateral regions of the hindbrain (Liu et al., 2009). Through development the expression of Pcdh19 become restricted to patches in the dorsal and/or lateral regions of the central nervous system (Emond et al., 2009, Liu et al., 2009), but no longer in the spinal cord (Liu et al., 2009). Pcdh19 expression was also detected in the eye primordium, developing retina, lens and optic vesicle (Liu et al., 2009). Interestingly, Pcdh19 expression in the nervous system, eye and ear was spatially and temporally regulated (Liu et al., 2009).

Pcdh19 has recently been identified as essential for early steps in brain morphogenesis in zebrafish (Emond et al., 2009). Disruption to Pcdh19 expression, through antisense morpholino oligonucleotides, leads to inhibition of convergent cell movements of the anterior neural plate resulting in severe disruption to early brain morphogenesis (Emond et al., 2009), with the hindbrain of morphant embryos disorganised and the midbrain-hindbrain boundary (MHB) disrupted. The anterior brain had ectopic folds of the neural tube (Emond et al., 2009). Morphants displayed a range of phenotypes which were dose dependent with increased doses of morpholino resulting in more severe phenotypes (Emond et al., 2009). The phenotype of the Pcdh19 morphants was similar to that of ncad mutants, which have a defect in convergence of the lateral neural plate (Hong and Brewster, 2006, Lele et al., 2002). Pcdh19 morphant embryos had a wider neural plate, suggestive of a defect in convergent movements (Emond et al., 2009). In addition Pcdh19 morphants have slightly shortened body axis and abnormal morphology of the eye (Emond et al., 2009). Defects resulting from disruption to Pcdh19 appear specific to future brain regions.
1.10 Project aims

Prior to the commencement of my Ph.D. our clinical collaborators has ascertained four additional families that they believed had EFMR. In collaboration with our research group they had also performed candidate gene screening on genes within the linkage interval, which they believed were good candidates, however no mutations were identified. This work has since been published (Scheffer et al., 2008). Individuals from these additional families were included in a large scale X-chromosome re-sequencing project aimed at identifying additional causative genes for XLID (Tarpey et al., 2009). \textit{PCDH19} was identified as the causative gene for EFMR. When \textit{PCDH19} was identified in EFMR there was very little information known about the precise function of this gene, therefore, the aims of my Ph.D. project were as follows:

1) To perform molecular characterisation of \textit{PCDH19} as the gene responsible for EFMR
2) To assess the likely protein consequences of the EFMR mutations
3) To investigate possible mechanisms responsible for the unusual inheritance pattern seen in EFMR families
4) To expand the spectrum of phenotypes caused by mutations in the \textit{PCDH19} gene, and
5) To perform microarray expression profiling on samples from EFMR patients in order to explore interacting pathways and formulate hypotheses for future work.
Chapter 2 – Materials and Methods

2.1 Sample Collection

2.2 Nucleic Acid Extraction

2.2.1 Genomic DNA Extraction from Blood

2.2.2 Genomic DNA Extraction from Saliva

2.2.3 Genomic DNA Extraction from Cultured Cells

2.2.4 Plasmid DNA Isolation

2.2.5 BAC Isolation

2.2.6 PCR Purification

2.2.7 RNA Extraction

2.3 DNA and RNA Quantification

2.4 First Strand cDNA Synthesis

2.4.1 Generation of cDNA from Patient Fibroblast RNA

2.4.2 Generation of cDNA from the RNA Tissue Panel

2.5 Polymerase Chain Reaction

2.5.1 Oligonucleotide Primers

2.5.2 PCR Conditions

2.5.3 PCR Product Visualisation

2.6 Real-Time Quantitative PCR (RT-qPCR)

2.6.1 RT-qPCR Primers

2.6.2 RT-qPCR Reaction

2.6.3 RT-qPCR Cycle Conditions
2.6.4 RT-qPCR ANALYSIS

2.7 SEQUENCING

2.7.1 BigDye Terminator Cycle Sequencing

2.7.2 Purification of Extension Products

2.7.3 Analysis of Sequencing

2.8 Assessment of the Conservation of Residues Mutated in EFMR

2.8.1 Conservation Analysis

2.9 Assessment of the Conservation of the Full PCDH19 Protein

2.9.1 Alignment of Full Length PCDH19

2.9.2 Assessment of the Conservation of Residues of the EC Repeat Domains of PCDH19

2.10 Alternative Splicing Analysis

2.11 Cloning

2.11.1 Preparation of Competent Cells

2.11.2 Ligation of PCR Products into pGEMT

2.11.3 Transfection of Vector DNA into Competent Cells

2.11.4 Selection of Positive Clones

2.12 Gateway® Cloning

2.12.1 Gateway Primers

2.12.2 Producing attB-PCR Products

2.12.3 BP Recombination Reaction

2.12.4 LR Recombination Reaction

2.13 Fibroblast Tissue Culture

2.13.1 Fibroblast Media

2.13.2 Establishment of Primary Fibroblast Cultures from Skin Biopsy

2.13.3 Splitting the Fibroblast Cells

2.13.4 Liquid Nitrogen Storage of Cells

2.13.5 Reviving Cells from Liquid Nitrogen

2.14 Western Blot Analysis
2.14.1 General Lysis Method for Cultured Cells 70
2.14.2 Bradford Protein Assay 70
2.14.3 Preparation of Protein Lysates for Gel Loading (Reducing Conditions) 71
2.14.4 Running the Gel (Reducing Conditions) 71
2.14.5 Western Blot Transfer 72
2.14.6 Ponceau Staining 73
2.14.7 Immunoblot 74
2.14.8 Enhanced Chemiluminescent Detection 74

2.15 Transient Transfection of PCDH19 75
2.15.1 Cell Preparation 75
2.15.2 Cell Transfection 76
2.15.3 Cell Harvest 76

2.16 Northern and Southern Blot Analysis 77
2.16.1 Prehybridization of Northern and Southern Blots 77
2.16.2 Probe Labelling Reaction 77
2.16.3 Washing of Hybridised Blots 78

2.17 Specific Methods Used in the Microarray Analysis 79
2.17.1 Fibroblast Culture for the Microarray Analysis 79
2.17.2 RNA Extraction 80
2.17.3 Data Analysis 80
2.17.4 Interpretation of the Biological Relevance of the Microarray Results Using DAVID 85
2.17.5 Interpretation of the Biological Relevance of the Microarray Results Using Ingenuity Pathways Analysis 87
2.1 Sample collection

Samples were collected with informed consent from affected and unaffected individuals as blood, saliva or skin punches.

2.2 Nucleic acid extraction

2.2.1 Genomic DNA extraction from blood

DNA was extracted from blood samples by the Molecular Diagnostic Laboratory, Division of Genetic Medicine, Women’s and Children’s Hospital, Adelaide. Genomic DNA was isolated from patient’s white blood cells using the QIAamp DNA blood maxi kit (Qiagen).

2.2.2 Genomic DNA extraction from saliva

Genomic DNA was extracted from patient’s saliva using the Oragene-DNA/saliva DNA extraction kit (DNAgenotek) by T. Williams (Epilepsy research group, Women’s and Children’s Hospital, Adelaide).

2.2.3 Genomic DNA extraction from cultured cells

DNA was extracted from tissues using phenolchloroform extraction.

2.2.3.1 Cell lysis by proteinase K treatment

Tissue samples were added to a proteinase cocktail outlined below. The 1.5 ml microcentrifuge tube lid was sealed with parafilm to prevent leakage, and rotated on a wheel overnight at 37°C.
**Proteinase cocktail:**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteinase K buffer</td>
<td>400µl</td>
</tr>
<tr>
<td>SDS (10%)</td>
<td>50µl</td>
</tr>
<tr>
<td>Proteinase K (1mg/ml)</td>
<td>50µl</td>
</tr>
</tbody>
</table>

**Proteinase K buffer:**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Concentration</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCL (pH 7.5)</td>
<td>1 M</td>
<td>10 ml</td>
</tr>
<tr>
<td>NaCl</td>
<td>5 M</td>
<td>2 ml</td>
</tr>
<tr>
<td>Na₂ EDTA (pH 8.0)</td>
<td>0.5 M</td>
<td>20 ml</td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td>968 ml</td>
</tr>
</tbody>
</table>

2.2.3.2 Phenol Chloroform Extraction

I. 500µl of a 1:1 Phenol : Water : Chloroform mix (BIORAD) was added to the digested tissue samples and mixed well by hand for 2 minutes. Samples were centrifuged at 13,000 rpm for 5 minutes.

II. The top aqueous layer of liquid was removed and mixed by hand with an additional 500µl of a 1:1 Phenol : Water : Chloroform mix (BIORAD) for 2 minutes, then centrifuged at 13,000 rpm for 5 minutes.

III. The top aqueous layer was removed and 1ml of ice-cold absolute EtOH added, samples were gently inverted several times to precipitate the DNA.

IV. Samples were centrifuged at 13,000 rpm for 10 minutes to pellet the precipitated DNA.

V. The DNA pellet was dried under vacuum then re-suspended in 100 µl of sterile distilled water.
2.2.3.3 **RNaseH treatment**

The extracted DNA was RNaseH (BioLabs) treated to remove any RNA contamination.

I. An additional 100 µl of sterile distilled water was added to the diluted DNA as well as 23µl of 10X RNaseH buffer and 1µl of RNaseH.

II. Samples were mixed and incubated at 37°C for 1 hour.

2.2.3.4 **Ethanol precipitation**

I. After the incubation, 25 µl of 3M Na-Acetate and 1 ml of ice cold absolute EtOH was added and samples were mixed to precipitate the DNA. Samples were centrifuged at 13,000 rpm for 5 minutes.

II. The resulting DNA pellet was washed with 70% EtOH then centrifuged for 10 minutes at 13,000 rpm.

III. The DNA pellet was then dried under vacuum and resuspended in 100 µl of sterile distilled water. The resulting DNA was analysed on a NanoDrop spectrophotometer (2.3 – DNA and RNA quantification) to obtain its concentration. Samples were stored at 4°C.

2.2.4 **Plasmid DNA Isolation**

The method used for plasmid DNA isolation was dependent on the amount of DNA required; extraction of plasmid DNA for validation that the plasmid contained desired inserts was performed using the Promega mini-preps DNA purification kit. However the QIAGEN plasmid midi purification kit and QIAfilters were used to generate large quantities of DNA for use in transfections. These procedures were performed according to manufacturer’s instructions as follows.
2.2.4.1 Promega Minipreps DNA purification kit

I. 10 ml bacterial cultures that had been grown overnight were centrifuged for 5 minutes at 5,000 rpm and the supernatant discarded.

II. Bacterial pellets were resuspended in 250 µl of cell re-suspension solution.

III. 250 µl of cell lysis solution was added and samples mixed by inversion, 4-6 times.

IV. 10 µl of Alkaline Protease Solution was added and samples were again mixed by inversion then left to incubate for 5 minutes at room temperature.

V. After the 5 minute incubation the lysis was inhibited through addition of 350 µl of neutralisation solution.

VI. Samples were mixed by inversion then centrifuged for 10 minutes at 13,000 rpm at room temperature.

VII. The cleared lysate was decanted into a miniprep spin column and centrifuged for 1 minute at 13,000 rpm at room temperature, to bind the plasma DNA to the spin column.

VIII. Flowthrough was discarded.

IX. The bound Plasmid DNA was washed by firstly adding 750 µl of wash solution and centrifuging the column for 1 minute at 13,000 rpm at room temperature, flow through was discarded. Followed by a second wash step with the addition of 250 µl of wash solution and centrifuging the column for 2 minute at 13,000 rpm at room temperature, flow through was discarded.

X. The bound plasmid DNA was eluted off the column through addition of 100 µl of nuclease-free water to the column followed by centrifugation of the column for 1 minute at 13,000 rpm at room temperature.

XI. The resulting plasmid DNA was analysed on a NanoDrop spectrophotometer to obtain the concentration, then stored at 4 °C.
2.2.4.2 QIAGEN Plasmid Midi Kit

I. 200 ml bacterial cultures that had been grown overnight under antibiotic selection were harvested by centrifugation at 5,000 rpm for 15 minutes at 4°C.

II. The supernatant was discarded and the bacterial pellets resuspended in 8 ml of buffer P1.

III. The bacterial pellets were resuspended through pipetting.

IV. Bacterial lysis was performed by addition of 8 ml of P2 buffer, samples were mixed by inverting the samples 4-6 times, samples were then incubated for 5 minutes at room temperature.

V. The lysis was terminated through addition of 8 ml P3 buffer and mixed immediately by inversion.

VI. Lysate was poured into the barrel of a QIAfilter cartridge and allowed to settle for 10 minutes at room temperature.

VII. During the incubation a QIAGEN-tip 100 was equilibrated by allowing 4 ml of QBT buffer to flow through the QIAGEN-tip by gravity flow.

VIII. The settled lysate was forced by a plunger through the QIAfilter cartridge and added into the QIAGEN-tip.

IX. Plasmid DNA was bound to the QIAGEN-tip by gravity flow of the lysate through the QIAGEN-tip.

X. The bound DNA was washed by two successive additions of 10 ml buffer QC which was allowed to flow through the QIAGEN-tip by gravity flow.

XI. Bound DNA was then eluted by the addition of 5 ml buffer QF.

XII. DNA was precipitated by addition of 3.5 ml of room temperature isopropanol, mixed and centrifuged immediately at 13,000 rpm for 30 minutes at 4°C.
XIII. The supernatant was then removed and 2 ml of room temperature 70% ethanol added to the DNA pellet to wash the DNA. Samples were then centrifuged at 13,000 rpm for 10 minutes at room temperature.

XIV. The supernatant was decanted and the DNA pellet air-dried under vacuum centrifugation.

XV. DNA was redissolved in 1 ml of sterile water, the resulting plasmid DNA was analysed on a NanoDrop spectrophotometer to obtain the concentration then the DNA was stored at -20 °C.

### 2.2.5 BAC isolation

The mouse BAC (Bacterial artificial chromosome) clone RP23-30I16 that contains PCDH19 was purchased from Australian Genome Research Facility (AGRF). DNA isolation was performed using QIAGEN® Plasmid Midi Kit and an adapted protocol outlined below:

I. Upon receipt of the BAC it was streaked out on a chloromphenacol agar plate. A single colony was then inoculated into 200 ml of LB media (see 2.11.1 – Preparation of competent cells for the recipe) and grown overnight at 37°C under chloramphenicol selection with vigorous shaking.

II. The following day the bacteria were harvested by centrifugation at 3,000 rpm for 20 minutes at 4°C.

III. The bacteria pellet was resuspended in 10 ml of buffer P1.

IV. 10 ml of buffer P2 was added to each sample and mixed thoroughly but gently through inverting 4-6 times. Samples were incubated at room temperature for 5 minutes.

V. 10 ml of buffer P3 was added to each sample and immediately mixed by inversion 4-6 times then incubated on ice for 15 minutes.
VI. Samples were centrifuged at 18,000 rpm for 30 minutes at 4°C to remove the cell debris.

VII. The supernatant containing the BAC was removed and re-centrifuged at 18,000 rpm for 30 minutes at 4°C.

VIII. A QIAGEN-tip 100 was equilibrated by applying 4 ml of buffer QBT and allowing it to flow through the column by gravity flow.

IX. The supernatant containing the BAC DNA was added to the QIAGEN-tip 100 and allowed to enter the column by gravity flow.

X. The QIAGEN-tip 100 was washed with 10 ml of buffer QC, which was allowed to flow through the column by gravity flow.

XI. The column was then re-washed with an additional 10 ml of buffer QC.

XII. The DNA was eluted with 5 mls of buffer QC, which had been pre-warmed at 65°C.

XIII. The eluted DNA was precipitated by adding 3.5 ml of room temperature isoproponol and mixed by vortexing then centrifuged at 13,000 rpm for 30 minutes at 4°C.

XIV. The resulting DNA pellet was washed with 1 ml of room temperature 70% ethanol and centrifuged at 13,000 rpm for 10 minutes and the supernatant removed.

XV. The DNA pellet was dried under vacuum and resuspended in ~100 μl of sterile distilled water.

XVI. The isolated BAC DNA was analysed on a NanoDrop spectrophotometer to obtain the concentration, 100-200 ng of BAC DNA was also run on a 1% agarose gel to ensure the quality of the BAC DNA and to ensure it has not been sheared in the extraction process.

XVII. BAC DNA was stored at 4°C.
2.2.6 PCR purification

2.2.6.1 QIAquick PCR Purification Kit

Purification of PCR products was performed using the QIAquick PCR Purification Kit (Qiagen) as per the manufacturer’s instructions, briefly:

Following agarose gel (see 2.5.3 – PCR product visualisation) confirmation that a PCR (see 2.5 – Polymerase chain reaction) had been successful and a single band of the desired size had been obtained the remaining PCR product was purified as follows:

I. 5 volumes of buffer PB were mixed with 1 volume of PCR product.

II. The resulting mixture was added to a QIAquick column in a collection tube and centrifuged for 60 seconds at 13,000 rpm.

III. The flow through was discarded and the column washed by adding 750 µl of buffer PE and centrifuged for 60 seconds.

IV. Flow through was discarded and the QIAquick column was washed again with 250 µl of buffer PE and centrifuged for 60 seconds at 13,000 rpm.

V. Flow through was discarded and the QIAquick column was centrifuged for an additional minute at 13,000 rpm to remove any remaining buffer.

VI. The DNA was eluted by adding 30 µl of elution buffer directly to the QIAquick column and left to stand at room temperature for 1 minute then centrifuged for 60 seconds at 13,000 rpm.

VII. The resulting DNA concentration was measured with a NanoDrop, then the DNA was stored at 4°C.
2.2.6.2 QIAquick Gel Extraction Kit

DNA was extracted from agarose gels using the QIAquick Gel Extraction Kit Protocol from QIAGEN as per the manufacturer’s instructions. In brief:

I. DNA was excised from a 1% gel using a scale and the fragment weighed.

II. 3 volumes of buffer QG was added to 1 volume of gel (100 mg ~ 100 μl) and incubated at 50°C for 10 minutes (or until the gel slice has completely dissolved).

III. 1 gel volume of isoproponol was added and the samples vortexed.

IV. The resulting sample was added to a QIAquick spin column and centrifuged for 1 minute at 13,000 rpm.

V. The supernatant was discarded.

VI. The column washed with 0.5 ml of buffer QG, centrifuged for 1 minute at 13,000 rpm.

VII. The supernatant was discarded.

VIII. The column was re-washed with 0.75 ml of buffer PE.

IX. Supernatant was discarded and the column centrifuged for an additional minute at 13,000 rpm.

X. DNA was eluted by adding 30 μl of elution buffer to the centre of the QIAquick column then left to sit at room temperature for 1 minute; the column was then centrifuged for 1 minute at 13,000 rpm.

XI. The resulting DNA was run on a NanoDrop to obtain the concentration, then stored at 4°C.
2.2.7 RNA extraction

RNA extractions were performed using Trizol (Invitrogen) and QIAGEN RNA isolation columns as per the manufacturer’s instructions.

I. Frozen cell pellets were resuspended in 1 ml of Trizol. The cells were homogenised through agitation by pipetting. Samples were incubated at room temperature for 5 minutes.

II. 200 μl of Chloroform (MERCK) was added and the samples shaken vigorously by hand for 30 seconds then incubated at room temperature for 3 minutes.

III. Samples were centrifuged for 15 minutes, 4°C, at 12,000 g.

IV. The upper aqueous phase containing the RNA to a fresh tube and 600 μl of 70% EtOH added and mixed well.

V. 700 μl of the resulting mixture was added to a RNeasy mini column (Qiagen), then centrifuged for 30 seconds at 13,000 rpm.

VI. The flow through was discarded and the column was subjected to DNase digestion using the RNase-Free DNase set (Qiagen).

VII. This was performed by adding 350 μl of buffer RW1 to the column and centrifuging it for 30 seconds.

VIII. 10 μl of DNase I stock solution was gently mixed with 70 μl of buffer RDD and the resulting 80 μl added to the RNeasy mini column and incubated at room temperature for 15 minutes.

IX. An additional 350 μl of buffer RW1 was added into the column and the column was centrifuged for 30 seconds at 13,000 rpm, the supernatant was discarded.

X. The RNeasy column was transferred to a new 2 ml collection tube and 500 μl of buffer RPE added to the column and centrifuged at 13,000 rpm for 30 seconds.
XI. An additional 500 μl of buffer RPE was added to the column and centrifuged for 2 minutes and the RNA eluted by adding 30 μl of RNase-free water to the column membrane and centrifuging for 30 seconds at 13,000 rpm.

XII. The resulting RNA was analysed on a NanoDrop spectrophotometer to obtain the concentration, 100-200ng of RNA was also run on a 1% agarose gel (Scientifix) to ensure the quality of the RNA.

2.3 DNA and RNA quantification

DNA and RNA concentrations were determined by measuring the optical density on a NanoDrop spectrophotometer ND-1000, running software version V3.3.0, as per the manufacturer’s instructions.

I. The NanoDrop pedestal and arm were cleaned with MQ-water.

II. 2 μl of MQ-water were used to initialise the instrument, this was then removed using a KimWipe.

III. The instrument was set to Nucleic acid or RNA depending on the sample being assessed.

IV. 2 μl of the appropriate elution buffer or MQ-water were used to ‘blank’ the instrument, this was then removed.

V. 2 μl of undiluted sample were used for quantification of the 260/280 nm readings and the DNA or RNA concentration.

VI. After all samples had been assessed the solution used to ‘blank’ the machine was re-assessed to ensure it gave a zero reading.
2.4 First strand cDNA synthesis

First strand cDNA synthesis was performed using SUPERSCRIPT™ III Reverse Transcriptase (Invitrogen) as per the manufacturer’s instructions. In brief:

2.4.1 Generation of cDNA from patient fibroblast RNA

I. In a 0.5 ml PCR tube 2 mg of total RNA was mixed with 1 μl of random hexamers (Geneworks), 1 μl of 10 mM dNTP mix (Invitrogen) and sterile distilled water to 13 μl.

II. This mixture was heated at 65 °C for 5 minutes then chilled on ice for at least 1 minute.

III. 4 μl of 5X First-Strand buffer, 1 μl of 0.1 m DTT was added to all reactions.

IV. 1 μl of Superscript™ III RT at 200 units/μl was added to the mixture (excluding negative controls where 1 μl of sterile distilled water was added); all reactions were mixed through pipetting.

V. Samples were incubated at 25 °C for 5 minutes, then at 50 °C for 1 hour.

VI. The reaction was terminated through incubating samples at 70 °C for 15 minutes.

VII. Following first strand cDNA synthesis 30 μl of sterile distilled water was added to all samples.

VIII. 1 μl of the cDNA mixture generated was generally used as a template for PCR using standard conditions (see 2.5 – Polymerase chain reaction).

IX. All cDNA samples generated were tested with a control PCR reaction using oligonucleotides to the EsteraseD gene, to confirm the first strand cDNA synthesis had been successful and that there was no gDNA contamination in the samples (this was assessed through a negative PCR result for the minus Superscript™ III RT reactions).

X. Samples were stored at -20°C.
The EsteraseD oligonucleotides used were:

Forward 5’ – GGA GCT TCC CCA ACT CAT AAA TGC C – 3’
Reverse 5’ – GCA TGA TGT CTG ATG TGG TCA GTA A – 3’

2.4.2 Generation of cDNA from the RNA tissue panel

RNA from a range of human tissues including specific brain regions were purchased from Clontech. 1mg of each of these were used to generate cDNA as per the procedure described above.

2.5 Polymerase chain reaction

2.5.1 Oligonucleotide Primers

Oligonucleotides for PCR were designed such that they contained close to 50% GC content and a melting temperature around 60°C. Oligonucleotides were usually 18-25 base pairs in length. The primers used for sequencing PCDH19 are outlined in the table below.
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’to 3’)</th>
<th>Size (bp)</th>
<th>Product (bp)</th>
<th>Annealing Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCDH19ex1-F1</td>
<td>TCGGAGGGGTGTGGAGAGG</td>
<td>19</td>
<td>428</td>
<td>64</td>
</tr>
<tr>
<td>PCDH19ex1-R1</td>
<td>CTGAACGACAATGCGGCCCA</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex1-F2</td>
<td>ATTAGCTGTACCTGCTGTGCA</td>
<td>21</td>
<td>507</td>
<td>64</td>
</tr>
<tr>
<td>PCDH19ex1-R2</td>
<td>AAAACTGGGCTCCCCAACC</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex1-F3</td>
<td>TGACCGACTCACAATGCAAACACCA</td>
<td>24</td>
<td>551</td>
<td>62</td>
</tr>
<tr>
<td>PCDH19ex1-R3</td>
<td>ACTATTCTGTGGGACGGAGCCTG</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex1-F4</td>
<td>GCCTTGGTCGGGTGCTGTGAT</td>
<td>21</td>
<td>480</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex1-R4</td>
<td>TAACCAGGAGGAGACCAAGGCG</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex1-F5</td>
<td>TACCAAGGCTTTTGTGCAAGAAGA</td>
<td>24</td>
<td>629</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex1-R5</td>
<td>AATCTACTTTGCTCCCTGCTCTCA</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex1-F6</td>
<td>CAGGTAATGGGCGAAGTCAGA</td>
<td>21</td>
<td>383</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex1-R6</td>
<td>AATCTCTGTGCTTTTTGTTCCCTAC</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex2-F</td>
<td>CTGAAATGGAAGGATGCTGA</td>
<td>22</td>
<td>352</td>
<td>62</td>
</tr>
<tr>
<td>PCDH19ex2-R</td>
<td>GTTTTGGACTGGGTAGGAG</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex3-F</td>
<td>GGAGAAAACCCGATGATGTAATG</td>
<td>22</td>
<td>529</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex3-R</td>
<td>GCGAGCACCTAAAGAGGAGA</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex4-F</td>
<td>CTCTTAAGTAGGAGCAGTTC</td>
<td>19</td>
<td>317</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex4-R</td>
<td>ATTACCAGTGTAGTTATTTAC</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex5-F</td>
<td>TCATTACAGAGGCGACATCATC</td>
<td>22</td>
<td>296</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex5-R</td>
<td>ACATTTCGGTTCTTTGGAGT</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex6-F1</td>
<td>TGGGGAGTAAACAGTGAATC</td>
<td>19</td>
<td>476</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex6-R1</td>
<td>CAGGGCAATGGTGTAAGAC</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCDH19ex6-F2</td>
<td>CTGAAAGGCAAGAGGACTG</td>
<td>19</td>
<td>429</td>
<td>60</td>
</tr>
<tr>
<td>PCDH19ex6-R2</td>
<td>TGGTGAGCAATAAAAACAAGA</td>
<td>21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**PCDH19** specific primers used for PCR amplification and sequencing.

The remaining primer sequences used are given throughout the text in the sections for which they are relevant to.

### 2.5.2 PCR conditions

PCRs were amplified in 20 µl or 50 µl reaction volumes using the 10X buffer supplied (100 mM Tris-HCl, 15 mM MgCl$_2$, 500 mM KCl (pH8.3 [20°C])) with Taq DNA Polymerase (Roche) as per the manufacturer’s instructions. Each PCR reaction contained the following.
<table>
<thead>
<tr>
<th>Reagent</th>
<th>Per 50 μl Vol</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA (50 ng/μl)</td>
<td>1 μl</td>
<td>0.5 ng/μl</td>
</tr>
<tr>
<td>10X reaction buffer</td>
<td>5 μl</td>
<td>1X</td>
</tr>
<tr>
<td>10 mM dNTP mix</td>
<td>1 μl</td>
<td>200 μM</td>
</tr>
<tr>
<td>Forward primer (10μM/μl)</td>
<td>1 μl</td>
<td>0.2 μM</td>
</tr>
<tr>
<td>Reverse primer (10μM/μl)</td>
<td>1 μl</td>
<td>0.2 μM</td>
</tr>
<tr>
<td>Taq polymerase (5 U/μl)</td>
<td>0.02 μl</td>
<td>0.4 U/nl</td>
</tr>
<tr>
<td>Sterile distilled water</td>
<td>To 50 μl</td>
<td></td>
</tr>
</tbody>
</table>

Standard PCR reactions were carried out using the cycling conditions as shown below. This was altered if the expected PCR product sizes were greater than 500 bp, by adding 30 seconds to the extension time per 500 bp. The annealing temperature was also adjusted depending on different primer sets (typically between 55-65°C).

94°C  5 minutes
94°C  30 sec
60°C  30 sec  x35 Cycles
72°C  30 sec
72°C  5 minutes

**2.5.3 PCR product visualisation**

PCR products were run on an agarose gel with ethidium bromide added for visualisation of the DNA. Typically 1.5% agarose (Scientifix) gels were used, 2 or 1% gels were also used for visualisation where required. 1kb+ DNA ladder (Invitrogen) was used as a size standard in agarose gels. Gels were photographed using an INGENIUS syngene Bio imaging capture system and GeneSnap image acquisition software version 7.05.01.
2.6 Real time quantitative PCR (RT-qPCR)

RT-qPCR was performed using the standard curve method and standardised to the control genes Actin-β. All samples and the standard curve samples were run in triplicate. The cDNA used to generate the standard curve was comprised of cDNA from all samples being assessed in a RT-qPCR experiment. A five dilution standard curve was performed in every real-time assay run. In general the dilution series used contained dilutions of: 1/10, 1/50, 1/100, 1/500 and 1/1000. RT-qPCR was performed using a StepOne Real-Time PCR System (Applied Biosystems).

2.6.1 RT-qPCR primers

The primers used to assess the expression of PDGFC, NTF3, ITGA3, ITGA6, MET, LAMB1, GAS7 and EFNA5 were purchased from PrimerDesign (www.primerdesign.co.uk/). These primer sequences are copyrighted and remain the intellectual property of PrimerDesign 2007 Ltd and are not allowed to be re-synthesised.

EFNA5 Forward 5’ – CCTCCTGGCGATGCTTTTG – 3’
EFNA5 Reverse 5’ – AGTTAGGTGGATCTCTGGTGTT – 3’

LAMB1 Forward 5’ – GAGAGGAAAGTCAGCGAGATAA – 3’
LAMB1 Reverse 5’ – GAGCCATCATTTCTGTAACATTT – 3’

PDGFC Forward 5’ – GGGAAGAACTAAAGAGAACGATA – 3’
PDGFC Reverse 5’ – TGCTTGGGACACATTGACATT – 3’

NTF3 Forward 5’ – AGGTAACAACATGGATCAAAGGA – 3’
NTF3 Reverse 5’ – AGGGTGCTCTGTAATTTTCC – 3’

ITGA6 Forward 5’ – ACAAGGATGGTGCAAGA – 3’
ITGA6 Reverse 5’ – TAAGACGAATTGGCTTCACATTATT – 3’

ITGA3 Forward 5’ – GATAACAGGAGATGCCACTTC – 3’
ITGA3 Reverse 5’ – CCCTGCAATATTCATTAATTCAAGT – 3’

MET Forward 5’ – AAAAGTTTACCACACAGTCAGATG – 3’
MET Reverse 5’ – AGTATTCGGGTTGTAGGAGTCT – 3’

GAS7 Forward 5’ – CGCCAGCCCATCCTTAGTA – 3’
GAS7 Reverse 5’ – ATGTGCTCGTGCTGGAAG – 3’

The primers used to assess expression of EPHA2, CDON, CXCR7 and mouse Pcdh19 were ordered from primer sequences deposited in the PrimerBank (www.pga.mgh.harvard.edu/primerbank) a public database of PCR primers for use in gene expression detection.

Primer sequences used:

EPHA2 Forward 5’ – AGAGGCTGAGCGTATCTTCAT – 3’
EPHA2 Reverse 5’ – GGTCCGACTCGGCATAGTAGA – 3’

CDON Forward 5’ – AAAATCCGGGAAAAATGGCTG – 3’
CDON Reverse 5’ – GAGCTTTCCGCAAATAGGTTC – 3’
CXCR7 Forward 5’ – CCAGTCTGGGTGGTCAGTCT – 3’
CXCR7 Reverse 5’ – CTCATGCACGTGAGGAAGAA – 3’

mPcdh19 Forward 5’ – TTGGCcTGGAAATAAAGACGC – 3’
mPcdh19 Reverse 5’ – GCGAAAGCTGTAATGTGACTGTG – 3’

The primers used to assess expression of mouse Efna5 and Epha2 were ordered from primer sequences deposited in the PrimerDepot (www.mouseprimerdepot.nci.nih.gov/) a quantitative real time PCR primer database by the National Institutes of Health.

Primer sequences used:

mEfna5 Forward 5’ – GGGCAGAAAACATCCAGGTA – 3’
mEfna5 Reverse 5’ – GGATGTGTGTGTTCCAGCCAG – 3’

mEpha2 Forward 5’ – CTCTGTGGATGGCAGGGTAT – 3’
mEpha2 Reverse 5’ – AGGCTACGAGAAGGTCGAGG – 3’

The primers used to assess expression of Actin-β, PCDH11X and PCDH11Y were designed in our laboratory.

Primer sequences used

Actin-β Forward 5’ – ATGGGTCAGAAGGATTCCTATGTG – 3’
Actin-β Reverse 5’ – TGTTGAAGGTCTCAAACATGATCTGG – 3’

PCDH11X Forward 5’ – GTAACAAAGTGCTACCTGGATGACT – 3’
PCDH11X Reverse 5’ – TCAACCTTTACTTTTCATCAG – 3’
PCDH11Y Forward 5’ – TACAACAAACTGTCACAAGTGTTT – 3’

PCDH11Y Reverse 5’ – TCAACCTTTACTTTCATCACA – 3’

2.6.2 RT-qPCR reaction

RT-qPCR reactions were performed using the iTaq SYBR Green super mix with ROX (BioRad).

RT-qPCR recipe

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Per 10 μl Vol</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA (1/20 dilution)</td>
<td>2 μl</td>
<td>0.5 ng/μl</td>
</tr>
<tr>
<td>2X SYBR master mix</td>
<td>5 μl</td>
<td>1X</td>
</tr>
<tr>
<td>Forward primer (10μM/μl)</td>
<td>1 μl</td>
<td>1 μM</td>
</tr>
<tr>
<td>Reverse primer (10μM/μl)</td>
<td>1 μl</td>
<td>1 μM</td>
</tr>
<tr>
<td>Sterile distilled water</td>
<td>To 10 μl</td>
<td></td>
</tr>
</tbody>
</table>

2.6.3 RT-qPCR cycle conditions

95°C  10 minutes

9°C   15 sec  } x40 Cycles

60°C  1 minute

Melt curve step

95°C  15 sec

60°C  1 minute

95°C  15 sec

Data for the melt curve was successively collected as the temperature increased from 60°C to 95°C.
2.6.4 RT-qPCR analysis

Analysis of the RT-qPCR results was performed using the StepOne Software V2.0 (Applied Biosystem). Here single outliers were removed from within the three replicates run for each sample. A real-time run was repeated if there was too much variation between triplicates as flagged by the software program. Real-time plates were also repeated if the $R^2$ of the standard curve for either the control gene ($Actin-ß$) or the particular gene being assessed was less than 0.98. Repeats were also performed if the efficiency (as calculated by the analysis software from standard curve results) was less than 90% or above 110%.

The data was then exported to Excel for further assessment. In Excel

I. The quantity mean and the quantity SD (standard deviation) for each fibroblast cell line being analysed was extracted for both the control gene ($Actin-B$) and the gene being assessed.

II. The quantity mean and quantity SD for each fibroblast cell line being analysed for a particular gene was normalised against the control gene ($Actin-B$). This was performed by dividing both the quantity mean and quantity SD for each fibroblast cell line by the quantity mean of $Actin-B$ obtained for that fibroblast cell line.

III. The normalised quantity mean value for each individual fibroblast cell line was then used to calculate the mean expression for all of the individuals in a group, including affected females, control females, transmitting males and control males.

IV. The standard error for each group was also calculated in Excel using the normalised quantity mean values.

V. P-values of the difference between the normalised quality mean values for the individuals in each group was calculated. The Excel formula “TTEST” was used to calculate the P-value. The normalised quality mean value for each individual in a group being assessed was used as “array1”, for example, all six affected EFMR females. The
normalised quality mean value for each individual in the group being compared to were entered as “array2”, for example, all three control females. Two-tailed distribution was selected and the type selected was three, which correlated to two-sample unequal variance.

2.7 Sequencing

Sequencing samples were electroporated by the ABI automated DNA sequencers at the IMVS Molecular Pathology Unit (Adelaide, Australia) as part of a paid service.

2.7.1 BigDye Terminator Cycle Sequencing

PCR products and cloned DNA were sequenced using the BigDye terminator cycle sequencing kit as per the suppliers (Perkin Elmer) instructions. PCR products were cleaned using the QIAquick PCR purification kit (see 2.2.6 – PCR purification). Plasmid DNA was isolated using “Promega Mini prep” (see 2.2.4 – Plasmid DNA isolation).

**Sequencing reaction**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Plasmid sequencing</th>
<th>PCR product sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid DNA</td>
<td>250 ng</td>
<td>-</td>
</tr>
<tr>
<td>Purified PCR product</td>
<td>-</td>
<td>100 ng</td>
</tr>
<tr>
<td>Primer</td>
<td>10 μM</td>
<td>10 μM</td>
</tr>
<tr>
<td>ABI BigDye™</td>
<td>1 μl</td>
<td>1 μl</td>
</tr>
<tr>
<td>BigDye Buffer V3-1</td>
<td>3 μl</td>
<td>3 μl</td>
</tr>
<tr>
<td>Sterile distilled water</td>
<td>To 20 μl</td>
<td>To 20 μl</td>
</tr>
</tbody>
</table>
Sequencing conditions

96°C  30 seconds
50°C  15 seconds
60°C  4 minutes

\( \text{x25 Cycles} \)

2.7.2 Purification of Extension Products

Sequencing reactions were cleaned up using isopropanol as follows.

I. 20 \( \mu \text{l} \) of sterile distilled water was added to the sequencing samples followed by 60 \( \mu \text{l} \) of isopropanol (Ajax Finechem Pty Ltd), samples were mixed by vortexing and incubated at room temperature for at least 15 minutes.

II. The extension products were then pelleted by centrifugation at 13,000 rpm for 20 minutes.

III. Supernatant was removed and the pellet washed with 62.5 \( \mu \text{l} \) of sterile distilled water and 187.5 \( \mu \text{l} \) of isopropanol, samples were mixed by vortexing, and centrifuged at 13,000 rpm for 5 minutes.

IV. The supernatant was removed and the pellet air-dried.

V. Samples were then sent to IMVS for sequencing.

2.7.3 Analysis of sequencing

Sequence analysis was performed using the SeqMan Pro program from the Lasergene DNA Star software package version 8.0.2.

I. Sequence products were aligned to their respective target sequences then analysed for discrepancies by eye.
II. Any discrepancies identified were cross-referenced to the known SNPs in the HapMap data set and our own sequence findings.

III. Nucleotide changes that were not known SNPs were interrogated further.

IV. To determine the predicted protein consequence of mutations identified, nucleotide changes identified were introduced into the full length \textit{PCDH19} sequence (GeneBank accession number EF676096) and translated into protein using the EditSeq program from the Lasergene DNA Star software package version 8.0.2.

V. The translated protein sequence was then compared to the reference protein sequence using the MegAlign program from the Lasergene DNA Star software package version 8.0.2.

2.8 Assessment of the conservation of residues mutated in EFMR

Once a novel nucleotide change in \textit{PCDH19} that resulted in amino acid changes had been identified the conservation of the affected residue was assessed.

2.8.1 Conservation analysis

I. The protein sequence flanking the mutated residue (~ 30 aas in total) were subjected to a BLAST-P search (http://blast.ncbi.nlm.nih.gov/Blast).

II. Matching sequences were extracted from the NCBI database and an EditSeq (protein) file was created for each one.

III. Sequences were then aligned using MegAlign typically using the Clustal-W option.

The likely consequence to protein function of any such changes were then assessed through alignment, using MegAlign, of the affected EC repeat domain with EC domain 1 from Pcdhα4, in which the important residues involved in calcium binding and formation of the hydrophobic core had been identified.
2.9 Assessment of the conservation of the full PCDH19 protein

2.9.1 Alignment of full length PCDH19

I. Protein sequences of all of the δ-protocadherins were extracted from NCBI and aligned to PCDH19 using MegAlign.

II. This alignment was uploaded to the Scorecons Internet based server (www.ebi.ac.uk) under standard settings.

III. The output from this program was imported into Excel and used to graph the conservation score of each amino acid in PCDH19.

2.9.2 Assessment of the conservation of residues of the EC repeat domains of PCDH19

I. A Clustal-W alignment was generated to align all of the six EC repeat domains of PCDH19.

II. Based on the protein alignment the conservation score for each amino acid in the six EC repeat domains were also aligned together.

III. The mean conservation score of the conservation score for all amino acids present at each specific position of the alignment was then calculated.

IV. Pcdhα4 and EC1 sequence was then added into the Clustal-W protein alignment in order to correlate the calcium binding residues and hydrophobic residues of Pcdhα4 with PCDH19. In order to identify likely important residues of PCDH19.
2.10 Alternative splicing analysis

To determine if any alternative splicing occurs in PCDH19 we downloaded the EST’s associated with PCDH19 and aligned using the SeqMan Pro program from the Lasergene DNA Star software package version 8.0.2. Primers were then designed which flanked exon 2 with the forward in exon 1 and the reverse in exon 3.

Primers used

PCDH19 Exon 1 F  5’- TCACGACCACGGCAAGACATC -3’
PCDH19 Exon 3 R  5’- GCTGAGTACTCCTATGGGCAT -3’

These primers were used to PCR amplify across exon 2 such that two different sized products (differing in size by 141 bp) would be produced depending on if exon 2 was expressed or not. PCR products were assessed by agarose gel electrophoresis (see 2.5.3 – PCR product visualisation). The PCR products were also sequenced using the PCDH19 Exon 1 F and PCDH19 Exon 3 R primers to verify the two products obtained were what the EST sequences had indicated.

2.11 Cloning

Cloning was used to generate the PCDH19 and PCDH11X/11Y northern blot probe and the mouse Pcdh19 in-situ probe.

2.11.1 Preparation of Competent Cells

I. A single colony of the required cells was obtained by streaking out a portion of the glycerol stock of the required cell type, which was grown overnight on a LB agar plate with the appropriate antibiotic resistance.
II. Once a single colony was inoculated into a 5 ml of LB broth with the appropriate antibiotic for the cell type and grown at 37°C with agitation overnight.

III. The following day, cells were diluted 1/200 into fresh media (1 ml of overnight culture into 200 ml of LB Broth) and incubated at 37°C with agitation until the OD$_{600}$ was near 0.3 (~2.5 hours).

IV. The OD was measured using a spectrophotometer (Pharmacia Biotech UltraSpec 3000) at a wavelength of 260 nm, using LB as a blank.

V. Cells were harvested by centrifugation at 2,250 rpm for 15 minutes at 4°C, and the supernatant discarded.

VI. Cells were resuspended in 2 times the pellet volume (~2-3 ml) of ice cold TSS (see recipe below).

VII. Cells were incubated on ice for 30 minutes then distributed into 200 µl aliquots and snap frozen in liquid nitrogen after which the samples were stored at -80°C.

<table>
<thead>
<tr>
<th>TSS</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyethylene glycol</td>
<td>2g</td>
</tr>
<tr>
<td>Dimethyl sulfoxide</td>
<td>1ml</td>
</tr>
<tr>
<td>1M MgCl$_2$</td>
<td>1ml</td>
</tr>
<tr>
<td>L-Broth</td>
<td>18ml</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L-Broth</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone-Bacto</td>
<td>10g</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5g</td>
</tr>
<tr>
<td>NaCl</td>
<td>10g</td>
</tr>
<tr>
<td>Water to</td>
<td>1L</td>
</tr>
</tbody>
</table>
2.11.2  Ligation of PCR products into pGEMT

Ligations were performed using the pGEM®-T Vector System (Promega) as follows:

**Reaction:**

<table>
<thead>
<tr>
<th></th>
<th>Standard reaction</th>
<th>Positive control</th>
<th>Background control</th>
</tr>
</thead>
<tbody>
<tr>
<td>2X Rapid Ligation Buffer</td>
<td>5µl</td>
<td>5µl</td>
<td>5µl</td>
</tr>
<tr>
<td>ATP (10µM)</td>
<td>1µl</td>
<td>1µl</td>
<td>1µl</td>
</tr>
<tr>
<td>pGEM®-T easy vector (50ng)</td>
<td>1µl</td>
<td>1µl</td>
<td>1µl</td>
</tr>
<tr>
<td>PCR product</td>
<td>1µl</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Control insert DNA</td>
<td>-</td>
<td>2µl</td>
<td>-</td>
</tr>
<tr>
<td>T4 DNA ligase</td>
<td>1µl</td>
<td>1µl</td>
<td>1µl</td>
</tr>
<tr>
<td>Sterile distilled water to final volume of</td>
<td>10µl</td>
<td>10µl</td>
<td>10µl</td>
</tr>
</tbody>
</table>

I. Reactions were set up as described and mixed by pipetting.

II. Reactions were incubated at 4°C overnight.

III. Positive controls were set up to assess if the ligation reaction had been successful, the background control was used to assess for contamination.

IV. Ligation reactions were subsequently transfected into the required competent cells.

2.11.3  Transfection of vector DNA into Competent cells

I. 2 µl of the ligation mixture was added to 50 µl of prepared competent cells, mixed gently, and incubated on ice for 30 minutes.

II. Samples were heated in a 42°C water bath for 45 seconds, then immediately placed in an ice slurry for 2 minutes.

III. 450 µl of SOC media (see below for recipe) was added and samples placed at 37°C with vigorous shaking for 45 minutes.

IV. 100-300 µl was plated out onto LB agar plates containing the appropriate antibiotic selection.
<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto Tryptone</td>
<td>20g</td>
</tr>
<tr>
<td>Bacto Yeast Extract</td>
<td>5g</td>
</tr>
<tr>
<td>5M NaCl</td>
<td>2ml</td>
</tr>
<tr>
<td>1M KCl</td>
<td>2.5ml</td>
</tr>
<tr>
<td>1M MgCl₂</td>
<td>10ml</td>
</tr>
<tr>
<td>1M MgSO₄</td>
<td>10ml</td>
</tr>
<tr>
<td>1M Glucose</td>
<td>20ml</td>
</tr>
<tr>
<td>Sterile distilled water</td>
<td>To make 1L</td>
</tr>
</tbody>
</table>

### 2.11.4 Selection of positive clones

A number (generally 10-20) of bacterial colonies that grew under the appropriate antibiotic selection were tested for the required insert by colony PCR using a vector specific and insert specific primer pair. For pGEM-T the vector specific primers used were T7 as a forward primer and SP6 as a reverse primer. For the Gateway cloning (see 2.12 – Gateway cloning) involving the vector pDONR™221 the vector specific primers used were T7 as a forward primer and M13 as a reverse primer.

Primer sequences used

- T7 Forward 5’ – TAATACGACTCACTATAG – 3’
- SP6 Reverse 5’ – ATTTAGGTGACACTATAGAAT – 3’
- M13 Reverse 5’ – CAGGAAACAGCTATGACC – 3’

Colony PCRs were performed as described above for standard PCR, however instead of adding 1 µl of DNA template, a small portion of a single bacterial colony was added to the mix. Positive colonies were identified by gel electrophoresis. Positive colonies were then grown up overnight in 10 ml LB cultures and plasmid DNA obtained as described above. The target DNA sequence being cloned was then PCR amplified and sequenced to ensure it was entirely correct.
2.12 GATEWAY® cloning

Gateway cloning was used to generate the full-length *PCDH19* construct used in transfection assays to test the PCDH19 antibodies. Full length *PCDH19* was introduced into two Gateway compatible destination vectors. These were a 3’ Myc tagged vector and a 3’ V5 tagged vector.

2.12.1 Gateway primers

To facilitate Gateway® cloning (Invitrogen) of PCR products into a Gateway® competent donor vector through a BP recombination reaction, primers were generated with *attB* sites (underlined in the sequences below) at the 5’ ends. The *attB* primers were used to add *attB* sites at the 5’ and 3’ end of *PCDH19*, these sites could then be used to recombine the full-length *PCDH19* PCR product into a gateway compatible vector. The *attB* sequences that are generated as part of the primers are shown below.

Forward primer:

5’ – **GGGGACAAAGTTTGTACAAAAAAGCAGGCTNN**-(18-25 gene-specific nucleotides) – 3’

*attB1*

Reverse primer:

5’ – **GGGGACCCACTTTGTACAAGAAGCTGGGTC**-(18-25 gene-specific nucleotides) – 3’

*attB2*

The gene specific nucleotides of the reverse primer were designed to be in frame with 3’ Myc or V5 tag gateway compatible destination vector.
2.12.2 Producing attB-PCR Products

Standard PCR conditions as described above (2.5 – Polymerase chain reaction) were used to generate attB-PCR products, however the proof reading polymerase Platinum Pfx was used to amplify PCDH19. An extension temperature of 68°C was used. PCR products were then purified by gel extraction (as previously described in 2.2.6 PCR purification).

2.12.3 BP Recombination Reaction

The Gateway® BP Clonase™ Enzyme Mix kit (Invitrogen) was used to transfer the attB-PCR product into an attP-containing vector (pDONR™221) thereby generating an entry clone. The BP recombination reaction was performed as follows.

I. The following components were added to a sterile 1.5 microcentrifuge tube and mixed at room temperature.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Negative control</th>
<th>Positive control</th>
<th>Clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>5x BP reaction buffer</td>
<td>2 µl</td>
<td>2 µl</td>
<td>2 µl</td>
</tr>
<tr>
<td>pEXP7@50ng/µl</td>
<td>-</td>
<td>1 µl</td>
<td>-</td>
</tr>
<tr>
<td>attB PCR product (50-100ng/µl)</td>
<td>-</td>
<td>-</td>
<td>1-3 µl</td>
</tr>
<tr>
<td>pDONR™221 vector (150ng)</td>
<td>1 µl</td>
<td>1 µl</td>
<td>1 µl</td>
</tr>
<tr>
<td>TE</td>
<td>To 8 µl</td>
<td>To 8 µl</td>
<td>To 8 µl</td>
</tr>
<tr>
<td>LR clonase enzyme</td>
<td>2 µl</td>
<td>2 µl</td>
<td>2 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>10 µl</td>
<td>10 µl</td>
<td>10 µl</td>
</tr>
</tbody>
</table>

II. The BP Clonase™ enzyme mix was removed from the -70°C freezer and thawed on ice for approximately 2 minutes after which it was briefly vortexed twice.

III. To each BP reaction, 2 µl of BP Clonase™ enzyme mix was added, and the reaction was vortexed briefly twice.

IV. The BP reactions were incubated at 4°C overnight.

V. The following day 2 µl of proteinase K was added to each reaction and incubated at 37°C for 10 minutes to terminate the reaction.
VI. Of the BP reactions, 1 μl was transformed into DH5α™ (Invitrogen) chemically competent cells (see 2.11.3 – Transformation of vector DNA into competent cells) and the rest was stored at -20°C for a maximum of 7 days.

VII. BP recombination success was confirmed by colony PCR.

VIII. BP clones were also sequenced with either pDONR™221 specific primers T7/SP6 (see 2.11.4 – Selection of positive clones) or with sequence specific primers such as the attB-PCR primers (see 2.12 Gateway cloning) to confirm that the clone was in frame with the forward primer and that no mutations had occurred.

### 2.12.4 LR Recombination Reaction

The Gateway® LR Clonase™ Enzyme Mix kit (Invitrogen) was used to transfer the attL-pDONR™221 entry clone into an attR-containing vector (for example pDEST™3.2) thereby generating an expression clone. The LR recombination reaction was performed as follows.

I. The following components were added to a sterile 1.5 microcentrifuge tube and mixed at room temperature.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Negative control</th>
<th>Positive control</th>
<th>Clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR reaction buffer</td>
<td>2μl</td>
<td>2μl</td>
<td>2μl</td>
</tr>
<tr>
<td>pENTR-gus @50ng/μl</td>
<td>-</td>
<td>1μl</td>
<td>-</td>
</tr>
<tr>
<td>Entry clone (200ng)</td>
<td>-</td>
<td>-</td>
<td>1-5μl</td>
</tr>
<tr>
<td>Destination vector (300ng)</td>
<td>1-5μl</td>
<td>1-5μl</td>
<td>1-5μl</td>
</tr>
<tr>
<td>TE</td>
<td>To 8μl</td>
<td>To 8μl</td>
<td>To 8μl</td>
</tr>
<tr>
<td>LR clonase enzyme</td>
<td>2μl</td>
<td>2μl</td>
<td>2μl</td>
</tr>
<tr>
<td>Total volume</td>
<td>10μl</td>
<td>10μl</td>
<td>10μl</td>
</tr>
</tbody>
</table>

II. The LR Clonase™ enzyme mix was removed from the -70°C freezer and thawed on ice for approximately 2 minutes after which it was briefly vortexed twice.

III. To each LR reaction, 2 μl of LR Clonase™ enzyme mix was added and the reaction was vortexed briefly twice.
IV. The LR reactions were incubated at 4°C overnight.

V. 2 μl of proteinase K solution was added to each reaction and incubated at 37°C for 10 minutes.

VI. Of the LR reactions, 1 μl was transformed into DH5α™ (Invitrogen) chemically competent cells (see 2.13.3 – Transformation of vector DNA into competent cells) and the rest was stored at -20°C for a maximum of 7 days.

VII. Colonies that had effectively undergone the LR reaction were determined through antibiotic resistance selection on agar plates with the appropriate antibiotic resistance for the vector

VIII. LR recombination success was confirmed by colony PCR (see 2.11.4 – Selection of positive clones). LR clones were re-sequenced to confirm that the clone was in the correct reading frame and that no mutations had occurred.

2.13 Fibroblast tissue culture

2.13.1 Fibroblast media

The fibroblast cells were maintained in RPMI with a number of reagents added as detailed in the table below.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount added</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPMI</td>
<td>500 ml</td>
</tr>
<tr>
<td>L-Glutamine</td>
<td>8 ml</td>
</tr>
<tr>
<td>Benzlpenicillin</td>
<td>0.5 ml</td>
</tr>
<tr>
<td>Foetal Calf Sera (FCS)</td>
<td>55 ml</td>
</tr>
</tbody>
</table>

2.13.2 Establishment of primary fibroblast cultures from skin biopsy

I. Upon receipt of a skin punch sample it was washed in sterile PBS in a Petri dish

II. Excess fat was cut off and the skin punch was cut into small fragments and transferred to a T25 flask.
III. The tissue fragments were distributed over the bottom of the culture flask.

IV. 5 mls of media were gently added to the cells (at this stage the media used was as described above except 20% FCS was used instead of 10%).

V. Cells were gassed with CO₂ and incubated at 37°C.

VI. Cells were closely monitored and the media replaced every 2 days until at least three large fibroblast outgrowths had projected from different skin fragments.

VII. Cells were then split into 2 T25 flasks (this procedure is described below, see 2.13.3 – Splitting the fibroblast cells).

### 2.13.3 Splitting the fibroblast cells

Cells were split generally 1 in 3 once they had reached a confluency of approximately 90%.

I. The media was removed from the fibroblast culture.

II. The cells were washed in sterile PBS to remove all traces of the culture media.

III. Trypsine was added to the cells (0.5ml in a T25, 1 ml in a T75 and 5ml in a T175 flask)

IV. Cells were left at room temperature for 5 minutes or until the cells had lifted off the surface of the flask.

V. Cells were distributed by tapping the culture flask.

VI. A homogenous solution of the cells was made by mixing the cells through pippetting with some of the culture media.

VII. One third of the resulting mixture was added to a new tissue culture flask.

VIII. Additional culture media was added to the flask to give a final volume of 5 ml for a T25, 15ml for a T75 and 25 ml of a T175 flask.

IX. Cells were gassed and placed back in the incubator at 37°C.
2.13.4 Liquid nitrogen storage of cells

As soon as a primary fibroblast cell line had been established aliquots of cells were stored in liquid nitrogen for future use.

I. Freezing down media was made which consisted of RPMI, 20% FCS and 10% filtered DMSO. This media was wrapped in foil to inhibit light interfering with the DMSO.

II. Cells were harvested by trypsinization (as described for splitting cells see 2.13.3 – Splitting the fibroblast cells) followed by pelleting the cells through centrifugation at 1,100 rpm for 5 minutes.

III. The media was removed and the cell pellet resuspended in approximately 1 ml of freezing down media.

IV. The resuspended cells were transferred to a 1.5 ml CryoTube.

V. The CryoTubes were placed in a cryo freezing chamber, which contained isoproponol.

VI. The cryo freezing chamber was then placed in a -80°C freezer overnight.

VII. The following day the CryoTubes were transferred out of the cryo-freezing chamber and into a liquid nitrogen dewier.

2.13.5 Reviving cells from liquid nitrogen

I. Cells were removed from the liquid nitrogen dewier and immediately placed in warm water (~37°C) and allowed to thaw.

II. The cell suspension was removed from the CryoTube and added to 10 ml of culture medium (for this procedure 20% FCS is used).

III. Cells were centrifuged at 1,100 for 5 minutes and the media removed.

IV. The cell pellet was resuspended in ~5ml of culture media with 20% FCS.

V. Cells were transferred to the appropriate sized tissue culture flask based on the size of the cell pellet obtained (typically T75).
VI. Additional media was added where needed (to make it up to 25ml for a T75 flask).

VII. Cells were gassed and placed in the incubator.

2.14 Western blot analysis

2.14.1 General Lysis Method for Cultured Cells

I. Cells were collected from a T175 flask by scraping the cells off with a 30 cm cell scraper. Cells were placed in 10 ml tube and centrifuged at 5,000 rpm for 10 minutes.

II. The culture media was removed and the pellet was resuspended in 50-100 μl of RIPA+ buffer (50 mM Tris-HCl (pH 7.4), 1% (v/v) Nonidet P-40, 150 mM NaCl, 1 mM EDTA, 0.25% (v/v) sodium deoxycholate, 1X Protease Inhibitor Cocktail (SIGMA), 1 mM NaVO₄, 1 mM NaF, 1 mM PMSF).

III. Resuspended cells were incubated on ice with agitation for 15 minutes and were then passed through a 21-gauge needle 10X.

IV. Lysed cells were centrifuged at 12,000 rpm for 15 minutes at 4°C.

V. The supernatant was transferred to a clean 1.5 ml screw-capped tube and stored at -70°C. The concentration of the protein obtained was assessed by Bradford assay (see 2.14.2 Bradford protein assay).

2.14.2 Bradford Protein Assay

I. In a 96 well plate, BSA (Bovine serum albumin) was used to create a protein standard curve at concentrations of 0 μg, 10 μg, 20 μg, 40 μg, 60 μg, 80 μg and 100 μg, where each standard was represented three times on the 96 well plate.

II. Each sample to be analysed was diluted 1/10 and 1/50 in the diluted Protein Assay reagent and was represented three times at each concentration on the 96 well plate.

Chapter 2- 70
III. 200 µl of Protein Assay reagent (BIO-RAD) that had been diluted 1 in 4 was added to each well, making sure that no bubbles formed.

IV. Concentration of protein in each well of the 96 well plate was measured and calculated at λ570 using a Apollo 8 LB 912 plate reader (Berthold Technologies) running MikroWin 2000 OEM version software.

V. The output of this analysis was used to calculate the protein concentration of each sample.

2.14.3 Preparation of Protein Lysates for Gel Loading (Reducing Conditions)

Protein lysates were mixed with 10X NuPAGE® Reducing Agent (Invitrogen) and 4X NuPAGE® LDS sample buffer (Invitrogen) before gel loading. Protein lysate samples were prepared in either 10 µl or 15 µl volumes depending on the cell lysis method and protein concentration. For example, each 10 µl reaction contained 1-6.5 µl of protein lysate, 1X NuPAGE® LDS and 1X NuPAGE® Reducing Agent.

2.14.4 Running the Gel (Reducing Conditions)

Protein lysates were run on precast 12 well NuPAGE® Novex 4-12% Bis-Tris gradient gels (Invitrogen) using the XCell SureLock™ Mini-Cell (Invitrogen) according to manufacturer’s instructions.

I. The precast 12 well NuPAGE® Novex 4-12% Bis-Tris gradient gel cassettes were removed from their sealed pouches and rinsed in distilled water.

II. The tape covering the slot on the back of the gel cassette was removed as was the comb separating the 12 wells.
III. The gel wells were rinsed with 1X MES running buffer (50 mM 2-(N-morpholino)ethane sulfonic acid (MES), 50 mM Tris Base, 3.47 mM SDS, 1.03 mM EDTA, pH 7.3) (Invitrogen).

IV. After assembling the XCell SureLock™ Mini-Cell, the upper buffer chamber was filled with 200 ml of 1X MES buffer containing 500 µl of NuPAGE® Antioxidant (Invitrogen).

V. The lower buffer chamber was half filled with 1X MES buffer.

VI. After loading the protein lysate samples, the gels were run at 140 volts for 80 minutes to 2 hours.

VII. The XCell SureLock™ Mini-Cell was then dismantled and the gels removed by cracking apart the plates with a bevelled Gel Knife (Invitrogen).

VIII. Both the wells and the foot of the gel were cut off with the Gel Knife and discarded.

2.14.5 Western Blot Transfer

I. To transfer the separated proteins in the NuPAGE® Novex 4-12% Bis-Tris gradient gels to a nitrocellulose membrane (PALL Life Sciences) the XCell II™ Blot Module (Invitrogen) was used with the XCell SureLock™ Mini-Cell.

II. Both the blotting pads and Whatman #5 paper were pre-wet in 1X transfer buffer (Invitrogen).

III. The nitrocellulose membranes were first pre-wet in distilled water before being pre-wet in 1X transfer buffer. The 1X transfer buffer (25 mM Bicine, 25 mM Bis-Tris, 1.03 mM EDTA, 50 µM chlorobutanol, pH 7.2) contained 10% (v/v) methanol (analytical grade) when transferring 1 gel or 20% (v/v) methanol when transferring 2 gels.

IV. The XCell II™ Blot Module was assembled with back plate of the apparatus ⇒ 2x blotting sponge ⇒ Whatman filter paper ⇒ nitrocellulose membrane ⇒ NuPAGE gel
Whatman filter paper ⇒ blotting sponge ⇒ Whatman filter paper ⇒ nitrocellulose membrane ⇒ NuPAGE gel ⇒ Whatman filter paper ⇒ 2x blotting sponge and the front of the apparatus.

V. The XCell II™ Blot Module chamber was filled with 1X transfer buffer and the outer buffer chamber was filled with ice-cold distilled water.

VI. The separated proteins in the gels were then transferred to the nitrocellulose membrane by running the XCell II™ Blot Module at 30 volts for 90 minutes.

2.14.6 Ponceau Staining

Nitrocellulose membranes were stained with ponceau stain to determine if protein transfer was successful.

I. Nitrocellulose membranes were washed in distilled water.

II. Nitrocellulose membranes were then covered by Ponceau S solution (SIGMA) for 1 minute.

III. Distilled water was used to wash the nitrocellulose membranes until a protein-banding pattern became clear.

IV. Once confirmed that protein transfer did occur, the nitrocellulose membranes were washed with shaking in 1X Tris buffered saline solution plus Tween 20 (1X TBST; 20 mM Tris Base, 137 mM NaCl, 0.05% Tween 20 (v/v)) until no stain remained.
2.14.7 Immunoblot

I. Nitrocellulose membranes were blocked overnight in a 1X TBST plus 5% milk (w/v) solution at 4°C.

II. Primary antibodies were diluted in 1X TBST plus 1% milk (w/v) and incubated with the nitrocellulose membrane for 1 hour at room temperature.

III. Nitrocellulose membranes were then washed 3X for 10 minutes with agitation in 1X TBST at room temperature.

IV. Secondary antibodies conjugated to horseradish peroxidase were diluted in 1X TBST plus 1% milk (w/v) and incubated with the nitrocellulose membrane for 1 hour at room temperature.

V. Nitrocellulose membranes were then washed 4X 10 minutes with agitation in 1X TBST at room temperature.

2.14.8 Enhanced Chemiluminescent Detection

The ECL detection kit (Amersham Biosciences) was used to detect proteins on the nitrocellulose membrane labelled with horseradish peroxidase conjugated to the secondary antibody.

I. Excess 1X TBST was removed from the nitrocellulose membranes by blotting the edges of the membrane on to a paper towel.

II. Equal amounts of detection reagent 1 and detection reagent 2 was mixed together and placed on the nitrocellulose membrane forming a meniscus for 1 minute at room temperature.

III. ECL detection reagents were removed and the excess drained off by blotting the edges of the membrane on paper towels.
IV. Nitrocellulose membranes were placed in a clear plastic sleeve located in a film cassette. The nitrocellulose membranes were then exposed to autoradiography film (AGFA) for 15 seconds, 30 seconds, 1 minute, 2 minutes and 5 minutes.

V. The autoradiography films were assessed and if necessary a new sheet of autoradiography film was exposed to the nitrocellulose membrane for a suitable length of time.

2.15 Transient transfection of PCDH19

Over expression of PCDH19 in 293T cells was performed for use in determining the specificity of the PCDH19 antibodies, as we had been unable to detect native protein expression in the fibroblast cells assessed.

2.15.1 Cell preparation

I. Once a T75 flask of 293T cells was confluent, the cells were harvested by trypsinization as described for fibroblast cells (see 2.13.3 Splitting the fibroblast cells).

II. The resulting cell pellet was resuspended in culture media to make a 10 ml homogenous cell suspension.

III. Cells were then counted in order to determine the cell concentration. 50 μL of the cell suspension was mixed with 50 μL of trypan blue.

IV. The resulting mixture was added on to a haemocytometer and cells present in the four outer squares were counted and the cell concentration deduced.

V. 4 x 10^5 cells were added to each well in a 6 well plate.

VI. The 6-well plate was gassed and placed at 37°C.
2.15.2 Cell transfection

I. The following day the transfection reactions were setup.

II. 1 µg of plasmid DNA was diluted in 200 µl of culture media (minus FCS and Benzlpenicillin) and mixed gently by inversion.

III. In a separate tube 4 µl of Lipofectamine 2000 (Invitrogen) was diluted in 200 µl of culture media (minus FCS and Benzlpenicillin) and mixed gently by inversion.

IV. The lipofectamine solution was added to the DNA solution and gently mixed by inversion, then left to incubate at room temperature for 20 minutes.

V. During the incubation period the media was removed from the cells and the cells washed in sterile PBS.

VI. 1 µl of culture media (minus FCS and Benzlpenicillin) was added to each well of cells.

VII. 400 µl of the lipofectamine/DNA mixture was carefully added to avoid agitation to the cells.

VIII. Cells were gassed and returned to the 37°C incubator.

IX. After 3 hours the cells were topped up with 1 µl of culture media (with 20% FCS and minus the Benzlpenicillin).

X. Cells were then incubated for 24 hours at 37°C.

2.15.3 Cell harvest

I. At the time of cell harvest cells were washed twice in PBS (2 ml per well of cells). All washes and additions of solutions to the cells were performed as gently as possible to avoid unnecessary agitation and loss of cells.

II. Cells were then harvested into a 1.5 ml tube by scraping with a 20 cm cell scraper.

III. Cells were pelleted by centrifugation for 5 minutes at 5,000 rpm, the supernatant was then removed and the cell pellets stored at -80°C until required.
2.16 Northern and Southern blot analysis

The \textit{PCDH19} and \textit{PCDH11X/11Y} probes were generated through cloning then tested for their specificity and cleanliness on a southern blot, which had previously been made and used by the Molecular Diagnostic Laboratory, Women’s and Children’s Hospital, Adelaide. Once it had been established that the probe was clean and specific it was run on the northern blot. The northern blots used were the Human Brain Multiple Tissue Northern (MTN™) Blot II and the Human Brain Multiple Tissue Northern (MTN™) Blot III; purchased from Clontech. These blots were hybridised as per the manufacturer’s instructions described in brief below.

2.16.1 Prehybridization of Northern and Southern blots

I. Northern and southern blots were prehybridised with at least 5mls of pre-warmed ExpressHyb (Clontech) solution at 68°C with continuous shaking for 30 minutes.

II. Once the probe had been made the ExpressHyb solution was removed from the blot and was replaced with ExpressHyb which contained the labelled probe.

2.16.2 Probe labelling reaction

Radioactive labelling of the probe was performed using the Amersham MegaPrime™ DNA labelling system as per the manufacturer’s recommendations, in brief.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount</th>
<th>Volume added</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe DNA</td>
<td>50ng</td>
<td>2μl</td>
</tr>
<tr>
<td>Random primer</td>
<td>10μM</td>
<td>5μl</td>
</tr>
<tr>
<td>Sterile distilled water</td>
<td>Make up to 20μl</td>
<td>14μl</td>
</tr>
</tbody>
</table>
I. The DNA probe was denatured by heating the samples to 95°C for 2 minutes then placed on ice for 5 minutes. The remaining reagents for the reaction were then added which included

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount</th>
<th>Volume added</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labelling Buffer</td>
<td>50ng</td>
<td>2µl</td>
</tr>
<tr>
<td>Enzyme</td>
<td>10µM</td>
<td>5µl</td>
</tr>
<tr>
<td>EASYtides™ α²²P dCTP</td>
<td>Make up to 20µl</td>
<td>14µl</td>
</tr>
</tbody>
</table>

II. Samples were incubated for 20 minutes at 37°C.

III. After incubation the reaction was terminated by the addition of 10 µl of 100 mM EDTA.

IV. The labelled probe was cleaned up by passing it through an Illustra™ ProbeQuant™ G-50 Micro Column by centrifugation at 2,000 rpm for 30 seconds.

V. The cleaned up probe was denatured by heating at 95°C for 2-5 minutes then chilled quickly on ice.

VI. The denatured probe was added to 5 ml of ExpressHyb solution and the resulting mixture added to the desired prehybridized blot (northern or southern) and incubated overnight at 65°C in a rotating incubator.

### 2.16.3 Washing of hybridised blots

Washing solutions

Wash solution 1

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSC</td>
<td>2X</td>
</tr>
<tr>
<td>SDS</td>
<td>0.05%</td>
</tr>
</tbody>
</table>
Wash solution 2

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSC</td>
<td>0.1X</td>
</tr>
<tr>
<td>SDS</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

I. Hybridised blots were rinsed twice in wash solution 1 for 10 minutes at room temperature.

II. Followed by two washes for 5 minutes each in wash solution 1 at 50°C.

III. Blots were washed finally for 10 minutes in wash solution 2 at 50°C.

IV. Between washes the blots were tested with the geiger counter to assess if the level of background radiation was low enough to stop washing and proceed with the exposure.

V. Once the hybridised blots had an acceptable level of background radiation they were sealed in a plastic sleeve and exposed to X-ray film for up to a month at -80°C.

VI. X-ray films were removed and processed at a number of time points during the exposure, these were used to assess if additional washing was required and how long the blot should be left exposing.

2.17 Specific methods used in the microarray analysis

2.17.1 Fibroblast culture for the microarray analysis

Fibroblast cell lines for all of the EFMR family members and controls we had access to were revived from liquid nitrogen storage. Cell numbers were amplified through successive cell splitting to obtain at least 4 T175 culture flasks of cells for each individual. These cells were amplified in a manner such that all of the cell lines were harvested at passage 10 or as close to passage 10 as possible. Cells were harvested as previously described and stored at -80°C until all the required cells had been obtained.
2.17.2 RNA extraction

RNA was extracted in two batches as per the procedure described previously. However due to the large quantities of RNA expected from the extraction the elution step was performed twice. The RNA was eluted by adding 30 μl of RNase-free water to the column membrane and centrifuging for 30 seconds at 13000 rpm. This stage was then repeated by addition of a further 30 μl of RNase-free.

The resulting RNA was analysed on a spectrophotometer to obtain the concentration, 100-200ng of RNA was also run on a 1% agarose gel to ensure the quality of the RNA. The extracted RNA was stored at -80°C.

Approximately 10 μg of RNA was sent on dry ice to the Australian Genome Research Facility (AGRF, Melbourne, Australia) for labelling and hybridisation to an Affymetrix Human Exon 1.0 ST array (Affymetrix, Santa Clara, CA, USA).

2.17.3 Data analysis

Quality assessment of the array results was performed by the AGRF, all samples submitted passed quality assessment and were used in the array analysis. Analysis was performed using Partek Genomic Suite – Turning data into discovery® version V6.4 and the probe set annotation file HuEx-1_0-st-v2.

The microarray results were returned to us as CEL files.
2.17.3.1 Importing CEL files into Partek

I. The CEL files to be analysed were imported into the Partek program through clicking on “File” selecting “import” then “Affymetrix” and finally “CEL files”.

II. The desired CEL files to be analysed were selected and moved to the import file list then “Next” was clicked.

III. In the following options box that appears the files were imported under customised settings. This was performed by clicking on the customise icon.

IV. From there samples were imported under the Partek default settings by ticking the Partek default settings option. This involved the use of; interrogating probes, include core meta-probeset, pre-background adjustment with adjustment for GC content, RMA background correction, quantile normalisation, log probes using base 2 and probe set summarisation using the mean.

V. A data spreadsheet was then automatically created.

2.17.3.2 Adding a factor to be interrogated

Factors to be interrogated in the data were added into the data spreadsheet through the addition of extra columns.

I. Additional columns were added to the data spreadsheet by going to “Edit” in the menu header at the top of the program.

II. Under “Edit” “Add Rows/Columns” was selected.

III. The “Add column” tab was selected.

IV. The column being added was labelled “Status” and the “Type” of “Categorical” was selected and then the “Attribute” of “Factor” was selected.

V. This introduced a new column into the data spreadsheet.
VI. The status either of each sample being analysed was added into the new column as either Affected (for EFMR females with a mutation in \textit{PCDH19}), Transmittting (for males with a mutation in \textit{PCDH19}) and control (for the male and female controls).

\textbf{2.17.3.3 Principal Component Analysis (PCA)}

I. Through the dropdown “Workflow” menu on the right hand side of the Partek window the option of “Gene Expression” was selected.

II. Then the “QA/QC” option opened and “Principal Component Analysis” selected.

III. This generated a graphical output of the PCA results.

IV. The parameters of the graph were changed to reflect the analysis we wanted performed. This involved selecting “Status” as the factor being displayed. Opening up the dropdown menu on the top right hand side of the graph window, and selecting “Status” achieved this.

V. Elipsoids were added to the graph by opening the “Edit” options on the main menu at the top of the window and selecting “Plot properties”.

VI. In the “Plot properties” window the “Ellipsoid” tab was selected.

VII. “Add Ellipsoid” was then selected and the factor “Status” was selected and added to the grouped variables column.

VIII. “Apply” was pressed and the ellipsoids added.

\textbf{2.17.3.4 Gene summary}

Gene summary analysis was performed on the data in order to integrate the expression data from all of the probes for a particular gene into a single representative expression value.

I. Opening up the “Tools” drop down menu and selecting the “Gene Summary” option performed the “Gene Summary” integration.
II. The gene summary was performed under the default Partek settings.

III. The output of this analysis is a new window containing a spreadsheet type layout of the data. All further analysis was performed on the summary data.

2.17.3.5 Generation of Dot plots and Box and Whisker plots of a genes expression level

The expression level of a gene identified in each subject being assessed in a particular analysis could be visualised in Partek by Dot plots and Box and Whisker plots. This was performed on the integrated gene summary data.

I. A Dot plot and a Box and Whisker plot was generated for a gene of interest by right clicking on the gene then selecting “Dot plot”.

II. In the new window containing the Dot plot “Status” was then selected from the drop down menu in the upper right hand corner.

2.17.3.6 Identification of differentially expressed genes

One-way ANOVA analysis was used to identify genes that were differentially expressed between the different status’s being analysed, for example, affected EFMR females vs control females. This analysis was performed on the gene summary data.

I. The “Analysis” option was opened in the “Gene Expression” section of the “Workflow” drop down menu.

II. “Detect differentially expressed genes” was selected from the “Analysis” options.

III. Once the “one-way ANOVA” window opens “Status” was selected and moved to the right hand ANOVA factor column by clicking on “Add factor”.

IV. A contrast was then added by clicking on the “Contrast” button, opening up a new window.
V. Here one of the two different status variables being interrogated, for example, affected and control, was added to the upper box of contrast level by selecting the status and then clicking on “Add contrast level”.

VI. The above step was repeated for the remaining status variable except this variable was added to the lower contrast level box.

VII. From there the “Add contrast” button was selected. From there the program returns to the ANOVA options window, here “Apply” was selected and the ANOVA analysis performed.

VIII. The results of the one-way ANOVA analysis were displayed in a new spreadsheet window.

2.17.3.7 Creation of a gene list

Gene lists were generated which contained genes with a P-value of less that or equal to 0.05 and 0.01.

I. Once the ANOVA analysis had been performed gene lists were generated from the Partek window containing the ANOVA results.

II. This was achieved by opening the “Gene Expression” section of the “Workflow” drop down menu and selecting the “Create gene list” option.

III. This opens up the Gene list option window, from there “Specify new criteria” was selected.

IV. The desired P-value was entered (either 0.01 or 0.05) and the “Unadjusted P-value of less than or equal to” was selected and the desired list generated.

V. These gene lists were saved as “txt” files and imported into Excel for further analysis.
2.17.3.8 Detecting the source of variation in the data

Detection of the source of variation in the data was performed in the Partek program on the full gene lists generated from the one-way ANOVA analysis and on the P-value >0.01 or P-value >0.05 gene lists.

I. Source of variation was analysed by selecting the “Plot variation” option in the “Genes expression” section of the “Workflow” dropdown menu.

II. Once this option had been clicked a new window with a pie chart appears displaying the source of variation.

III. From there the bar chart option was selected and a bar chart of the variation was generated.

2.17.4 Interpretation of the biological relevance of the microarray results using DAVID

The methods used for analysis of the biological relevance of the microarray results were based on a publication by Huang da (Huang da et al., 2009).

2.17.4.1 Uploading gene lists to DAVID

I. This was achieved by going to the DAVID web address http://david.abcc.ncifcrf.gov/home.jsp and clicking on “Start Analysis” on the header.

II. A gene list of interest, in the form of gene symbols, was copied from excel and pasted in the “Gene list manager panel”

III. The appropriate gene identifier of “Official Gene Symbol” was selected.

IV. The gene list option was ticked as the list type.

V. The gene list was submitted.
VI. Using the “background” tab in the “Gene list manager panel” the list of available Affymetrix background lists was displayed and the background HuEx-1_0-st-v2 selected.

2.17.4.2 Functional annotation analysis

Once the gene list of interest had been successfully uploaded “Functional annotation clustering” was performed by clicking on the “Functional annotation clustering” icon.

Exploration of the enriched cluster results obtained involved a number of steps:

I. The fold enrichment score was added to the results displayed. This was achieved through expanding the options section by clicking on the + next to “options”. From there the “fold enrichment” option under display was ticked and then the new results obtained by clicking the “return using options” button.

II. Matrix representations of the interactions between the genes and functional annotations identified in an enriched cluster were generated through clicking on the matrix image displayed in the region of headings above each enriched cluster identified.

III. A list of all of the genes associated with an enriched cluster was obtained through clicking on the G also displayed in the region of headings above each enriched cluster identified.

IV. The genes associated with each specific functional annotation involved in an enriched cluster were obtained through clicking on the blue bar associated with each annotation function.

V. Specific details about any of the genes identified in these clusters was obtained by clicking on a gene from either of the above two lists.

VI. Clicking on the RG associated with any gene from the aforementioned gene lists revealed a list of related genes.
The level of stringency associated with the enriched clustal analysis was also altered by clicking on the drop down menu next to “classification stringency” and altered as required.

2.17.5 Interpretation of the biological relevance of the microarray results using Ingenuity Pathways Analysis

The microarray data was also analysed for biological relevance using Ingenuity Pathways Analysis Software. The outcome of this analysis was similar to that obtained by using the DAVID program. Due to access issues with the Ingenuity Pathways Analysis Software the DAVID software was the predominant program used.

Ingenuity Pathway Analysis was performed using Ingenuity IPA version 7.6, Ingenuity Systems Inc.

I. Under the analysis options “Core” was selected and the “analyse dataset” option clicked.

II. Data was then uploaded to the Ingenuity software by selecting the required excel spreadsheet of the significantly altered gene list being assessed.

III. Under the “upload data settings”
   a. The file format selected was the “flexible format option”.
   b. “Yes” was selected for the question “contains column heading?”
   c. HUGO gene symbol was selected as the “Identifier type”
   d. The “array platform used” was added by selecting Human exon 1.0ST array from the Affymetrix options.
   e. Identifiers were then selected for each column of uploaded data with the first column labelled ID, the second observation 1 – P-value and the third column identified as observation 1 – fold change.
f. The “save and create analysis” button was selected and the data was uploaded.

g. Once the data had been uploaded the “Create Core Analysis” window appears

h. All options were left as the programs default settings except for in the “general settings” tab the “functions/pathways/tox list analysis reference set” was changed to the Affymetrix – Human Exon 1.0 ST Array data set.

i. No expression values parameters were selected.

IV. “Run analysis” was then selected

V. Once the analysis results had been returned the “Summary”, “Networks”, “Functions” and “Canonical pathways” tabs were all interrogated.

VI. Under the “Canonical pathways” tab pathways of interest were selected by clicking on the bar on the bar chart that corresponded to the pathway of interest.

VII. “View pathway” was then selected to obtain a diagram of the pathway with genes highlighted that were present in the gene list submitted for analysis.
Chapter 3 – Gene identification and molecular characterisation of EFMR

SECTION 1

3.1 INTRODUCTION AND BACKGROUND RESULTS
3.2 BACKGROUND RESULTS
  3.2.1 EFMR GENE IDENTIFICATION
  3.2.2 ANNOTATION OF THE FULL LENGTH PCDH19 GENE AND PROTEIN

SECTION 2

3.3 PCDH19 CHARACTERISATION
  3.3.1 ALTERNATIVE SPlicing OF PCDH19
  3.3.2 EFMR MUTATIONS
  3.3.3 INCOMPLETE PENETRANCE
  3.3.4 PREDICTED FUNCTIONAL CONSEQUENCE OF THE PROTEIN TRUNCATION MUTATIONS
  3.3.5 PREDICTED FUNCTIONAL CONSEQUENCE OF THE MISSENSE MUTATIONS
  3.3.6 PCDH19 EXPRESSION
  3.3.7 EFMR BIOLOGICAL MECHANISM

3.4 DISCUSSION
  3.4.1 MUTATION IDENTIFICATION
  3.4.2 EXPRESSION
  3.4.3 A ROLE FOR PCDH11Y IN EFMR?
  3.4.4 HEMIZYGOUS DELETION OF PCDH19 IN AN AFFECTED MALE
  3.4.5 CFNS MOLECULAR MECHANISM
  3.4.6 CONCLUSION

3.5 CONTRIBUTIONS
Preamble

This results chapter is broken up into two sections; the first consists of background results that were performed prior to my involvement in the project. These results are outlined as they set the scene for the work performed in my Ph.D. project. Section two details work conducted after I became involved in the project. This section contains work performed by others as part of a major collaborative effort, work conducted by others is acknowledged at the end of this chapter in a detailed list of contributions.

Section 1

3.1 Introduction and background results

Epilepsy and Mental Retardation Limited to Females (EFMR) is an intriguing X-linked disorder affecting heterozygous females and sparing hemizygous males (Fabisiak and Erickson, 1990, Juberg and Hellman, 1971, Ryan et al., 1997). The unusual pattern of inheritance seen in large multigenerational families, with the presence of a number of affected females and transmission occurring through unaffected males, is characteristic of EFMR.

The ~25 cM EFMR linkage interval contained ~246 genes of which 150 genes were believed to be expressed in the brain and were therefore possible candidate genes for EFMR (Scheffer et al., 2008). Included in the EFMR linkage interval are also 9 known XLID genes; *BRWD3, SRPX2, TIMM8A, NXF5, PLP, PRPS1, ACSL4, PAK3* and *DCX* (figure 3.1) (Chiurazzi et al., 2008).
Figure 3.1 Known XLID genes in the EFMR linkage interval

Ideogram of the X-chromosome with the position of the EFMR linkage interval, as indicated by the teal vertical line. The known XLID genes contained within this interval are indicated and their approximate position shown on the expanded illustration of the EFMR linkage interval (Chiurazzi et al., 2008).

3.2 Background results

3.2.1 EFMR gene identification

Additional families that were highly suggestive of EFMR based on their clinical history and pattern of inheritance were identified; 2 from Australia and 2 from Israel (Scheffer et al., 2008). Linkage analysis was carried out with the aim of identifying recombination events that could narrow down the previously known ~25 cM or 34 Mb linkage interval for EFMR (Ryan
et al., 1997). Unfortunately no such recombinants were identified and the existing interval could not be reduced. Whilst none of the families analysed was of sufficient size to independently confirm linkage to Xq22, the absence of recombination events within the region and their combined LOD score was consistent with the gene defect in these families mapping to the same chromosomal location as for EFMR (Ryan et al., 1997, Scheffer et al., 2008).

Four genes, *PCDH11X*, *DIAPH2*, *NAP1L3* and *SRPX2* from within the linkage interval were selected by Scheffer *et al* as good candidate genes based on their localisation and known functions (Scheffer et al., 2008). No mutations were identified in any of the candidate genes screened. A summary of the genes screened in the EFMR probands is outlined in table 3.1.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Cytoband</th>
<th>Protein name</th>
<th>Biological function</th>
<th>Associated disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>PCDH11X</em></td>
<td>Xq21.2</td>
<td>Protocadherin 11, X-linked</td>
<td>Cell adhesion</td>
<td>-</td>
</tr>
<tr>
<td><em>DIAPH2</em></td>
<td>Xq22</td>
<td>Diaphanous, drosophila, homolog, 2</td>
<td>Cell motility</td>
<td>Premature ovarian failure, sterility</td>
</tr>
<tr>
<td><em>NAP1L3</em></td>
<td>Xq21.3-q22</td>
<td>Nucleosome assembly protein 1 – like 3</td>
<td>Nucleosome assembly</td>
<td>-</td>
</tr>
<tr>
<td><em>SRPX2</em></td>
<td>Xq22.1</td>
<td>Sushi repeat containing protein, X-linked, 2</td>
<td>Signal transduction, growth factor?</td>
<td>X-linked rolandic seizures associated with oral and speech dyspraxia and mental retardation (RESDX)</td>
</tr>
</tbody>
</table>

**Table 3.1 Negative candidate genes screened in EFMR patients**

In light of the large number of additional potential candidate genes remaining in the EFMR linkage, interval probands from three of the EFMR families characterised in (Scheffer et al., 2008) were included in the large scale re-sequencing project conducted by the International Genetics of Learning Disability group (IGOLD) (Tarpey et al., 2009). This comprehensive X-
chromosome exon re-sequencing project was performed with the aim to identify novel XLID genes in a cohort of 208 families where all, at that time known, XLID genes had previously tested negative (Tarpey et al., 2009).

Large scale re-sequencing of 719 X-chromosome genes in the probands from three of the EFMR families (figure 3.2) identified a single missense nucleotide change, c.2012 C>G (p.S671X) in protocadherin 19 (PCDH19) in one of the EFMR families (Family 3, figure 3.3). This single base substitution was predicted to result in premature truncation of PCDH19 protein at amino acid S671X. The c.2012 C>G nucleotide change was not identified in 208 male probands from suspected XLID families or in ~750 control X-chromosomes (consisting of 350 male and 200 female controls). No other potentially deleterious nucleotide changes were identified in this proband in any of the other 718 X-chromosome genes sequenced. This nucleotide change in PCDH19 looked promising as the causative mutation for EFMR, in this individual, as the nucleotide change was predicted to result in premature termination of the protein. However no potentially deleterious sequence changes were identified in PCDH19 in the remaining two probands from families 1 and 2 (figure 3.2).
Figure 3.2 Pedigrees of three suspected EFMR families

Illustrating the characteristic inheritance pattern, with multiple affected females and the disorder passed on through unaffected transmitting males. Families 1 and 2 are families obtained from Australia, family 3 was obtained from Israel. The proband for each family is indicated by an arrow. The three families had been described in greater detail previously (Scheffer et al., 2008).

Figure 3.3 Identification of a missense change in PCDH19

Sequence chromatogram of the missense change identified in PCDH19 in family 3, one of the suspected EFMR families. The specific missense change identified is indicated by an arrow above the sequence chromatogram. The pedigree of this family is also illustrated with the proband sequenced indicated by an arrow.
The identification of a mutation in \textit{PCDH19} in one of the EFMR families came as a surprise. Firstly \textit{PCDH19} was not initially considered as a candidate gene primarily because it mapped outside the EFMR linkage interval (Ryan et al., 1997). \textit{PCDH19} was originally positioned at Xq13.3 (Wolverton and Lalande, 2001). Wolverton \textit{et al} were the first to identify \textit{PCDH19} as a novel protocadherin based on database searches for sequence homology to the conserved domain, CM2, which had previously been identified in other \(\delta\)-protocadherins (Wolverton and Lalande, 2001). The characteristic CM1 and CM2 domains of protocadherins and the high level of conservation of the CM2 domain across the \(\delta2\)-protocadherins is demonstrated in figure 3.4. Wolverton \textit{et al} constructed a 6.1 kb putative \textit{PCDH19} cDNA transcript which mapped to the PAC clone, RP11-99E24, through high-throughput genomic sequence (htgs) database searches (Wolverton and Lalande, 2001). Clone RP11-99E24 is an X-chromosome clone containing the WI-14790 marker at Xq13.3. The chromosomal location of \textit{PCDH19} was determined as Xq13.3 and this was entered into the National Centre for Biotechnology Information (NCBI). However further investigations into \textit{PCDH19} identified that the Ensemble and UC Santa Cruz Genome Browser (UCSC) databases have \textit{PCDH19} located at Xq22, which is within the EFMR linkage interval.

The second inconsistency was that a putative mutation was found in only one of these families. Given that the c.2012 C>G (p.S671X) change in \textit{PCDH19} was a clear protein truncation mutation this gene was investigated further before exclusion.

\textbf{3.2.2 Annotation of the full length \textit{PCDH19} gene and protein}

Investigations into \textit{PCDH19} revealed that both the NCBI referred chromosomal location and the \textit{PCDH19} (gene ID 57526, NM_020766) open reading frame (ORF) were incorrect (Appendix A, Dibbens et al., 2008). Comparative sequence analysis of \textit{PCDH19} homologs and paralogs revealed that the annotation of the human \textit{PCDH19} ORF was incomplete. This
was also consistent with a previous report (Wolverton and Lalande, 2001). Following further work on the gene it was concluded that the NCBI database annotated \textit{PCDH19} ORF was missing 1497 bp at the 5’ end of the gene, encoding 499 aa at the N-terminus, as illustrated in figure 3.4. The completed ORF and protein of \textit{PCDH19} was constructed and annotated, and submitted to GenBank under the accession number EF676096 (Appendix A, Dibbens et al., 2008).

\textit{PCDH19} genomic sequence (GeneBank accession AL355593) encompasses 168065 bp and is oriented in a telomere to centromere direction. The genomic sequence transcribes into a processed 9765 bp mRNA sequence including; a 1676 bp 5’UTR, 3447 bp of core coding sequence and a 4642 bp 3’UTR. The ORF consists of six exons, the largest being the 2147 bp exon 1, which encompasses over half of the ORF. The 141 bp exon 2 is subject to alternative splicing (this will be discussed in further detail later, section 3.3.1). The ORF translates into an 1148 aa protein.
**Figure 3.4 Genomic, gene and protein structure of PCDH19**

*PCDH19* is positioned on the X-chromosome in a telomere to centromere orientation. The exon/intron structure of *PCDH19* is illustrated here on the background of its genomic sequence (GenBank accession AL355593). The position of the *PCDH19* exons/introns is not drawn to scale. The size of individual *PCDH19* introns in nucleotide base pairs [bp] is illustrated. The six individual *PCDH19* exons are labelled and their size indicated. Alternative splicing of *PCDH19*, which involves exon 2 is indicated by dotted lines below the exon structure, connecting exon 1 to 3 (alternative splicing will be discussed in further detail later). The lower panel illustrates the protein structure of PCDH19 with the amino acid position of the protein domains indicated. PCDH19 consists of a signal peptide (SP), six extracellular cadherin repeats (EC1-6), a transmembrane domain (TM), and known cytoplasmic domains (CM1 and CM2). The alternatively spliced exon 2 and the 499 aa of PCDH19 that were not previously annotated in the NCBI database are also illustrated on the protein structure. (Appendix A, Dibbens et al., 2008)

Analysis of PCDH19 protein revealed that it is composed of a 23 aa signal peptide, six extracellular cadherin repeats ranging in size from 106 – 115 aa, a 27 aa transmembrane domain and a 743 aa cytoplasmic domain containing the CM1 and CM2 domains. The CM1 and CM2 domains are characteristic features of δ protocadherins, as illustrated in figure 3.5 (Wolverton and Lalande, 2001). The CM1 domain consists of 27 aa while the CM2 domain consists of 17 aa that are highly conserved across the δ-family with respect to their identity and position (figure 3.5).
Clustal-W alignment of the members of the δ2-protocadherin subfamily illustrates the characteristic CM1 and CM2 domains which define this subfamily. The high level of conservation of the CM2 domain that was used to identify the existence of PCDH19 as a novel member of the δ-protocadherin sub-group. These domains are represented in a teal box for the CM1 domain and a purple box for the CM2 domain.

**Figure 3.5 Characteristic CM1 and CM2 domains of δ-protocadherins**
Section 2

This was where the EFMR project was up to when I commenced my Ph.D. project.

3.3 *PCDH19* characterisation

3.3.1 Alternative splicing of *PCDH19*

Alignment of *PCDH19* expressed sequences tags (ESTs), obtained from the UniGene cluster Hs 4993, which was used in the full annotation of *PCDH19* ORF also suggested that there could be alternative splicing of exon 2 (figure 3.6). Primers flanking the predicted alternative exon 2 were designed, by the candidate, with the forward primer positioned at the end of exon 1 and a reverse primer in exon 3. Reverse-transcriptase PCR (RT-PCR) amplification of cDNA from a range of human tissues confirmed the presence of two isoforms of *PCDH19* (figure 3.6). Both isoforms are present in all of the tissues analysed. The predominant isoform is the E2 isoform. Sequence analysis confirmed that these are the two isoforms predicted by the EST alignment.
Figure 3.6 Alternative splicing of *PCDH19* exon 2

A, Alignment of *PCDH19* EST’s obtained from the NCBI database reveals a possible alternatively spliced exon, with only some EST’s containing the additional sequence (in red).
B, Agarose gel results obtained from PCR amplification of cDNA from a range of human tissues both the short and long isoforms of *PCDH19* were amplified in all tissues analysed indicating that exon 2 is alternatively spliced in all of these tissues.

Alternatively spliced exon 2 is situated at c. 2146A – 2286G which corresponds to amino acids p.S716- R763. To date there is no functional domain attributed to the coding sequence of this exon, which is situated 11 amino acids into the cytoplasmic domain of *PCDH19*. Interestingly, exon 2 alternative splicing has so far been identified only in *PCDH19*. It is yet to be determined if there are any distinct roles for each of the isoforms.

### 3.3.2 EFMR mutations

After annotation of full length *PCDH19* mRNA the additional coding region was sequenced in the probands from family 1 and 2, as well as the proband from an additional family (family
4, figure 3.7). Sequence analysis identified 3 additional potentially deleterious changes in PCDH19 in these families (figure 3.7). Two of these changes, c.253 C>T and c.2030_2031insT (figure 3.7) were predicted to result in premature protein truncation of PCDH19, p.Q85X and p.L677FfsX41 respectively, while the remaining change, c.1322 T>A, was predicted to result in a missense amino acid change p.V441E. All four changes segregated with the disorder in all four families (Appendix A, Dibbens et al., 2008).

Figure 3.7 Identification of additional missense and nonsense changes in PCDH19

Sequence chromatogram of the additional missense and nonsense changes identified in PCDH19 in three suspected EFMR families. The specific change identified in each family is indicated by an arrow above the sequence chromatograms. The pedigrees of these families are illustrated with the probands sequenced for each family indicated by an arrow.

As the next step we analysed samples from the original EFMR family first published in 1971 by Juberg et al (Juberg and Hellman, 1971) and reanalysed by Ryan et al (Ryan et al., 1997). We identified a single cytosine base insertion, c.1091_1092insC, which is predicted to result...
in protein truncation 10 aas after the tyrosine at position p.366, p.Y366LfsX10 (Appendix A, Dibbens et al., 2008) (figure 3.8). We also sequenced the proband from a new EFMR family from Ireland through collaboration. In this family we identified a single base pair deletion c.357delC, which is predicted to result in a premature termination of p.K120RfsX3 (Appendix A, Dibbens et al., 2008) (figure 3.8). Both amino acid changes segregated with the disorder in the respective families.

Figure 3.8 Nonsense changes identified in PCDH19 in an American and Irish family

Pedigrees of these two additional families are illustrated. Family 5 was obtained from Ireland, while family 6 is the original EFMR family from the USA. Sequence chromatogram illustrating nonsense changes identified in PCDH19 in these suspected EFMR families is shown. The specific nonsense change identified in each family is indicated by an arrow above the sequence chromatograms. The pedigrees of these families are illustrated with the probands sequenced for each family indicated by an arrow.

With the identification of unique, deleterious base pair changes in PCDH19 in all six EFMR families we were, at this point, confident we had identified the EFMR gene. Subsequently, we started to screen patients with epilepsy and / or ID with or without a family history, to see if
we could identify additional individuals and families with EFMR. In the initial phases of this follow up screen performed by the candidate, the details of which will be discussed in results chapter 4, a further missense change c.1671C>G, p.N557K was identified in two affected sisters from family 7 (figure 3.9).

**Figure 3.9 Additional small EFMR family identified**

Pedigree of an additional EFMR family identified by follow up PCDH19 screening is illustrated. Family 7 was obtained from Australia. Sequence chromatogram illustrates the missense change identified in PCDH19 in this family ascertained from sequencing unrelated females with epilepsy and ID. The nonsense change identified in is indicated by an arrow above the sequence chromatogram. The pedigree of this family is illustrated with the probands for each family indicated by an arrow.

In total we have identified six PCDH19 mutations, in known, large EFMR families, as well as an additional mutation in a small family. Five of the seven mutations identified are protein truncation mutations and the remaining two are missense mutations. All mutations cluster in exon 1 of PCDH19, which encodes six extracellular, cadherin repeat domains. A summary of these seven PCDH19 mutations is presented in figure 3.10. The candidate identified the mutation in family 7.
Figure 3.10 Schematic representation of PCDH19 protein illustrating the mutations identified in EFMR

Representative diagram of the PCDH19 protein illustrating the signal peptide (purple, aa 1-23), extracellular cadherin repeats (EC - teal, aa 24-678), transmembrane domain (TM – pink, aa 679-705), cytoplasmic region (blue, aa 706-1448), alternatively spliced exon 2 (E2 – dark blue, aa 717-763) and conserved cytoplasmic domains (CM – light blue, CM1 aa 902-936 and CM2 aa 956-972). The positions of the mutations identified in PCDH19 in families with EFMR are shown. The reporting of these results appears in (Appendix A, Dibbens et al., 2008).

3.3.3 Incomplete penetrance

The identification of causative mutations in EFMR families enables direct assessment of their penetrance in females and the proportion of mutation carriers who have EFMR. Incomplete penetrance was first identified in the original EFMR family (family 6), where an obligate transmitting male gave rise to three affected daughters and one unaffected daughter (Ryan et al., 1997, Fabisiak and Erickson, 1990, Juberg and Hellman, 1971) (family 6, IV-12, highlighted in figure 3.11), it could not be determined if this female was an obligate carrier as she only had male children. Sequencing of PCDH19 in this unaffected female revealed that she does have the c.1091_1092insC (p.Y366LfsX10) mutation (Appendix A, Dibbens et al., 2008). Family 5 also contains a non-penetrant female (III-2) (highlighted in figure 3.11). In the six EFMR families analysed we have identified only two females (highlighted in figure 3.11), out of 68 analysed, that have the familial mutation in PCDH19, but are classified as
unaffected (based on their original assessment). Therefore EFMR has an estimated female penetrance of 97%. Penetrance in males remained at 0%.

Figure 3.11 Incomplete penetrance in EFMR

EFMR pedigrees in which incomplete penetrance occurs. Unaffected females who carry a mutation in \textit{PCDH19} are illustrated in a purple box.

3.3.4 Predicted functional consequence of the protein truncation mutations

The five premature protein truncation mutations identified in \textit{PCDH19} in EFMR all occur in the EC cadherin repeat domains (figure 3.10). Protein products from the protein truncation mutations, if they are produced, will lack the transmembrane and cytoplasmic domain of PCDH19. These truncated proteins are likely to be secreted from the cell and unlikely to function as normal in cell-cell adhesion.

3.3.5 Predicted functional consequence of the missense mutations

Whilst we were confident that mutations in \textit{PCDH19} were causative of EFMR we assessed the likely functional consequence of the two missense mutations we had identified (p.V441E
and p.N557K). We initially queried these two missense changes in the online programs *PolyPhen* (http://genetics.bwh.harvard.edu/pph/) and *Sorting intolerant from tolerant v.2* (SIFT) (http://blocks.fhcrc.org/sift/SIFT_seq_submit2.html). These programs predict computationally the possible impact of amino acid replacements on the structure and function of a given protein, and can be used to predict whether missense changes are likely to be tolerated in a given protein. To achieve this the programs search for similar sequences, align these sequences, and calculate scores based on the amino acid variation present at a given position in closely related sequences.

The two programs give slightly different result read outs, *PolyPhen* gives a prediction about a specific amino acid change submitted, whilst SIFT v.2 generates a list of all the amino acids predicted to be tolerated and not tolerated for every amino acid position in a given protein. *PolyPhen* analysis predicted that both p.V441E and p.N557K were “probably damaging”. SIFT v.2 predicted that neither of the substituted amino acids (441E and 557K) would be tolerated in the protein. This further supports the causative involvement of *PCDH19* mutations in EFMR.

To examine the conservation of the mutated residues visually we performed a clustal-W alignment of PCDH19 against various protocadherins and PCDH19 from various species (figure 3.12). Val441 is invariant across PCDH19 from the various species assessed; it is also highly conserved across the protocadherin family. Asn557 is invariant across both the protocadherin family members assessed and PCDH19 from various species. The high level of conservation of these two residues suggests that they are important to the correct functioning of PCDH19.
Partial alignment of human PCDHs and orthologs of PCDH19 from various species, showing the high level of conservation of the residues affected by the two missense mutations, p.V441E (top) and p.N557K (bottom) illustrated in the teal boxes. p.Val441 is highly conserved and in close proximity to the calcium-binding acidic residues (illustrated in brackets). p.Asn557 is invariant and one of the essential residues for calcium ion binding (Morishita et al., 2006, Patel et al., 2006).

Even though the two missense mutations had been predicted to be harmful we needed further evidence in favour of their pathogenic nature. To further confirm the importance of these two residues I aligned the six EC repeat domains of PCDH19 with EC1 of classical cadherins (C, E and N) and PCDHα4. The crystal structure of the EC1 domain of classical cadherins (C, E and N) (Patel et al., 2006) and PCDHα4 (Morishita et al., 2006) have been determined along with the residues involved in forming the adhesive interface and calcium binding regions (Patel et al., 2006, Morishita et al., 2006) (figure 3.13). Figure 3.13 shows the alignment of the six EC repeat domains of PCDH19 to human and mouse Pcdhα4. Mouse Pcdhα4 is
included in addition to human PCDHα4 as Morishita et al. (Morishita et al., 2006) annotate the adhesive interface and calcium binding regions of EC1 for the mouse protein. Analysis of the alignments along with the functional annotation revealed that both V441 and N557 are likely to be involved in calcium binding to the EC domain.

Figure 3.13 Extracellular cadherin domain repeat alignment

Clustal-W alignment of the six extracellular cadherin (EC) repeat domains of PCDH19, with the well characterised mouse and human PCDHα4 EC1. The residues shaded in grey are residues that differ from the consensus sequence. Residues shaded in a teal box are those that have been identified in mouse Pcdha4 EC1 to be involved in calcium binding (Morishita et al., 2006). Residues in a purple box are residues that have been identified to be involved in the adhesive interface. The two amino acids affected by the two missense mutations identified in EFMR families are circled in red, with the mutation annotation written below the alignment.
Valine 441 of EC4 of PCDH19 is equivalent to Isoleucine at p.96 in EC1 of C-, E- and N-Cadherin (Morishita et al., 2006, Patel et al., 2006) or p.Val95 of EC1 of CNR/Pcdhα 4 (Morishita et al., 2006). The Val residue is conserved both across the six EC domains of PCDH19 (figure 3.13) and across various protocadherins and protocadherins of various species (figure 3.12). Whilst the Val residue is not conserved in the classical cadherins, a highly conserved Isoleucine residue substitutes it (figure 3.14). Val and Ile are both non-polar, neutral pH amino acids with similar hydropathy index scores or 4.2 and 4.5 respectively. The missense mutation p.V441E is therefore not likely to be tolerated as the substituted amino acid, glutamic acid (E), is polar, has a negative pH and a hydropathy index score of -3.5.

Figure 3.14 Alignment of PCDH19 EC4 and 5 with the well characterised EC1 domain of classical cadherins and PCDHα4

Clustal-W alignment of the EC1 domain of: classical cadherins (C, E and N) and protocadherin α4; with the EC4 and EC5 domains of PCDH19, which contain the domains mutated in the missense mutations p.V441E and p.N557K respectively. Arrows indicate residues involved in the formation of the βF and βG sheets, navy for classic cadherins and teal for protocadherins (Morishita et al., 2006, Patel et al., 2006). Residues involved in calcium binding are illustrated in purple (Morishita et al., 2006, Patel et al., 2006). Residues shaded in black are those that differ from the consensus sequence. Residues p.V441 and p.N557 which are mutated in EFMR families are illustrated in red and circled, with the mutations annotated below the alignment. Residue p.N557 is in the calcium binding site while p.V441 is in close proximity to the calcium binding site.

The second amino acid change identified, p.N557K, is also predicted to disrupt calcium binding. This mutation affects an invariant asparagine (Asn; N) residue (figure 3.13).
Alignment of EC5 of PCDH19, to EC1 of classical cadherins and CNR/Pcdhα4, reveals that p.N557 is equivalent to Asn102 in EC1 of C-, E- and N-Cadherin (Morishita et al., 2006, Patel et al., 2006) and Asn101 of EC1 of CNR/Pcdhα4 (Morishita et al., 2006) (figure 3.14). Whilst this mutation does not directly interfere with an amino acid in the calcium binding domain it is likely to disrupt an amino acid in the βG strand, which is in close proximity to the calcium binding domain (as illustrated in figure 3.15).

![Figure 3.15 Overview of CNR/Pcdhα4 EC1 domain structure](image)

**Figure 3.15 Overview of CNR/Pcdhα4 EC1 domain structure**

These images were taken from Morishita et al. (Morishita et al., 2006). **A**, A ribbon diagram of the representative NMR structure of EC1 of CNR/Pcdhα4. Illustrated in a purple box is residue p.V95 of EC1 that is equivalent to p.V441 of EC4 which is mutated in PCDH19 in one of the EFMR families. **B**, Greek key topology of the EC1 domain of CNR/Pcdhα4. The arrow indicates the approximate position of residue p.V95.

**3.3.6 PCDH19 expression**

With the identification of the causative mutations in PCDH19 in EFMR families we aimed to enhance our understanding of the expression pattern of PCDH19 and subsequently the function of PCDH19. It was important to demonstrate, with respect to EFMR, that PCDH19 is expressed in the brain, in which regions and at which developmental stages. To investigate the expression pattern of PCDH19 we used northern blot analysis, in-situ hybridisation, and RT-PCR.
Initially, we aimed to identify which specific brain regions expressed PCDH19. This information was sought as it could help to narrow down regions of the brain requiring in-depth imaging investigation in the EFMR patients.

To investigate PCDH19 expression in the brain we performed northern blot analysis of multiple human brain regions using a PCDH19 mRNA specific probe. The probe used was a 374 bp fragment of exon 6 encompassing nucleotides c.2884-3257 of the human PCDH19 ORF (EF676096). Northern blot analysis identified a ~ 9.8 kb transcript as expected, based on our annotation of the full length PCDH19 mRNA sequence of 9754 bp (Genebank accession AK355593). The expression was evident in cerebellum, cerebral cortex, occipital pole, frontal lobe, temporal lobe, amygdala, hippocampus and brain (whole) (figure 3.16). The probe also detected additional mRNA species of; ~6.75 kb, ~4.5 kb and ~4.2 kb. We believe these were either non-specific mRNAs or PCDH19 mRNA degradation products. There is no evidence to support the presence of alternative PCDH19 isoforms of these sizes.

![Figure 3.16 Tissue expression analysis of PCDH19 RNA](image)

RNA blot analysis of PCDH19 in various human brain tissues. Asterisks indicate the PCDH19 transcript of ~9.8 kb. Brackets indicate non-specific binding of the PCDH19 probe and/or PCDH19 degradation products.

Following the northern blot results we selected five specific brain regions cerebral cortex, occipital pole, frontal lobe, hippocampus and brain (whole), for quantitative real time PCR
(RT-qPCR) (figure 3.17). RT-qPCR analysis validated the northern blot results confirming the highest level of PCDH19 expression in the frontal lobe.

![Figure 3.17 PCDH19 expressions in a range of brain regions](image)

**Figure 3.17 PCDH19 expressions in a range of brain regions**

Expression of PCDH19 was assessed in human brain regions through RT-qPCR using the standard curve method and normalising expression levels to the control gene Actin-β.

### 3.3.6.2 Pcdh19 in situ hybridisation results

The study of PCDH19 expression was further extended to mouse in-situ hybridisation analysis of the developing mouse central nervous system (CNS). Previous mouse in situ analysis of Pcdh19 expression identified a tissue specific pattern of expression during embryogenesis that was unique to Pcdh19. We concentrated our analysis on the developing central nervous system.

The candidate generated a Pcdh19 in-situ probe encompassing 986 bp of exon 1. A digoxigenin-labeled anti-sense RNA probe was then generated from this construct and used to label embryonic day 15.5 embryonic heads and dissected P2 brain sections (figure 3.18), this
work was conducted and analysed by our colleagues (P. Thomas and E. Sutton, School of Molecular & Biomedical Science, University of Adelaide).

*Pcdh19* expression was found to be widespread in both embryonic and adult brain including the developing cortex and hippocampus. Specifically *Pcdh19* is expressed in a widespread asymmetrical pattern in the embryonic forebrain and frequently localised to discrete cell clusters within the cortex, thalamus and hypothalamus (figure 3.18A, B indicated by arrowheads). In the cortex expression was restricted to the cortical plate (figure 3.18C) and extended medially into the intercerebral fissure (figure 3.18D). Robust expression was also detected in the ganglionic eminence that abuts the dorsolateral wall of the lateral ventricles (figure 3.18E, asterix). At this stage hippocampal expression was not observed on the medial edge of the lateral ventricle in the presumptive hippocampus (figure 3.18B and E). Analysis of anterior forebrain sections revealed *Pcdh19* expression in the epithelial lining of the nasal cavity (consistent with the previous report (Gaitan and Bouchard, 2006)) and in the olfactory bulbs (figure 3.18F). At postnatal day 2, *Pcdh19* expression was maintained in discrete regions of the cortex and the thalamus however, unlike the embryonic brain, expression was also observed in the hippocampus (figure 3.18G-I). In the cortex, expression was restricted to a band of cells that spanned layers II-IV (Figure 3.18J and K, arrows) whilst the most prominent *Pcdh19* signal was observed in the CA1 and CA3 regions of the hippocampus (figure 3.18 H and I). *Pcdh19* transcripts were not detected in white matter tracts including the corpus callosum (figure 3.18H).
Figure 3.18 Expression of Pcdh19 in the developing mouse CNS.
Expression at embryonic day 15.5 (a-f) and postnatal day 2 (g-l) (representative of two males and two females studied). (a,b) Adjacent coronal sections through the hippocampal region stained with hematoxylin and eosin or processed for Pcdh19 in situ, respectively. (c-e) Higher magnification images of the boxed regions in b. Arroceheads in c, Pcdh19 expressing cells within the cortex; asterix in e, dorsolateral wall of the lateral ventricle. (f) Coronal section through the olfactory bulb highlighting Pcdh19 expression in the nasal epithelium. (g,h) Adjacent coronal sections through the mid-hippocampal region stained with hematoxylin and eosin or processed for Pcdh19 in situ, respectively. (i) A brain section more posterior than that in h, highlighting Pcdh19 expression. (j-l) Higher-magnification images of the regions boxed in g and h, as indicated. Arrows in j,k, Pcdh19 expression within cortical layers II-IV. Cx, cortex; CxP, cortical plate; Hn, hippocampal neuroepithelium; lv, lateral ventricle; Th, thalamus; Hy, hypothalamus; icf, intercerebral fissure; Ob, olfactory bulbs; Ne, nasal epithelium. Scale bars in a,b,f-l, 200 μM; in c-e, j-l, 50 μM.
3.3.6.3 *PCDH19* tissue expression – RT-PCR

*PCDH19* expression in additional human adult tissues was analysed by the candidate by RT-PCR. Expression of *PCDH19* was identified in all tissues analysed (figure 3.19), with the highest level of expression identified in total brain, foetal brain, cerebellum and testis. Relatively low levels of expression were identified in adrenal gland, fibroblast, spleen and lung.

![Figure 3.19 Tissue expression of *PCDH19*](image)

Agarose gel results of PCR amplification of *PCDH19* in cDNA generated from a range of human tissues. PCR amplification of Actin-β was used to assess the quality of the tissue cDNA.

The presence of *PCDH19* expression in skin fibroblast cells, lead us to obtain skin punches from EFMR patient and control individuals and generate fibroblast cultures. Whilst the level of expression of *PCDH19* in fibroblast cells is low we postulated that this cell type will be useful in further molecular analyses because it is an adherent cell type which can be obtained from living patients.

3.3.7 EFMR biological mechanism

Ryan et al have put forward four suggestions to explain the EFMR inheritance pattern: a Y-chromosome homologue that protects males, metabolic interference, a defect in X-
inactivation or that the defective gene in EFMR is not required for the development of the male brain (Ryan et al., 1997). Now that we have identified the PCDH19 gene we set out to re-assess the validity of these hypotheses.

3.3.7.1 Metabolic interference

This mechanism was put forward because it can only occur in individuals expressing two alleles of a given locus, in this case females with two X chromosomes. So far all of the EFMR females we have identified with mutations in PCDH19 are heterozygous for PCDH19 mutations.

In EFMR the combination of normal and mutant protein is unlikely to occur in a single cell unless PCDH19 escapes X-inactivation. That is why we set out to determine whether PCDH19 is or is not subject to X-inactivation. As we demonstrate, PCDH19 is subject to X-inactivation (see section 3.3.6.2.1). In this case every cell in a female only expresses a single copy of PCDH19. The required combination of normal and mutant protein can still be achieved, as PCDH19 is a transmembrane protein, which is believed to function thorough homophilic binding of the extracellular EC domains from adjacent cells. For this to happen the mutant protein ought to be expressed and incorporated into the membrane.

We have generated three peptide specific sheep polyclonal human PCDH19 specific antibodies. Different C and N-terminal regions were selected to facilitate the identification of truncated PCDH19 (figure 3.20). We also purchased two commercial PCDH19 antibodies (figure 3.20). Initially we studied fibroblast cells from an affected individual (Family 2, V-7, figure 3.7) with the p.Q85X mutation, which we predicted due to its small size would not be maintained in the cells. Unfortunately all attempts to use these antibodies on patient fibroblasts in immunofluorescence or western blot analysis failed likely due to the low level
of expression of PCDH19. We determined that the antibodies were PCDH19 specific by performing transient transfections and over expressing V5-tagged PCDH19 in fibroblast cells (figure 3.21).

Figure 3.20 Schematic representation of the PCDH19 antibodies

This schematic illustrates the five PCDH19 antibodies we had available for use, named: 201, 205, Abnova, Atlas and 209. Antibodies depicted in purple are the three antibodies we generated; the two antibodies depicted in navy were purchased commercially. The PCDH19 protein domain which each antibody detects is demonstrated by the coloured protein fragment beneath each antibody. The position of the amino acid residues that comprised the immunogenic sequence recognised by the antibody, are stated for each of the 5 antibodies. The two truncation mutations that we had fibroblast cell lines for are illustrated below the protein schematic.
Figure 3.21 Western blot analysis of PCDH19 expression

Western blot analysis in combination with over expression of PCDH19 achieved through transient transfection was used to test the specificity of the PCDH19 antibodies. Each panel illustrates the results obtained from a antibody, the antibody used is stated below each panel. PCDH19 transfected refers to 293T cells transiently transfected with full length PCDH19. The asterisk highlights the 126kD PCDH19 transcript. PCDH19 expression was not detected in 293T cells or fibroblasts, and could only be visualised in the transfected cells.

Failing to detect the wild type and truncated PCDH19 protein we decided to focus on PCDH19 mRNA with the aim to investigate the stability of mRNA containing the premature termination mutations. Premature termination codon (PTC) containing mRNAs, are typically recognised by the nonsense mediated mRNA decay (NMD) surveillance complex and efficiently degraded. Figure 3.22 briefly describes the process of NMD. By eliminating the PTC-containing mRNA, NMD functions to prevent the production of potentially deleterious truncated proteins (reviewed in Maquat, 2004). We aimed to assess if the PTC containing mRNA transcript with the c.253C>T (p.Q85X) mutation was subject to NMD.
Chapter 3

Figure 3.22 Nonsense mediated mRNA decay (NMD)

Nonsense mediated decay is a process that recognises the presence of PTC and degrades the mutant mRNA ensuring that no truncated and potentially harmful proteins are produced. This is an illustration of how the NMD surveillance and degradation mechanism is believed to function. The left panel depicts the scenario that occurs in normal transcripts that are devoid of a premature termination codon (PTC). In the spliceosome (top row) exon junction complexes (EJC) are deposited at the exon junctions as the pre-mRNA is spliced together (second row). During the primary round of translation these EJC are displaced from the mRNA by the translating ribosome and translation is allowed to proceed to the end of the mRNA (third panel) and the encoded protein is produced (bottom panel). The right panel illustrates how NMD works in the presence of a PTC (as illustrated here by a pair of scissors). In this scenario the translating ribosome stops at the PTC (third panel) and therefore does not displace all of the EJCs, the NMD surveillance system detects the presence of these EJCs and initiates the degradation of the PTC containing mRNA, resulting in no protein being produced (bottom panel). Figure adapted from (Holbrook et al., 2004).

I carried out sequence-based analysis of both the genomic DNA (gDNA) and complementary DNA (cDNA) obtained from patient fibroblasts (Family 2, VI-7 figure 3.7). Sequence analysis of the gDNA showed, as expected, heterozygosity for the c.253C>T (p.Q85X) mutation (figure 3.23). However, the cDNA showed only the wild-type c.253C allele. This suggested that the mutant c.253C>T allele is degraded by NMD.
Figure 3.23 mRNA containing the c.253V>T (p.Q85X) mutation in *PCDH19* is subject to nonsense-mediated RNA decay

Nonsense mediated RNA decay of mutant *PCDH19* transcripts. Sequence chromatogram from EFMR-affected female from family 2 showing the detection of the mutation c.253C>T in genomic DNA (gDNA) (top), the absence of the mutant sequence of the mutant sequence in fibroblast complementary DNA (cDNA) (middle) and the presence of both mutant and wild-type cDNA after treatment of fibroblasts with cyclohexamide (bottom), which inhibits the pioneer round of translation and thus NMD. The position of the mutation is boxed.

To confirm NMD degradation of the c.253C>T allele, we treated the fibroblast cells with 50 μg/ml of cyclohexamide for 6 hrs. Cycloheximide is a known inhibitor of translation (Carter et al., 1995), thereby inhibiting NMD. Sequence analysis of the patient cDNA treated with cycloheximide identified both alleles, thereby confirming that the c.253C>T (p.Q85X) mutation is subject to NMD (figure 3.23).
We also assessed the expression levels of *PCDH19* using RT-qPCR. We wanted to assess the effect of NMD with regard to the level of *PCDH19* expression in this individual's fibroblasts and how the expression level changes with the inhibition of NMD. The RT-qPCR results (figure 3.24) identified that after cycloheximide treatment there was an increase in the expression of *PCDH19* by 4.8 fold (P-value 0.007). This result further supports our conclusion that this mutation is subject to NMD.

![Figure 3.24 PCDH19 expression in response to cyclohexamide treatment](image)

**Figure 3.24 PCDH19 expression in response to cyclohexamide treatment**

Graphical representation of *PCDH19* expression in fibroblast cells from a control (purple) and from a patient with the p.Q85X mutation (teal), under normal growth conditions and with the addition of cyclohexamide (an inhibitor of NMD). *PCDH19* expression was assessed by using RT-qPCR and the standard curve method, expression levels were normalised to the control gene Actin-β.

We also had access to fibroblast cell lines from one of the Israeli families (Family 3 figure 3.3) with the c.2012C>G (p.S671X) protein truncation mutation Here we were fortunate to have fibroblast cell lines from two affected females (Family 3, VI-1 and VI-2, figure 3.3) and an unaffected transmitting male (Family 3, V-6, figure 3.3). Given the results we had obtained
for the p.Q85X mutation demonstrating that it was subject to NMD we wished to assess if the 
p.S671X mutation was also subject to NMD.

Initial semi-quantitative RT-PCR indicated that the mutant PTC containing c.2012C>G 
mRNA is subject to NMD as there were very low levels present in the transmitting male 
(figure 3.25A). Subsequent sequence analysis of gDNA and cDNA of two affected females 
identified that both were heterozygous for the c.2012C>G mutation in their gDNA. The 
mutant G allele was also present in the cDNA of both females; however, it was at a 
considerably lower level than in the gDNA (figure 3.25). These experiments suggested that 
the c.2012C>G mutation is also subject to NMD; however, the degradation of its mRNA is 
not as efficient as that of the c.253C>T (p.Q85X) mutation.
Figure 3.25 mRNA containing the c.2012C>G (p.S671X) mutation in PCDH19 is subject to nonsense-mediated RNA decay

Nonsense mediated RNA decay of mutant PCDH19 transcripts. A, Semi quantitative PCR amplification of the PCDH19 amplicon containing the p.S671X mutation, as illustrated on an agarose gel, from gDNA and cDNA from a transmitting male and two affected females. Demonstrating substantially reduced levels of product in the cDNA of the carrier male due to NMD of the mutant transcript. B, Sequence chromatogram from two EFMR-affected females from family 3 showing the detection of the mutation c.2012C>G in genomic DNA (gDNA) (top), the absence of the mutant sequence of the mutant sequence in fibroblast complementary DNA (cDNA) (bottom). The position of the mutation is boxed. Sequence chromatograms from each affected females are illustrated either side of the vertical line. The position of the p.S671X mutation c.2012C/T is boxed on each sequence chromatogram and indicated by arrows.

The NMD driven degradation of the PTC containing PCDH19 implies that the mutated protein is not produced in sufficient amounts in these cells. Therefore a dominant-negative effect becomes an unlikely explanation for the molecular mechanism underlying EFMR. A dominant negative mechanism remains possible in the case of females with missense mutations in PCDH19. These mutations do not have premature termination codons and
therefore will not be subject to NMD. The possibility remains that the mutant PCDH19 expressed on one cell could interact with and inhibit the function of the wild type PCDH19. Given the high degree of similarity in the inheritance pattern and phenotype seen in patients with non-sense and missense mutations in PCDH19 we believe the same molecular mechanism will be responsible for EFMR in both mutation types making a dominant negative effect an unlikely explanation for the molecular mechanism underlying EFMR.

There is another potential explanation for the absence of the mutant transcript that has to be considered before the hypothesis of metabolic interference can be conclusively ruled out. Skewed X-inactivation could account for the absence of an allele of PCDH19 in patient cDNA. Whilst X-inactivation cannot account for the results obtained with cycloheximide treatment of the p.Q85X containing fibroblasts, it still needs to be ruled out as a possibility, especially with respect to the p.S671X mutation (see section 3.3.6.2).

3.3.7.1.1 PCDH19 expression in EFMR females

Intriguingly, the PCDH19 mRNA containing the c.253C>T mutation had a 5 fold (P-value 0.002) higher expression when compared to the control fibroblast cell line(s) analysed. Given that we know this individual has random X-inactivation (see section 3.3.6.2.2) we would predict that approximately half of her cells will express wild type PCDH19 and the other half will express the mutant PCDH19. We know that the mutant transcript is efficiently degraded by NMD (see section 3.3.6.1) and therefore approximately half of the cells will be devoid of PCDH19. We were therefore expecting to see a much lower level of expression of PCDH19 in this individual with regard to a control.

To investigate this further, we assessed PCDH19 expression in a number of individuals with various PCDH19 mutations and additional controls. RT-qPCR revealed that 2 out of the 9
individuals with *PCDH19* mutations had very high levels of PCDH19 expression, whilst the remaining seven individuals had similar levels to controls (figure 3.26). There is no obvious relationship between the two individuals with the highest levels of *PCDH19* expression. They have different mutations and different types of mutations, one with a protein truncation and one with a missense mutation. There was also a significant amount of variation in *PCDH19* expression present between different individuals from the same family with the same mutation (figure 3.26). Concluding from this, it appears that *PCDH19* expression in the fibroblast cell lines is highly variable and also given small numbers of patients and controls no meaningful correlations could be drawn. The expression level of *PCDH19* in an individual did however appear to be robust as similar expression levels were observed over replicate experiments involving independent cell extractions from different cell cultures. The robust highly variable expression identified between individuals could be the result of trans acting modifiers; however, this remains to be proven.

The assessment of the respective PCDH19 protein levels was not achievable due to low expression (as described in section 3.3.6.1).
Figure 3.26 *PCDH19* expression in patient and control fibroblasts

Graphical representation of *PCDH19* expression in fibroblast cells of nine individuals with *PCDH19* mutations (as defined on the X-axis) and two controls (navy). The individuals with *PCDH19* mutations are split according to the type of mutation they have, with teal bars for protein truncation mutations and purple bars for missense mutations. Error bars represent the standard error obtained.

3.3.7.2 Defect(s) of X-inactivation

One of the previous hypotheses to explain the unusual EFMR inheritance, suggested that, a defect in X-inactivation resulting in females having “functional dichotomy”, whereby in EFMR females there is a genetic defect that affects normal X-inactivation such that genes that are normally subject to X-inactivation fail to inactivate correctly leading to over-expressing of X-linked genes, could cause the EFMR phenotype (Ryan et al., 1997).

We have identified (see section 3.3.6.1) that one allele (the wild type allele) was predominantly (p.S671X, figure 3.25) and exclusively (p.Q85X, figure 3.23) present in the cDNA of heterozygous females from two EFMR families with two different protein truncation mutations. While this was interpreted as mutant RNA degradation by NMD, the
exclusive or predominant expression of only one allele could also have occurred as a result of skewed X-inactivation, whereby the wild type allele was preferentially expressed. In order to rule out skewed X-inactivation as an explanation for EFMR we had to demonstrate that *PCDH19* is subject to X-inactivation and that the affected females are not subjected to preferential X-inactivation of their normal allele.

### 3.3.7.2.1 Is *PCDH19* subject to X-inactivation?

*PCDH19*’s chromosomal location at Xq22.1 places it near, but not within the human specific region of X-Y homology. Genes located within this region tend to escape X-inactivation, in order to maintain similar gene dosage between females and males. Whilst there has been a lot of work done on the X-inactivation status of genes within the X-Y homology region, not much is known about the genes that flank this region, including *PCDH19*.

To assess whether *PCDH19* is subject to X-inactivation we isolated DNA and RNA from seven different female patients with balanced X;autosome translocations, 5 of these were extracted from fibroblast and 2 from lymphoblast cell lines. We took advantage of these patients as there was published evidence for most of them showing that the normal X-chromosome was late replicating and thus preferentially inactivated (Carrel and Willard, 1999, Gecz et al., 1999, Kalscheuer et al., 2003b). Typically in carriers of balanced X;autosome translocations the normal X-chromosome is preferentially inactivated to ensure proper dosage compensation of X-linked genes (as illustrated in figure 3.27). Alternatively in other cases, such as unbalanced translocations involving deletions or duplications, the abnormal X-chromosome portions are inactivated to ensure correct dosage is achieved from the active normal X-chromosome (as illustrated in figure 3.27) (Carrel and Willard, 1999). Therefore, if *PCDH19* is subject to X-inactivation, we should see only one allele of *PCDH19* expressed.
Figure 3.27 Schematic representation of X-inactivation in the presence of balanced and unbalanced X-chromosome; autosome translocations

Purple illustrates two autosomal chromosome homologues and teal the two X-chromosome homologues. A dashed box illustrates the chromosomal regions that are silenced by X-inactivation. The central panel depicts the scenario that occurs when there is a balanced X;autosome translocation, whereby no genetic information is lost or duplicated. Here the normal X-chromosome is preferentially inactivated and silenced. The final panel depicts the scenario that occurs when there is an unbalanced translocation and genetic material is lost or gained. In this scenario the abnormal X-chromosome portions are silenced to ensure any gained sequence is also silenced, and that any lost sequence is not completely silenced.

To be able to do this we needed to identify informative, heterozygous single nucleotide polymorphisms (SNPs) within the PCDH19 mRNA in these X;autosome translocation patients. Such SNP(s) would then be used to assess how many alleles of PCDH19 are expressed. PCDH19 has low a level of nucleotide variation with only one common SNP, c.1627 C>T (p.L543L), which occurs in approximately 20-30% of individuals (see chapter 4 for details). Sequence analysis of gDNA from the X;autosome translocation patients identified that 2/7 were informative, as they were heterozygous for the c.1627 C>T (p.L543L) SNP in PCDH19. The two informative cell lines, GM4628 karyotype 46,X,t(X;22)(q12;p11) and GM00089 karyotype 46,X,t(X;19)(q22;q13.3) were previously investigated for similar purpose by Carrel et al (Carrel and Willard, 1999) as case number 68 and 51 respectively.
They found both of these patients exhibited late replication and silencing of one of their X-chromosomes. If \( PCDH19 \) is subject to X-inactivation one of the alleles of \( PCDH19 \) should be inactivated and the individual’s cells should therefore be hemizygous for the SNP. Sequencing of cDNA from these two informative patients cell lines showed that both patients only expressed one allele, c.1627T (figure 3.28). Given that these females have only one of their X-chromosomes active and only one of their two \( PCDH19 \) alleles were expressed in their cDNA, then \( PCDH19 \) is subject to X-inactivation.

![Figure 3.28 PCDH19 is subject to X-inactivation.](image)

Sequence chromatogram from two X;autosome translocation patients showing the detection of the heterozygous SNP c.1627C/T in gDNA (top) and an absence of the C allele cDNA (bottom) in both patients. Sequence chromatograms from each translocation female are illustrated either side of the vertical line. The position of the c.1327C/T SNP is boxed on each sequence chromatogram and indicated by arrows.

3.3.7.2.2 X-inactivation studies of females from EFMR families

Once we had proven that \( PCDH19 \) is subject to X-inactivation we assessed the pattern of X-inactivation in our EFMR females. Previous X-inactivation studies conducted by Ryan et al on leucocytes of the original EFMR family indicated that the proportion of subjects with
predominant X-skewing did not differ significantly between EFMR females and controls (Ryan et al., 1997). Therefore there is no evidence to dispute that females with the c.1091_1092insC (p.Y366LfsX10) mutation have normal X-inactivation.

We extended this analysis to four further EFMR families (Families 1-4, figure 3.3 and 3.7). We performed a standard X-chromosome inactivation assay involving methylation sensitive cleavage of gDNA followed by PCR amplification of highly polymorphic regions of the X-chromosome (Contributed by Melissa M. Thovin et al., 2002). We were unable to directly assess the X-inactivation status of *PCDH19* itself as there was no suitable polymorphism we could use. Instead we tested the X-inactivation through two X-chromosome markers; a polymorphic (CAG)$_n$ region flanking the fragile(X) gene locus *FRAXA*, and the highly polymorphic (CAG)$_n$ region of the 5’ end of the coding region of the androgen receptor gene *AR*. This analysis (table 3.2) found two affected females (one with the p.V441E and one with the p.L677fsX41 mutation) and one unaffected female who showed skewed X-inactivation at the *FRAXA* locus. No individuals demonstrated skewed X-inactivation at the *AR* locus. Overall the results demonstrate that there is no significant skewing of X-inactivation among affected or unaffected members of the EFMR families tested. Whilst there were two affected females with skewed X-inactivation, there was no overall trend towards abnormal skewing in the affected individuals. One of the females skewed at the *FRAXA* locus did not exhibit skewing at the *AR* locus. A third locus would need to be tested in this individual to ascertain if she does actually have skewed X-inactivation. This was not performed due to the lack of overall trend towards skewing in the vast majority of affected females.
<table>
<thead>
<tr>
<th>Family</th>
<th>Phenotype</th>
<th>Mutation</th>
<th>FRAXA Result</th>
<th>AR Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>78/22</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>34/66</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>91/9</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>14/86</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>79/21</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>75/25</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>57/43</td>
<td>Non-informative</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>Non-Informative</td>
<td>71/29</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>95/5</td>
<td>67/33</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>Non-Informative</td>
<td>49/51</td>
</tr>
<tr>
<td>1</td>
<td>Affected</td>
<td>V441E</td>
<td>Non-Informative</td>
<td>49/51</td>
</tr>
<tr>
<td>2</td>
<td>Affected</td>
<td>Q85X</td>
<td>Non-Informative</td>
<td>43/57</td>
</tr>
<tr>
<td>2</td>
<td>Affected</td>
<td>Q85X</td>
<td>53/47</td>
<td>50/50</td>
</tr>
<tr>
<td>2</td>
<td>Affected</td>
<td>Q85X</td>
<td>Non-Informative</td>
<td>71/29</td>
</tr>
<tr>
<td>2</td>
<td>Affected</td>
<td>Q85X</td>
<td>68/32</td>
<td>54/46</td>
</tr>
<tr>
<td>2</td>
<td>Affected</td>
<td>Q85X</td>
<td>59/41</td>
<td>Non-informative</td>
</tr>
<tr>
<td>3</td>
<td>Unaffected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Affected</td>
<td>S671X</td>
<td>82/18</td>
<td>Non-informative</td>
</tr>
<tr>
<td>3</td>
<td>Affected</td>
<td>S671X</td>
<td>53/47</td>
<td>Non-informative</td>
</tr>
<tr>
<td>3</td>
<td>Affected</td>
<td>S671X</td>
<td>42/58</td>
<td>Non-informative</td>
</tr>
<tr>
<td>4</td>
<td>Unaffected</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>70/30</td>
<td>18/82</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>78/22</td>
<td>Non-informative</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>5/95</td>
<td>Non-informative</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>49/51</td>
<td>57/43</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>65/35</td>
<td>Non-informative</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>79/21</td>
<td>Non-informative</td>
</tr>
<tr>
<td>4</td>
<td>Unaffected</td>
<td></td>
<td>31/69</td>
<td>24/76</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>Non-Informative</td>
<td>55/45</td>
</tr>
<tr>
<td>4</td>
<td>Affected</td>
<td>L677FfsX41</td>
<td>Non-Informative</td>
<td>Non-informative</td>
</tr>
</tbody>
</table>

**Key**
- Skewed & affected
- Skewed & unaffected

### Table 3.2 X-inactivation status of individuals from four of the EFMR families

This table outlines the results obtained for the X-inactivation testing conducted on individuals from four of the EFMR families. The FRAX result and AR results are presented as a ratio. This ratio represents the proportion of gDNA that contains each allele. Non-informative results were obtained when individuals were homozygous for the given locus.

The identification of normal X-inactivation in the EFMR females from our four EFMR families further validates the results of Ryan *et al* in that EFMR females have normal X-inactivation (Ryan et al., 1997). Therefore we can eliminate defect(s) of X-inactivation as a possible mechanism to explain the EFMR inheritance pattern.
3.3.7.3 Y-chromosome rescue

A further speculation to explain the EFMR inheritance pattern suggested that a functional homologue of the disease locus exists on the Y-chromosome and is able to rescue males. There was initial support for this theory as the EFMR linkage interval encompasses a 4 Mb region of X/Y homology. However there is no direct Y chromosome paralog of \textit{PCDH19}, thereby eliminating this as a possible explanatory mechanism for the EFMR inheritance pattern. Whilst there is no paralog of \textit{PCDH19} there is a related non-paralogous protocadherin gene on the Y-chromosome, \textit{PCDH11Y}, which could possibly play a role in rescuing males from EFMR. Figure 3.29 illustrates the relationship between PCDH11X / 11Y and PCDH19, with respect to their δ-sub class.

![Figure 3.29 Characteristic domains of the δ Protocadherin family](image)

Clustal-W alignment of a region of the cytoplasmic domain of the δ-protocadherin family. With members of the δ1 subgroup indicated by the top bracket and the δ2 subgroup members in the bottom bracket. The characteristic domains of this family are illustrated; the CM1 domain in teal, CM2 domain in purple and the CM3 domain in navy. Residues shaded in black are those that do not match the consensus sequence.
3.3.7.3.1 *PCDH11X/Y* northern blot results

To investigate *PCDH11X/Y* expression in detail, and how its expression correlated with or differs from *PCDH19* we probed the human multiple tissue northern blots with a *PCDH11X/Y* specific probe.

Due to the high degree of similarity between *PCDH11X* and *11Y* we were unable to design probes specific for each gene, and instead had to settle for a probe which would detect both *PCDH11X* and *11Y*. Northern blot analysis identified an approximately 9.5 kb transcript. Expression of PCDH11X/Y was evident in; cerebral cortex, occipital pole, frontal lobe, temporal lobe, putamen, amygdala, and lowly expressed in the hippocampus and whole brain (figure 3.30). Comparison with *PCDH19* expression identified similarities as well as differences. Differences were evident in the cerebellum, where expression was restricted to *PCDH19*, and in the putamen, where expression was restricted to *PCDH11X/Y*. Apart from those differences the remaining overlaps in the expression between *PCDH19* and *PCDH11X/Y* suggested that the rescue hypothesis could be plausible.
Figure 3.30 Tissue expression analysis of PCDH19 RNA by Northern blot

RNA blot analysis of PCDH11X/Y in various human brain tissues. An arrow head indicates the ~9.5kb PCDH11X/Y transcript. The PCDH19 northernblot results are also demonstrated for ease of comparison.

3.3.7.3.2  PCDH19 and PCDH11X/Y brain region real time

To further investigate the plausibility of the PCDH11Y rescue hypothesis we selected five brain regions, based on the northern blot analysis, to investigate further by RT-qPCR. Here we were able to distinguish between PCDH11X and 11Y expression. To distinguish between PCDH11X and 11Y we took advantage of a 13 bp deletion that is present in PCDH11Y with respect to PCDH11X. PCDH11X and 11Y forward primers were designed over this deletion region (illustrated in figure 3.31) and the reverse primers were designed with a mismatching base between PCDH11X and 11Y at the 3’ end of the primer. Sequence analysis of PCR products was used to validate the specificity of the primers.
Figure 3.31 Genomic structure of PCDH11X and 11Y

An ideogram of the genomic structure of PCDH11X (top ideogram) and PCDH11Y (second ideogram). The 5’ UTRs are illustrated in teal, coding exons in purple and 3’ UTR in navy. Illustrated in the zoomed in region is the PCDH11X and PCDH11Y specific primers used to assess expression levels of these two genes by RT-qPCR. PCDH11X and PCDH11Y sequences are illustrated in navy and purple respectively. The primer sequences are illustrated in the boxes.

RT-qPCR results (figure 3.32) identified increased expression of PCDH11Y with respect to 11X in the hippocampus and increased expression of PCDH11X with respect to 11Y in the cerebral cortex. The remaining regions analysed had similar expression levels between PCDH11X and 11Y.
Expression of these protocadherins was assessed in human brain regions through RT-qPCR using the standard curve method and normalising expression levels to the control gene Actin-β. Navy bars represent PCDH11X expression, purple bars represent PCDH11Y and teal bars PCDH19.

We also compared the expression levels of PCDH11X and 11Y with PCDH19. Here we identified that PCDH11X and 11Y have higher expression in the frontal lobe than PCDH19. There were also differences present in the cerebral cortex and the hippocampus, where PCDH11X and PCDH19 had similar expression levels, whilst PCDH11Y had a lower and higher expression level respectively. These results support the notion that PCDH11X and 11Y do have differential expression patterns even with their high degree of similarity. However there appears to be a higher degree of similarity between PCDH11X and PCDH19 than between PCDH11Y and PCDH19. If PCDH11Y is capable of rescuing mutations in PCDH19 but PCDH11X is not due to expression level differences, we would expect a higher degree of correlation between PCDH19 expression and PCDH11Y. However the lack of correlation in expression could be because we are not looking in the specific brain regions or at the specific
developmental time point that are involved in the EFMR phenotype. Therefore rescue by the non-paralogous \( PCDH11Y \) could still occur.

3.3.7.3.3 \textit{In situ} hybridizations

Our colleagues (J. Bomar and D. H. Geshwind, Neurology Department and Semel Institute for Neuroscience and Behaviour, David Geffen School of Medicine), through \textit{in situ} hybridisation analysis of the human brain, further investigated the differential expression of \( PCDH11X, 11Y \), and \( PCDH19 \). They confirmed an overlap in the expression pattern across \( PCDH11X/Y \) and \( PCDH19 \), with all three genes expressed in developing cortical plate, amygdala and subcortical regions and in the ganglionic eminence. Closer inspection of the amygdala revealed \( PCDH19 \) was expressed in lateral nuclei, whereas \( PCDH11X/Y \) was more medial. Of importance with respect to the Y-chromosome rescue theory was the identification of sexually dimorphic expression of \( PCDH11X/Y \) in the caudate nucleus. Once again, specific \( PCDH11X \) and \( PCDH11Y \) probes were unable to be generated due to the high degree of similarity between the two genes. However, analysis using a \( PCDH11X/Y \) probe across male and female brains was used to assess differences in expression between \( PCDH11X \) expressing females and \( PCDH11X/Y \) expressing males. \( PCDH11X/Y \) expression was high in the caudate nucleus of females but virtually absent in males. The overlapping expression between \( PCDH11X/Y \) and \( PCDH19 \) in combination with the sexually dimorphic expression of \( PCDH11X/Y \) in the human brain, suggests that \( PCDH11Y \) could possibly function as a compensatory mechanism in males.
3.4 Discussion

3.4.1 Mutation identification

We have identified unique nucleotide changes in *PCDH19* in all six previously identified EFMR families (Ryan et al., 1997, Scheffer et al., 2008). We also identified a further unique nucleotide change in a screen of a cohort of 86 females with isolated or familial cases of epilepsy with or without intellectual disability (discussed further in chapter 4). Out of these seven nucleotide changes five are predicted to result in premature termination of the PCDH19 protein while the remaining two result in single amino acid substitutions.

Identification of the NMR structure of CNR/Pcdhα4 EC1 allowed us to investigate the likely consequence of the missense mutations we had identified in *PCDH19*. Protocadherins are believed to be composed of two β sheets packed face to face. One β sheet is composed of four β strands (βA, βC, βF, and βG) and the second β sheet is composed of the remaining three β strands (βB, βD, and βE) (Morishita et al., 2006, Patel et al., 2006). The p.Val95 residue in Pcdhα4 EC1 that is equivalent to p.Val441 in EC4 of PCDH19, is situated within the βG strand (figure 3.15). The introduction of a Glutamic acid in this position may disrupt the formation of the βG strand. Given that the βG strand is packed in between the βA and βF strands the introduction of the glutamic acid could also disrupt the overall formation of the β sheet. In addition to disrupting correct protein formation, the p.V441E mutation is also likely to disrupt calcium binding. Residue p.V441 in EC4 of PCDH19 is situated in close proximity to the highly conserved calcium-binding domain (figure 3.13 and 3.14). The missense mutation p.V441E could therefore disrupt calcium binding and thus the adhesive function of the EC4 domain.
The second missense mutation identified, p.N557K, is also predicted to disrupt calcium binding. The equivalent residue EC1 of Pcdhα4 and a number of cadherins has been identified as an important amino acid in the largest of the calcium binding regions identified (Morishita et al., 2006, Patel et al., 2006). The introduction of the positively charged amino acid, Lysine, in to the calcium-binding region is likely to inhibit binding of positively charged calcium. We predict this mutation will also result in a complete loss of function of PCDH19.

The two missense mutations identified (p.V441E and p.N557K) in PCDH19 affect highly conserved residues in close proximity to the highly conserved calcium binding domain. Whilst the EC1 domain of classical cadherins was originally thought to be entirely responsible for the homophilic adhesive activity of cadherins, there is increasing evidence highlighting the importance of other EC domains (Boggon et al., 2002, Chien et al., 2008, Tsuiji et al., 2007). Clinically, there are no discernable differences between affected females with protein truncation mutations and missense mutations. We hypothesise that these missense changes result in functional knockout of PCDH19 protein. We currently do not know what the phenotypic effect, if there will be any, that would result from less severe mutations. This will only be elucidated with continued mutation screening of \textit{PCDH19} in additional patient cohorts.

In total we have identified 5 protein truncation and 2 missense amino acids changes in \textit{PCDH19} in 7 unrelated EFMR families. We have presented considerable evidence in support of these nucleotide changes being highly deleterious to \textit{PCDH19} function (Appendix A, Dibbens et al., 2008).
3.4.2 Expression

*PCDH19* is a member of the cadherin family of calcium-dependant cell-cell adhesion molecules. Protocadherins are predominantly expressed in the nervous system where they are believed to dictate the survival and synaptic organisation of different neuronal populations (Morishita and Yagi, 2007). Whilst a significant amount of work has been conducted in an attempt to understand the functioning of protocadherins it has mainly centred on the clustered protocadherin sub-family of protocadherins. Consequently at the time these experiments were conducted relatively little was known about *PCDH19*, in particular its brain expression pattern.

Non-clustered δ-PCDHs (including *PCDH19*) are generally expressed in a variety of additional tissues (Hirano et al., 1999, Kuroda et al., 2002, Makarenkova et al., 2005, Murakami et al., 2006, Wolverton and Lalande, 2001, Yamamoto et al., 2000, Yoshida and Sugano, 1999). Expression of *PCDH19* has previously been identified by RT-PCR in human; brain, kidney, lung, heart and trachea (Wolverton and Lalande, 2001). We also identified expression of *PCDH19* in all brain regions analysed, skeletal muscle, spleen, testes, uterus, small intestines, placenta, saliva and fibroblasts. Confirming that *PCDH19* is expressed in the nervous system and a variety of additional tissues, this expression profile is characteristic of the non-clustered δ-PCDHs.

*PCDH19* expression has also been assessed by mouse in situ hybridization assessing *PCDH19* expression in the developing embryo (Gaitan and Bouchard, 2006). *PCDH19* expression was detected in day 9 - 12.5 embryos in the mesonephros and in the neuroepithelium of the forebrain and midbrain. Expression was also identified later in development in additional neuronal tissues; including the neuronal retina, nasal epithelium and spinal cord and in non-neuronal tissues; the collecting duct and differentiating nephrons.
of the metanephros in the glandular stomach, exocrine pancreas and hair follicles (Gaitan and Bouchard, 2006).

The highest level of *PCDH19* expression identified by northern blot and confirmed by RT-qPCR was the frontal lobe. This fits well with the identification by Ryan et al of defects in the cortical architecture of the frontal lobe in an EFMR patient (Ryan et al., 1997). A frontal lobe surgical specimen was obtained from an affected EFMR female and analysed using neurofilament protein immunohistochemistry. Results revealed a disruption of the normal cortical architecture with dysplastic neurons in the cortex and underlying white matter, as well as abnormal morphology of individual cortical neurons (Ryan et al., 1997). The high level of *PCDH19* expression and defects in the frontal lobe of an affected EFMR female suggested a possible role for *PCDH19* in ensuring proper cortical architecture formation in the frontal lobe. Interestingly, with respect to the suggested role for *PCDH11X/Y* in compensating for mutations in *PCDH19*, expression of *PCDH11X* and *PCDH11Y* were higher than *PCDH19* in the frontal lobe.

More detailed brain expression information was sort to identify regions of the brain, in addition to the frontal lobe, that could be disrupted in EFMR patients. Initial investigations into the overall structure and morphology of 14 affected female brains by means of MRI and CT-scan identified that, with the exception of one individual, they all appeared normal (Scheffer et al., 2008). The exception was an affected female from family 1 (Family 1, III-11, figure 3.7) who had a retro-cerebral arachnoid cyst (Scheffer et al., 2008). Therefore we were looking to extend the previous *in-situ* analysis and identify brain regions which could warrant further imaging investigations in the EFMR patients.
**PCDH19 in situ** expression analysis identified expression in the developing cortex, hippocampus, thalamus, hypothalamus and has extended previous findings (Gaitan and Bouchard, 2006) by demonstrating that *PCDH19* is most prominently expressed in the frontal lobe. Expression in embryonic and adult stages is widespread. The expression pattern of *PCDH19*, with the highest level of expression predominantly in the brain, fits well with the clinical presentation of mutations in *PCDH19* with the clinical features of EFMR being confined to brain functioning. The lack of observable defects in additional tissues, which express *PCDH19*, suggests that normal expression is essential only for the correct functioning of the brain.

### 3.4.3 A role for *PCDH11Y* in EFMR?

*PCDH11X* had initially been considered as a strong candidate gene for EFMR and had been sequenced in a number of EFMR female probands; however, no mutations were identified (Scheffer et al., 2008). *PCDH11XY* is of interest in EFMR due to the expression of *PCDH11Y* exclusively in males. *PCDH11Y* is also a member of the same subgroup of protocadherins as *PCDH19*, the δ-protocadherin subgroup. This subgroup is characterised by the presence of the conserved CM1 and CM2 domains, as illustrated in figure 3.29. It is also subdivided into two further groups’ δ1 and δ2 on the basis of the presence or absence of a third conserved domain the CM3 domain and a seventh EC-cadherin repeat. *PCDH11Y* is a member of the δ1 as it contains the CM3 domain and seven EC-cadherin repeats, whilst *PCDH19* which does not contain the CM3 domain and only six EC-cadherin repeats is a member of the δ2 subgroup (figure 3.29). *PCDH19* and *PCDH11Y* are therefore relatively closely related protocadherins.

It is possible that *PCDH11Y* is involved in rescuing males with mutations in *PCDH19* from the EFMR phenotype. For this theory to be plausible we would expect there to be overlap in
the expression patterns of \textit{PCDH19} with \textit{PCDH11Y} expression, such that \textit{PCDH11Y} could rescue cells with \textit{PCDH19} mutations. There would also have to be some difference in terms of expression pattern or functionality between \textit{PCDH11X} and \textit{PCDH11Y}, for \textit{PCDH11Y} to be able to rescue cells with \textit{PCDH19} mutations in males while \textit{PCDH11X} is unable to rescue cells with \textit{PCDH19} mutations in females.

\textit{PCDH11X} and \textit{11Y} are 98.1\% identical at the nucleotide level and 98.3\% identical at the amino acid level; however, the small differences between these two genes may have a significant impact on their function, in terms of differences in their spatial and temporal expression. To assess the plausibility of \textit{PCDH11Y} being involved in rescuing males from EFMR we investigated the expression of \textit{PCDH11X} and \textit{11Y} with respect to one another and \textit{PCDH19}. The overlapping expression between \textit{PCDH11X/Y} and \textit{PCDH19} in combination with the sexually dimorphic expression of \textit{PCDH11X/Y} in the human brain, suggests that \textit{PCDH11Y} could possibly function as a compensatory mechanism in males.

In light of the identification of the causative gene for EFMR we propose an alternative hypothesis for the mechanism behind the unusual inheritance of this disorder. We propose a model involving cellular interference in combination with functional rescue by a related but non-paralogous protocadherin(s) gene.

\textit{PCDH19} is subject to X-inactivation, and this inactivation occurs randomly in EFMR females resulting in mosaic population of cells with approximately half of the cells expressing functional \textit{PCDH19} while the other half are without functional \textit{PCDH19}. Males are however hemizygous for the \textit{PCDH19} mutations and will therefore have a homogenous population of cells without \textit{PCDH19} function. We speculate that the presence of \textit{PCDH19} positive and \textit{PCDH19} negative (or \textit{PCDH19} non-functioning) cells in females may instigate wrong cell-
cell connectivity and networks in the developing brain. This hypothesised scrambled cell-cell communication may then manifest at a clinical level as epilepsy and/or ID.

In males the *PCDH19* dependant communication lines are not instigated. In their absence a likely alternative route, through the use of, for example, activation of a nonparalogous protocadherin gene. *PCDH11Y* is the only male specific protocadherin and therefore the most obvious candidate nonparalogous protocadherin capable of rescuing males from the effects seen in female heterozygotes.

*PCDH11Y* does however have an X-chromosome paralog *PCDH11X*. Whilst these two paralogs *PCDH11X* and *11Y* are highly similar (Blanco et al., 2000) there are differences in the brain expression patterns of these genes (Blanco et al., 2000). This suggests that they do not function identically. A number of the brain regions where *PCDH11X* and *11Y* are expressed also show sexual dimorphism (Blanco et al., 2000, Vanhalst et al., 2005). It is possible that these differences in brain expression of *PCDH11X* and *11Y* may account for their differential ability to compensate for mutations in *PCDH19* in males and females.

### 3.4.4 Hemizygous deletion of *PCDH19* in an affected Male

Depienne et al have recently published a male patient with a 1Mb hemizygous deletion of Xq22.1, which encompassed a single gene, *PCDH19* (Depienne et al., 2009a). This male has epilepsy and moderate to severe ID, and was initially referred for genetic testing for Dravet syndrome (DS), but was found to be negative for point mutations and microchromosomal deletions and duplications of *SCN1A*. This male was described as having PCDH19-associated DS (PCDH19-DS). The identification of a male affected by epilepsy and ID with a deletion of *PCDH19* excludes the possibility of a Y-chromosome rescue factor in males.
Out of the 19 families identified with mutations in *PCDH19* (EFMR and PCDH19-DS patients) there is only one affected male, which is a mosaic with a hemizygous deletion of *PCDH19*. Otherwise, in these 19 families there are 24 unaffected transmitting males with point mutations in *PCDH19*. Depienne et al used fluorescent in-situ hybridisation (FISH) analysis of skin fibroblasts of this male to demonstrate that he was a mosaic with ~53% of his fibroblasts containing wild type *PCDH19* and ~47% with the *PCDH19* deletion (Depienne et al., 2009a). This situation therefore replicates what is seen in heterozygous females with EFMR.

The identification of an affected, mosaic male with *PCDH19* mutation has implications for the study of the molecular mechanism of EFMR by excluding the hypothesis of a male specific factor contribution. It also reinforces the postulated hypothesis of scrambled, cell-cell communication as the most likely basis for the unusual EFMR inheritance.

For the speculated model of cellular mosaicism to explain the unusual inheritance of EFMR it still requires compensation for mutations in *PCDH19* to occur in the hemizygous males, and presumably in a hypothetical homozygous female as they lack functional *PCDH19*. This compensation could in principal occur from any member of the protocadherin family. Identification of the specific neural cells that cause the EFMR phenotype will aid the identification of which of the more than 70 protocadherins (reviewed in Morishita and Yagi, 2007) are capable of compensating for mutation in *PCDH19*. Suitable animal models involving, for example, knockout of *Pcdh19* and various other protocadherins could also aid the identification of the compensatory gene. If cellular mosaicism is the mechanism involved in EFMR, a homozygous mutation in *PCDH19* in a female should not have a detrimental effect. This might be testable through a suitable animal model, since ascertainment of a homozygous human female is highly unlikely given the rarity of EFMR.
3.4.5 CFNS molecular mechanism

The cellular mosaicism mechanism predicted to be responsible for EFMR inheritance appears highly similar to the mechanism proposed for Craniofrontonasal Syndrome CFNS. CFNS also exhibits counterintuitive sex specific phenotypic severity with females more severely affected than males. However, unlike in EFMR the CFNS males are affected. The mechanism that has been proposed to explain the sex-dependant manifestation of CFNS is termed “cellular interference” (Wieacker and Wieland, 2005).

\textit{EFNB1}, the gene mutated in CFNS, is on the X-chromosome and is subject to X-inactivation (Carrel et al., 1999). This X-inactivation occurs randomly in affected females (Wieland et al., 2004), thereby setting up a heterogeneous population of \textit{EFNB1} positive and negative cells in females. It is this mixed population of cells, which is believed to be the basis of CFNS. Wieacker and Wieland (Wieacker and Wieland, 2005) first proposed the term ‘cellular interference’ to explain the mechanism responsible for CFNS. Cells with mutations in \textit{EFNB1} present surrogate ephrin molecules for interaction with adjacent ephrin-receptor expressing cells. The surrogate ephrin likely instigates different signalling pathways, thereby setting up cell compartments with different “cross talk”. These differences in cellular cross talk are predicted to disrupt the sharp segmental interfaces, resulting in defective boundary formation (Wieacker and Wieland, 2005, Twigg et al., 2006, Wieland et al., 2004). In mutant hemizygous \textit{EFNB1} males a homogenous population of \textit{EFNB1} negative cells are present. While an absence of \textit{EFNB1} function still needs to be compensated for, the outcome is a homogeneous cell population and thus better cellular cross talk leading to a less severe clinical outcome (Wieland et al., 2004, Twigg et al., 2006, Wieacker and Wieland, 2005).
Heterozygous EFNB1\textsuperscript{KO/+} mice validate these suggestions. Mutations in \textit{Efnb1} have different effects in heterozygous females when compared to hemizygous males and homozygous females. Targeted inactivation of \textit{Efnb1} in mice resulted in perinatal lethality, edema, defective body wall closure, and skeletal abnormalities (Compagni et al., 2003). An additional phenotype, preaxial polydactyly was exclusively present in the heterozygous female EFNB1\textsuperscript{KO/+} but was not seen in hemizygous male or homozygous female mice (Compagni et al., 2003). This is consistent with the cellular interference model.

Complete knock out of EFNB1 in mice recapitulated these findings. Polydactyly was identified in 100\% of heterozygous females and was nearly always restricted to digits I and II either in the forelimbs or hindlimbs (Davy et al., 2004). Expression of \textit{Efnb1} in the limb bud of heterozygous females was strikingly different to wild-type littermate controls, with alternative bands of \textit{Efnb1} positive and \textit{Efnb1} negative cells (Davy et al., 2004). Davy et al (Davy et al., 2004) suggest that the atypical expression pattern was not a result of random X-inactivation as this process occurs earlier in development and should result in a much finer distribution of \textit{Efnb1} positive and negative cells. They believed the size and distribution of the \textit{Efnb1} positive patches were suggestive of a disruption to cell sorting in heterozygous females in which \textit{Efnb1} positive cells segregated away from \textit{Efnb1} negative regions (Davy et al., 2004). In support of this hypothesis was the finding that expression of \textit{Efnb1} in the limb bud at an earlier time point of E8.5 in heterozygous females was uniformly distributed. This suggests that the banding of \textit{Efnb1} positive and \textit{Efnb1} negative cells is an active process that occurs over time. These results indicate that mosaic loss of \textit{Efnb1} is the underlying cause of the polydactyly phenotype (Davy et al., 2004).

Compagni et al previously suggested based on the findings from the heterozygous \textit{Efnb1}\textsuperscript{KO/+} mice that mosaic expression of an X-linked gene can affect biological function leading to
more severe phenotypes in heterozygous females. Disorders with unusual severity in heterozygous females should be investigated for defects in cell adhesion and sorting as a result of the clonal pattern of X-inactivation (Compagni et al., 2003). The finding that a gene involved in cell-cell adhesion is defective in EFMR (Appendix A, Dibbens et al., 2008) females reinforces this suggestion.

The more we learn about CFNS and EFMR / PCDH19-DS the more parallels between these disorders we identify. Both of these disorders are X-linked disorders where females are affected whilst carrier-transmitting males are less severely affected in the case of CFNS or unaffected in EFMR / PCDH19-DS (with the exception of the mosaic PCDH19 deleted male). Both PCDH19 and EFNB1 are transmembrane proteins involved in cell-cell adhesion and migration. Cellular interference has been proposed as the molecular mechanism to explain the unusual inheritance pattern of these two disorders with mosaicism invoked to explain the one male with PCDH19 deletion (Depienne et al., 2009a). The vast majority of mutations in both disorders occur in the extracellular domain of the respective proteins. To date all known PCDH19 mutations cluster in the extracellular domain (Depienne et al., 2009a, Appendix A, Dibbens et al., 2008) and 74 of the 77 different EFNB1 mutations are also in the extracellular domain (Wallis et al., 2008). The location of these mutations fits well with the proposed cellular interference mechanism and the functions of these proteins. For cellular interference to occur there has to be cues on the cell surface specifying similarities and differences between cells which ultimately determine whether or not they interact.

CFNS and EFMR females also present with a wide range of intra and inter familial variation in expressivity (Wieland et al., 2004, Wieland et al., 2002), which does not necessarily correlate to mutation type. X-inactivation of EFNB1 in tissues relevant to the disease has been suggested as a possible mechanism responsible for the variability in phenotype. A high degree
of X-inactivation towards the wild type allele would be predicted to result in a less severe phenotype (Compagni et al., 2003). However, investigations into the level of cellular mosaicism resulting from X-inactivation in CFNS females have been contradictory. With the identification of a 48% level of mutation in blood in a female with minimal clinical manifestations of CFNS (she was classified as unaffected, but investigated as her daughter had CFNS), while a female in another family had classical CFNS with a level of mutation in the blood of only 13% (Twigg et al., 2006). The poor correlation between the level of mosaicism and clinical severity is likely due to an inability to assess the level of mosaicism that occurs in the relevant tissues at the relevant developmental time point. Differences in the level of mosaicism between tissues can be considerable, the female with a 48% level of mutation in blood only had a 22% level of mutated cells in hair roots (Twigg et al., 2006).

The variation in the expressivity in CFNS and EFMR patients could also be the result of additional modifier genes. Modifier genes have previously been implicated in the phenotypic heterogeneity present in epilepsy. In particular modifier genes have been implicated in Dravet syndrome (DS) and Genetic (generalised) Epilepsy with febrile seizures plus (GEFS+) arising from mutations in SCN1A (reviewed in Scheffer et al., 2009). A similar scenario could be true for CFNS and EFMR, although ascertaining specific modifier genes involved in these disorders will be challenging.

To date there has only been one EFNB1 mutant mosaic male identified with a missense mutation in EFNB1 (Twigg et al., 2006). He was investigated for mosaicism in EFNB1 as he had fathered an unaffected and an affected daughter with CFNS. This male had low levels of mosaicism, with a 3% mutation level in blood and 2% in hair roots, and was presumed to have gonadal mosaicism due to fathering an affected and unaffected daughter. This mosaic male has mild features of the CFNS male-carrier phenotype, rather than the more severe
CFNS-like phenotype (Twigg et al., 2006). It is interesting to note that even with this very low level of mutation, <5%, this male was still affected, with mildly dysmorphic facial features and nasal asymmetry (Twigg et al., 2006).

The high degree of apparent mechanistic similarity between EFMR and CFNS represents a basis for further investigations. Knowledge gained from one may aid in a better understanding of the mechanisms in the other. It will also be fascinating to see if there are any additional similarities such as in the signalling pathways disrupted in these disorders.

### 3.4.6 Conclusion

We have identified causative mutations in *PCDH19* in all the currently known EFMR families. We have also determined that *PCDH19* is predominantly expressed in the brain. From the original four suggestions put forward to explain the unusual inheritance pattern of EFMR (Ryan et al., 1997), we can rule out metabolic interference and defective X-inactivation as possible mechanisms. The hypothesis of the Y-chromosome factor rescue was not excluded by our studies; however, recent identification of an affected mosaic male (Depienne et al., 2009a) rules this out. The remaining hypothesis, that the defective gene in EFMR is not required for development of the male brain, is true. We know that the transmitting males with protein truncation mutations in *PCDH19* are likely to be devoid of *PCDH19* protein and thus *PCDH19* null. Therefore *PCDH19* is not required for correct functioning of the male brain and likely even female brain if homozygously deleted or non-functional (which in practical terms can only be tested using an animal model). We propose a model involving cellular mosaicism in combination with functional rescue by a related but non-paralogous protocadherin gene (not including *PCDH11Y*). The discovery of an association between mutations in *PCDH19* and epilepsy and ID opens up involvement of a new family of genes in these conditions.
3.5 Contributions

Work performed by the candidate

- Assessment of alternative splicing of exon 2 of $PCDH19$.
- Conservation analysis and likely functional consequences of the mutated residues in $PCDH19$.
- $PCDH19$ expression analysis, via northern blot, RT-PCR and RT-qPCR analysis.
- Generation of the $PCDH19$ in situ probe.
- Assessment of the native protein expression of $PCDH19$ and over expression of $PCDH19$ for assessing the $PCDH19$ antibodies. The candidate performed this work with assistance from C. Shoubridge (Department of Genetic Medicine, Women’s and Children’s Hospital).
- The assessment of nonsense mediated decay; however, the tissue culture for these experiments was performed by L. Vandeleur (Department of Genetic Medicine, Women’s and Children’s Hospital).
- RT-qPCR assessment of $PCDH19$ expression in EFMR females.
- The experiments involved in determining $PCDH19$ is subject to X-inactivation. Once again the tissue culture for these experiments was performed by L. Vandeleur (Department of Genetic Medicine, Women’s and Children’s Hospital).
- Assessment of the expression of $PCDH11X/11Y$ by northern blot.
- Assessment of the expression of $PCDH11X$ and $11Y$ by RT-qPCR.

Identification of additional EFMR families

I. Scheffer (Epilepsy Research Centre and Department of Medicine, University of Melbourne) was the principal clinician involved with the EFMR project. S. Turner, C. Derry and S.
Berkovic (Epilepsy Research Centre and Department of Medicine, University of Melbourne) Z. Afawi, A. Mazarib, M. Neufeld (Department of Neurology, Tel Aviv Sourasky Medical Centre, Israel), S. Kivity (Department of Neurology, Schneider Children’s Medical Centre, Israel), D. Lev, T. Lerman-Sagie (Metabolic Neurogenetic Clinic, Wolfson Medical Centre, Israel), A. Korczyn (Molecular Neurogenetics Unit, Centre for Human Genetic Research, Massachusetts, USA), E. Haan (Department of Genetic Medicine, Woman’s and Children’s Hospital), S. Ryan (AstraZeneca, Delaware, USA), S McKee (Northern Ireland Regional Genetics Service, Belfast, UK) identified families and provided clinical information.

**PCDH19 gene identification and preliminary characterisation performed prior to my arrival**


Additional *PCDH19* screening and confirmatory *PCDH19* sequencing was performed and supervised by M. Bayly and L. Dibbens respectively (Department of Genetic Medicine, Women’s and Children’s Hospital).

The full length *PCDH19* sequence was constructed by J. Gecz and L. Dibbens (Department of Genetic Medicine, Women’s and Children’s Hospital).

**Additional contributions**

- M. Shaw (Department of Genetic Medicine, Women’s and Children’s Hospital) performed the X-inactivation assessment of the EFMR families.
• M. Bayly (Department of Genetic Medicine, Women’s and Children’s Hospital) designed the PCDH11X and PCDH11Y specific primers.

• P. Thomas and E. Sutton (School of Molecular and Biomedical Science, University of Adelaide) performed the mouse in-situ hybridization with a probe generated by the candidate.

• The human in-situ work was performed by J. Bomar and D.H. Geshwind (Program in Neurogenetics and Neurobehavioral Genetics David Geffen School of Medicine, University of California at Los Angeles, USA).

• D. Brooks and E. Parkinson-Lawrence (Division of Health Sciences, School of Pharmacy and Medical Sciences, University of South Australia) assisted with the epitope design for the three PCDH19 antibodies we generated.

• Mimotope generated the PCDH19 antibodies as a commercial service. Antibody purification was performed by C. Shoubridge (Department of Genetic Medicine, Women’s and Children’s Hospital).
Chapter 4 – PCDH19 screening

SECTION 1 155

4.1 INTRODUCTION 155
4.2 COHORTS SCREENED 156
4.2.1 RETT SYNDROME PATIENT COHORT 156
4.2.2 AUTISM SPECTRUM DISORDER PATIENT COHORT 157
4.2.3 FEMALES WITH EPILEPSY, WITH OR WITHOUT INTELLECTUAL DISABILITY 157
4.3 RESULTS 157
4.3.1 PCDH19 SCREENING 157
4.3.2 NUCLEOTIDE VARIATION OF PCDH19 160
4.3.3 CLINICAL DESCRIPTION OF THE PCDH19 POSITIVE CASES 162
4.4 DISCUSSION 165
4.5 CONCLUSION 170

SECTION 2 171

4.6 INTRODUCTION 171
4.7 RESULTS 172
4.7.1 LARGE USA FAMILY 172
4.7.2 COMPARISON OF MUTATIONS IDENTIFIED IN PCDH19 IN EFMR AND PCDH19-DS 175
4.7.3 OVERALL CONSERVATION AND MUTATION POSITIONS IN PCDH19 179
4.8 DISCUSSION 182
4.8.1 LARGE USA FAMILY 182
4.8.2 COMPARISON OF MUTATIONS IDENTIFIED IN PCDH19 IN EFMR AND PCDH19-DS 183
4.8.3 OVERALL CONSERVATION AND MUTATION POSITIONS IN PCDH19 184
4.9 CONCLUSION 187
4.10 CONTRIBUTIONS 189
Preamble

This results chapter is broken up into two sections; the first consists of follow up screening we performed to identify additional \( PCDH19 \) mutations. This work has now been published in The Journal of Medical Genetics (Hynes et al., 2009). A copy of the paper is in the appendix (Appendix B). The contributions made by all authors are outlined at the end of this chapter.

The second section describes the identification of a new \( PCDH19 \) mutation in a large family from USA. Finally, in this chapter I compare the mutations identified in EFMR to those identified recently in Dravet syndrome (Depienne et al., 2009a).

Section 1

4.1 Introduction

Subsequent to the identification of mutations in the \( PCDH19 \) gene as the cause of EFMR (Appendix A, Dibbens et al., 2008), we performed additional screening in individuals with clinical features that overlap those of EFMR. The aim was to test our hypothesis that EFMR as defined by \( PCDH19 \) mutations might represent a cause of seizures and ID in females more frequently than currently recognised. Once detected this might allow us to expand on the EFMR phenotype.

EFMR is easily identified by inspection of large pedigrees where the phenotype occurs in multiple affected females connected through unaffected male relatives. However, in smaller pedigrees, with just a few affected and isolated cases, EFMR due to mutations in \( PCDH19 \) may well be an under-recognized cause of seizures and ID. The characteristic unique inheritance pattern is not obvious in small pedigrees. The prediction of a broader phenotype
has recently been validated by the identification of $PCDH19$ mutations in 12 unrelated individuals (11 females and one mosaic male) with severe epileptic encephalopathies resembling Dravet syndrome (Dibbens et al., 2008). Six were proven de novo and five were familial. Of these familial, four had unaffected carrier fathers and one had a father with mild intellectual disability (Depienne et al., 2009a).

Clinical assessment of 27 females with EFMR found that whilst seizures beginning by age 3 years were the hallmark of this disorder, there were several additional clinical features (Scheffer et al., 2008). Sixty seven percent of affected females had ID or were of borderline intellect. Early development varied from normal to abnormal; developmental regression commonly occurred with seizure onset (Scheffer et al., 2008). A number of psychiatric features were identified in affected females: obsessive features were present in 33%, aggressive behaviour in 26% and autism spectrum disorders (ASD) in 22% (Scheffer et al., 2008).

Here, we searched for a broader spectrum of phenotypes caused by $PCDH19$ mutations, by performing mutation screening in three cohorts of females with clinical features that overlap those of known EFMR females. These three cohorts consisted of Rett syndrome (RS) females, females with ASD and females with epilepsy with or without ID.

### 4.2 Cohorts screened

#### 4.2.1 Rett syndrome patient cohort

Females with EFMR often have normal early developmental milestones and regress in infancy (Juberg and Hellman, 1971, Ryan et al., 1997, Scheffer et al., 2008). They exhibit autistic features and intellectual disability later in life. Thus the course of EFMR resembles
that of RS, including RS girls with early onset seizures. RS females from the Australian Rett database (ARSD) (Weaving et al., 2003), who did not have mutations in MECP2 or CDKL5/STK9, were selected (n=42). This cohort can be considered an enriched cohort for identifying additional EFMR cases as it consists of females with epilepsy and ID where MECP2 and CDKL5 mutations have already been removed.

### 4.2.2 Autism spectrum disorder patient cohort

In view of the prominent autistic features observed in girls with EFMR (Scheffer et al., 2008), a cohort of girls with ASD was selected from the Autism Genetic Resource Exchange (AGRE) database of the Cure Autism Now Foundation, USA (n=50) and from our patient cohorts (n=7). Individuals screened were selected based on the pedigree information; each proband was selected from a pedigree with multiple affected females across one or more generations and no affected males.

### 4.2.3 Females with epilepsy, with or without intellectual disability

In 51/86 cases, seizure onset was under 3 years, in 6/86 cases seizure onset was after 3 years. The age of seizure onset was unknown in the remaining 29/86 cases. No other cause for epilepsy was identified.

### 4.3 Results

#### 4.3.1 PCDH19 screening

**Rett syndrome cohort**

Sequencing of PCDH19 in 42 females with RS identified no mutations. The RS female cohort had previously been screened for mutations in MECP2 and CDKL5/STK9 and all individuals
were negative for mutations in either gene. This RS cohort was therefore a highly enriched cohort for finding additional causative genes for RS, but \( PCDH19 \) was not one of them.

**Autism spectrum disorder cohort**

We did not identify any mutations in \( PCDH19 \) in 57 index females with autism spectrum disorder (ASD) from families with multiple affected females and no affected males. This suggests that \( PCDH19 \) is not a significant contributor to ASD alone, in females.

**Epilepsy, with or without intellectual disability cohort**

Mutation screening in this cohort of 86 females identified two missense changes in \( PCDH19 \) (2/86 or 2.3%). Both changes were identified in the cohort of 51 females with documented seizure onset before 3 years. One of the females was of normal intellect and one had moderate ID. One of the changes, c.1671C>G (p.N557K), and the associated pedigree have previously been reported (Family 7) (Appendix A, Dibbens et al., 2008). Here we report detailed clinical data on this family (see section 3.3.2).

The second change identified was a novel heterozygous missense change c.826T>C (p.S276P). The patient's parents did not carry the change (figure 4.1A) and biparental testing using six microsatellite markers confirmed that the pedigree structure was correct (figure 4.1B). Thus, this girl carries a \textit{de novo} change; however, the possibility of gonadal mosaicism in one of the parents cannot be excluded. Neither of the unaffected sisters was available for genetic analysis.
Figure 4.1 De novo missense mutation in an affected female

(A) Sequence chromatograms of PCDH19 exon 1 illustrating the c.826T>C mutation identified in the proband (II-2, bottom panel) but absent in the parents (I-1 and I-2), the two unaffected sisters (II-1 and II-3) were unavailable for genetic analysis. (B) Pedigree of the family; black symbol denotes the affected girl and the arrow indicates the proband. (C) Schematic representation of the PCDH19 protein with all PCDH19 mutations identified to date. Mutations identified in EFMR patients are illustrated above the protein schematic (Dibbens et al., 2008), with the mutation identified in this investigation illustrated in bold. Mutations identified in the Dravet-like (DS) patients (Depienne et al., 2009a) are illustrated below the schematic. (D) Partial alignment of human PCDHs and orthologs of PCDH19 from available species shows a high degree of conservation of the serine 276 residue, indicated in the teal box. (E) Clustal W alignment of the EC1 domain of classical cadherins (C, E, and N) and protocadherin α4 with the EC3 domain of PCDH19 which contains the S276 residue mutated in the de novo case (circled in red). Arrows indicate residues involved in the formation of the βC sheet, navy for cadherins and teal for protocadherins. Residues involved in the adhesion interface of cadherins and the formation of a hydrophobic cluster in protocadherins are illustrated in the purple box (Patel et al., 2006, Morishita et al., 2006). p.S276 is adjacent to a highly conserved residue involved in the formation of a hydrophobic cluster (p.Y275).
The c.826T>C nucleotide change is predicted to cause replacement of the serine (S) amino acid at position p.276 with a proline (P) (figure 4.1E). Serine p.276 is highly conserved across a range of protocadherins and PCDH19 of various species (figure 4.1C). Serine p.276 is equivalent, based on alignment to the EC1 domains of classical cadherins and protocadherin α4, to residues Y37 and F38 respectively, which are involved in the formation of the βC strand and adjacent to a highly conserved residue important for the formation of a hydrophobic cluster (figure 4.1D) (Morishita et al., 2006, Patel et al., 2006). Based on the serine p.276 conservation, location within the extracellular domain of PCDH19 and its absence from control chromosomes, we postulate that this change has functional significance and thus represents a novel PCDH19 mutation.

### 4.3.2 Nucleotide variation of PCDH19

Sequence analysis of PCDH19 also identified 6 silent and 1 missense variants (figure 4.2A, Table 4.1). Three silent changes; c.402C>A (p.I134I), c.1137C>T (p.G379G) and c.1627C>T (p.L543L) had previously been identified as part of the HapMap project, while the remaining variants had not previously been reported.

<table>
<thead>
<tr>
<th>Nucleotide change</th>
<th>Protein consequence</th>
<th>Type of change</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c.6 G&gt;A</td>
<td>p.E2E</td>
<td>Silent</td>
<td>1.23</td>
</tr>
<tr>
<td>c.402 C&gt;A</td>
<td>p.I134I</td>
<td>Silent</td>
<td>7.80</td>
</tr>
<tr>
<td>c.655 C&gt;T</td>
<td>p.L219L</td>
<td>Silent</td>
<td>0.44</td>
</tr>
<tr>
<td>c.1137 C&gt;T</td>
<td>p.G379G</td>
<td>Silent</td>
<td>3.43</td>
</tr>
<tr>
<td>c.1627 C&gt;T</td>
<td>p.L543L</td>
<td>Silent</td>
<td>20.62</td>
</tr>
<tr>
<td>c.1683 G&gt;A</td>
<td>p.P561P</td>
<td>Silent</td>
<td>0.43</td>
</tr>
<tr>
<td>c.2873 G&gt;A</td>
<td>p.R958Q</td>
<td>Missense</td>
<td>0.46</td>
</tr>
</tbody>
</table>

The line dictates the distinction between the extracellular domain of PCDH19 (above the line) and the cytoplasmic domain (below the line).
Figure 4.2 Population variation and conservation of PCDH19

(A) Schematic representation of the PCDH19 protein with all non-disease causing variants identified to date. Silent variants are illustrated above and missense variants below the protein schematic. (B) Conservation score of PCDH19 relative to all other δ-protocadherins. Conservation scores were obtained from Scorecons – www.ebi.ac.uk. Each bar on the graph represents the conservation score of a single amino acid in PCDH19. Grey bars illustrate amino acids in the extracellular domain of PCDH19 and black the cytoplasmic domain. Missense mutations (n=7) are illustrated as black lines on top of the graph. Missense variants (n=2) are illustrated as grey lines. The PCDH19 protein schematic under the graph illustrates the position of functional domains of PCDH19. EFMR missense mutations are indicated with a hash (#), Dravet-like syndrome mutations are indicated with a star (*) and variant missense changes in grey.
Analysis of the conservation of PCDH19 among the δ protocadherins demonstrates that the extracellular domain of PCDH19 in particular is highly conserved (figure 4.2B). The level of conservation is much lower for the cytoplasmic domain. This had previously been noted for the protocadherin family in general (Sano et al., 1993). To date all missense changes identified in the extracellular domain of PCDH19 cause EFMR or PCDH19-DS (Appendix A, Dibbens et al., 2008, Depienne et al., 2009a) (figure 4.1E). The missense change c.3319C>G (p.R1107G), which occurs in the cytoplasmic domain, had previously been identified in healthy female individuals as a rare variant (Depienne et al., 2009a). We also identified a missense change c.2873G>A (p.R958Q) in the cytoplasmic domain of PCDH19 in a female with ASD; however, this is unlikely to be a causative mutation because as part of our large scale X-chromosome exon re-sequencing effort (Tarpey et al., 2009) we identified this sequence variant in an unrelated family with X-linked intellectual disability (data not shown). The p.R958Q change did not segregate with the intellectual disability and was present in three unaffected females. It appears that whilst missense changes in the extracellular domain of PCDH19 cause EFMR or PCDH19-DS, this particular missense change in the cytoplasmic domain of PCDH19 is tolerated in normal individuals.

4.3.3 Clinical description of the PCDH19 positive cases

4.3.3.1 Small family (Family 7 in (Appendix A, Dibbens et al., 2008))

Proband

The 25 year old librarian was selected for screening because she had a history of infantile-onset seizures. At 18 months, she had two simple febrile generalized tonic-clonic seizures. At 3 years 1 month, she started having afebrile right hemiclonic seizures lasting up to 5 minutes. Following an allergic reaction to carbamazepine, she was treated with valproate and later required the addition of clonazepam. At 8.5 years, she presented with atonic drop attacks in
clusters and received clobazam with valproate, as her behaviour had deteriorated on clonazepam. Her last seizure occurred at 12.5 years and she was weaned off antiepileptic therapy at 13.5 years. A postictal EEG study at 4 years showed biposterior high voltage slowing more marked on the left without epileptiform activity. MRI was not performed.

At 10 years, she was diagnosed with Asperger syndrome. She had obsessional features such as crawling out of bed very carefully in the morning so that her bed would not be messy and her bedroom was very tidy. Her full scale IQ showed that she was of average intellect; however, she had significant difficulties with vocabulary and visuospatial tasks. Despite these findings, she coped well academically at school but experienced social difficulties. She completed a university degree.

**Sister of the proband**

Her 23 year old sister presented with right hemiclonic seizures at 2 years. Her seizures began with a scream, then a fixed gaze and clonic activity of the right arm and face, and sometimes involved her leg. After 6 months on carbamazepine, she was changed to valproate. Seizures were controlled by 6 years and medication was weaned at 10 years. Seizures recurred after 8 months, and she was commenced on valproate with the later addition of lamotrigine. By 12 years, seizures began with a 10 second cephalic aura of feeling strange, nauseated and dizzy without true vertigo. Her head and eyes would deviate to the right and sometimes she would have right hemiclonic activity. Seizures would cluster with 20 in one day. Seizures remained highly refractory until they came under control at 16 years when levetiracetam was added to carbamazepine and lamotrigine.

Her EEG at 11 years showed a photoparoxysmal response with high voltage polyspike wave at a number of flash frequencies. Video-EEG monitoring at 14 years captured eight seizures
which began by the patient calling out to her mother, staring blankly, followed by head and eye deviation to the left and oral automatisms. The EEG showed an ictal rhythm which began with right frontotemporal slowing evolving to sharply contoured theta maximal at T4, which spread to involve the right hemisphere. MRI of the brain at 13 years was normal apart from mild ventriculomegaly.

Her early motor developmental milestones were normal but she spoke single words late at 15 months and put two words together after 2 years. She did not regress. She was diagnosed with attention deficit hyperactivity disorder at 6 years and was treated with dexamphetamine into adult life. Neuropsychological assessment at 13 years showed that she was of borderline intellect with good preservation of visuospatial function and average performances in general knowledge and non-verbal concept formation. She had difficulties in both verbal and visual domains of memory function. She also showed planning and organisation problems. She attended a special school and then completed a certificate in caring for the elderly as a young adult. There was no further family history of seizures or intellectual disability.

4.3.3.2 De novo case

This 7 year old girl presented with her first seizure at one year of age, five days after receiving her measles, mumps and rubella immunization. She had clusters of afebrile generalized tonic-clonic seizures (GTCS) and later had seizures with fever. Her EEG at 14 months showed left central and frontal epileptiform activity, but was normal at 4 and 6 years. By 7 years, she had episodes of convulsive status epilepticus, which responded to phenytoin in combination with lamotrigine and clobazam. She had previously tried valproate. Early developmental milestones were normal but by 4 years, she had moderate ID. At 7 years, she received a diagnosis of an autism spectrum disorder; she had not shown neurodevelopmental regression. Examination showed no focal neurological signs or dysmorphic features. MRI of
the brain was normal. She was the second child of unrelated Vietnamese parents and had two normal sisters, aged 3 and 9 years. There was no family history of seizures or intellectual disability.

### 4.4 Discussion

We sought to ascertain the frequency of PCDH19 mutations in females who share overlapping clinical characteristics with those that occur in EFMR (Scheffer et al., 2008, Ryan et al., 1997). The most typical phenotype seen in EFMR females comprises a normally developing infant whose seizures begin under 3 years (mean age of 14 months), sometimes with developmental regression at seizure onset. Two thirds of affected females have borderline intellect or intellectual disability, which varies from mild to profound. Autistic features may be prominent (Scheffer et al., 2008). We selected three cohorts based on specific phenotypic features to ascertain whether PCDH19 was a significant cause of these phenotypes.

First, a number of phenotypic similarities exist between RS and EFMR. RS (MIM 312750) is a severe neuro-developmental disorder which predominantly affects females. RS is characterised by developmental regression between 6 and 18 months of age and the development of other features including stereotypic hand movements and hyperventilation. Additional characteristics include seizures (Scala et al., 2005) and autistic features (Hagberg et al., 1983, Young et al., 2008). Mutations in either MECP2 (Amir et al., 1999) or CDKL5 (Tao et al., 2004, Weaving et al., 2004, Scala et al., 2005) are found in the majority of girls with RS; however approximately 25% of patients with RS do not carry a mutation of either gene (Downs et al., 2008). Therefore it is likely that additional genes contribute to RS (Hagberg et al., 1983). We used an enriched RS cohort, where mutations of MECP2 and
CDKL5 had been excluded, to screen for PCDH19 mutations. Despite the phenotypic similarities shared between females with RS and EFMR, we did not identify any mutations in PCDH19. This suggests that PCDH19 is unlikely to make a significant contribution to RS. The possibility remains that mutation in PCDH19 could be a rare cause of RS; or from the clinical perspective some EFMR females could rarely be diagnosed as RS. Whilst there are a number of phenotypic similarities between EFMR and RS females, the mutant males in these two disorders show opposite phenotypes. EFMR transmitting males do not have seizures or ID whilst the rare RS males with MECP2 mutations are more severely affected, than their affected sisters, mutations in MECP2 in males are generally incompatible with life (Hardwick et al., 2007). Thus, the absence of PCDH19 mutations in the RS cohort, together with the opposing phenotypes in hemizygous males carrying mutations, strongly suggests that unrelated molecular mechanisms are responsible for these two disorders.

Second, psychiatric features are prominent in females with EFMR. In particular, ASD and/or autistic features have been identified in 22% of EFMR females (Scheffer et al., 2008). ASD has a significant genetic component with at least 76 genes or chromosomal loci proposed to be associated with ASD (reviewed in Abrahams and Geschwind, 2008). None of the associated genes account for more than 1-2% of cases to date (reviewed in Abrahams and Geschwind, 2008). The PCDH19 gene has not previously been associated with autism; however, two close sequence relatives PCDH9 and PCDH10 (also δ protocadherins) have recently been implicated as potential candidate genes for autism (Marshall et al., 2008, Morrow et al., 2008). Despite this, we found no evidence that PCDH19 causes isolated ASD within the relatively small cohort that we screened. We did nevertheless find that the PCDH19 mutations identified in our third cohort occurred in association with Asperger syndrome in one proband and ASD in the other, confirming that there does in fact exist some level of association between ASD and EFMR.
Lastly, we screened 86 females with epilepsy with or without ID where the age of seizure onset was known to be less than 3 years in 51 cases. Half the affected females had a family history of seizures; however, their families did not show the characteristic inheritance pattern of EFMR over multiple generations as their pedigrees were too small. This screen allowed us to identify the smallest EFMR family to date with just two affected girls in a single generation and a normal transmitting father who carried a novel de novo *PCDH19* mutation (Appendix A, Dibbens et al., 2008). Taken together, our previous (Appendix A, Dibbens et al., 2008) and current screening ascertained two *PCDH19* mutations out of 86 females from small kindreds and sporadic cases. Thus EFMR is not a prominent cause of seizures and intellectual disability in females; however, a frequency of 2.3% means that EFMR must be considered sufficiently common to be part of the differential diagnosis given its major genetic counselling implications. Daughters of a female with a *PCDH19* mutation have a 50% risk of having a mild to severe EFMR phenotype. Their brothers will have a 50% risk of transmitting a *PCDH19* mutation on to their offspring. Since in most cases the mutation will have been inherited from a parent, this will have implications for other family members and this could have implications for other branches of the family.

The new case was a novel de novo p.S276P missense mutation detected in a sporadic female with refractory tonic-clonic seizures and intellectual disability. This girl’s phenotype poses a striking contrast to the proband of the small EFMR family who had relatively easily controlled infantile seizures and was of normal intellect. Previously cases of EFMR have been recognised based on its distinctive inheritance pattern of affected females through several generations connected by normal transmitting males. *PCDH19* testing should therefore be considered as one of the options in females with seizure onset within the first three years and with developmental delay or intellectual disability.
Recently Depienne and co-authors emphasized that females presenting with a severe encephalopathy resembling Dravet syndrome may have PCDH19 mutations (Depienne et al., 2009a). They noted that the PCDH19 cases had a later age of onset, less frequent status epilepticus, myoclonic and absence seizures and a lesser degree of intellectual disability than cases with classical Dravet syndrome with a mutation in the sodium channel gene SCN1A (Depienne et al., 2009a). These PCDH19 mutation positive cases closely resemble our severe EFMR cases. Indeed, our experience is that the severe EFMR phenotypes, while sharing some features with Dravet syndrome, lack the characteristic temporal evolution of Dravet syndrome (Carranza et al., 2009). In Dravet syndrome seizure onset occurs at around 6 months, followed by frequent convulsive seizures over the next year typically involving episodes of febrile status epilepticus. Between 1 and 4 years, other seizure types emerge and include partial, absence, myoclonic and atonic seizures. Seizures typically remain refractory. Children with Dravet syndrome develop normally in the first year of life, then their development slows and cognitive impairment is usual. This is quite different from the temporal evolution of EFMR.

The three new EFMR positive girls described here do not have Dravet syndrome. The sisters with the familial mutation had seizure onset at 18 to 24 months of age with hemiclonic seizures in both and febrile generalized tonic-clonic seizures in one. Seizures were brought under control in both by mid-adolescence and they were of normal intellect. Our third patient with a de novo PCHD19 mutation presented at one year with afebrile GTCS, which is rather late for Dravet syndrome. She only had convulsive attacks and her episodes of status epilepticus occurred late at 7 years. Thus, detailed phenotypic information shows that her picture is not that of Dravet syndrome although her outcome is similarly severe.
The p.S276P mutation affects a residue in EC3 of PCDH19, which is highly conserved across a number of orthologs and paralogs. The p.S276 is predicted, based on the crystal structure of the extracellular (EC) 1 domain of cadherins (Patel et al., 2006) and protocadherin-α (Morishita et al., 2006), to be involved in the formation of the βC strand (figure 4.1D). The EC repeat domains of cadherins and protocadherins are believed to be composed of two β sheets packed face to face. One β sheet is composed of four β strands (βA, βC, βF and βG) and the second β sheet is composed of the remaining three β strands (βB, βD and βE) (Morishita et al., 2006, Patel et al., 2006). The p.Y275 residue, which is adjacent to the mutated p.S276 residue, is equivalent to residues p.Y37 (Patel et al., 2006) and p.F38 (Morishita et al., 2006) in EC1 of cadherins and protocadherins respectively (figure 4.1D). p.Y37/p.F38 is an important residue involved in the adhesion interface of cadherins and the complementary hydrophobic core of protocadherins. We predict that the p.S276P mutation will disrupt the formation of the βC strand destroying the hydrophobic core of the EC3 domain. Whilst the EC1 domain of classical cadherins was originally thought to be entirely responsible for the homophilic adhesive activity of cadherins, there is increasing evidence highlighting the importance of the other EC domains including EC3 (Chien et al., 2008, Tsuiji et al., 2007). We therefore believe that this mutation results in a functional knockout of PCDH19.

The extracellular domains of protocadherins appear to be crucial for normal function. The amino acid sequence of the extracellular domains of protocadherins is highly conserved, whilst the sequence of the cytoplasmic domains shows more variation between the different protocadherin subclasses (Sano et al., 1993). Our screening of PCDH19 in 186 females analysed as part of this investigation, failed to identify any normal missense variation in its extracellular domain while 2 missense changes have been identified in the cytoplasmic domain (Depienne et al., 2009a). The high level of conservation and the absence of normal
missense variation in the extracellular domain of PCDH19 suggest strong evolutionary constraints acting on this part of the protein. This finding also has implications for PCDH19 diagnostic testing, which should be prioritised to the extracellular and the transmembrane domains. However truncation mutations in the cytoplasmic domain, which are subject to nonsense-mediated decay (NMD), are likely to result in EFMR, therefore the cytoplasmic domain should still be screened for mutations.

Further to this, Mateus et al identified that missense mutations in the extracellular domain or the juxtamembrane domain (which affect the integrity of the extracellular domain) of Cadherin 1 (CDH1) increase cell motility and EGFR activation while missense mutations in the cytoplasmic domain have no effect on EGFR activation or cell motility (Mateus et al., 2009). An interesting correlate is also seen in the Xenopus laevis paraxial protocadherin (PAPC), where the extracellular and transmembrane domains are required for its cell sorting function, whilst the cytoplasmic domain is not (Chen et al., 2007). Whether the EFMR phenotype may therefore occur as a result of defects in cell sorting and or motility, is yet to be investigated.

### 4.5 Conclusion

This section describes a sporadic female with EFMR with a mutation in PCDH19 and gives further clinical details for family 7 (Appendix A, Dibbens et al., 2008). Our recent report of six new families (Appendix A, Dibbens et al., 2008, Scheffer et al., 2008) led us to conjecture that EFMR could be an under-diagnosed disorder due to the lack of key indicators from pedigree structure not visible in small families. Here we relied solely on phenotypic attributes to select the cohort from which we identified the first sporadic affected female with EFMR. In light of this finding, mutation analysis of PCDH19 should be considered in sporadic females.
with seizures beginning under three years of age, in particular in the context of developmental delay with or without ID.

Section 2

4.6 Introduction

A large family from the USA, which partly fit the inheritance pattern seen in EFMR was referred to our laboratory for PCDH19 testing. It contained 11 females affected by various forms of epilepsy (figure 4.3): four females diagnosed with seizures and learning problems, three females with seizures in childhood, three females with seizures, and one female with some seizures beginning in her teens, which then ceased. We identified an additional novel PCDH19 mutation in this family.

During the additional screening as reported in section 1 of this chapter, work was published which identified an additional 11 PCDH19 mutations. These were identified in patients with epileptic encephalopathies resembling Dravet syndrome (DS) (Depienne et al., 2009a). Dravet syndrome (OMIM #607208) is a severe myoclonic epilepsy of infancy. Depienne et al screened 73 patients (45 females and 28 males) and identified 9 different heterozygous mutations in PCDH19 in 11 female probands (two of the mutations were present in two probands) (Depienne et al., 2009a). The authors describe these patients with PCDH19 mutations as having PCDH19 associated Dravet Syndrome (PCDH19-DS) as they have a more severe phenotype than those described in the EFMR families. They suggest that these patients have a syndrome that effectively sits, in terms of severity, between that of EFMR as previously described (Juberg and Hellman, 1971, Ryan et al., 1997, Scheffer et al., 2008) (as the least severe syndrome) and classical cases of Dravet syndrome (as the more severe syndrome) (Claes et al., 2003). Further assessment of the data from the EFMR families
suggest that the PCDH19-DS patient’s phenotype correlates well with that of the more severely affected EFMR females (Carranza et al., 2009).

In light of the apparent differences in the severity of phenotype caused by mutations in PCDH19 in PCDH19-DS and EFMR patients, we set out to investigate the possibility of genotype-phenotype relationships. Results obtained from this analysis lead us to assess the degree of conservation of the amino acids mutated in EFMR and PCDH19-DS.

4.7 Results

4.7.1 Large USA family

Sequence analysis of the coding exons of PCDH19 in three affected females (III-10, III-13 and III-14, indicated by a teal arrow, figure 4.3) identified a G>A heterozygous change c.1123 G>A (figure 4.3) in the three individuals. This nucleotide change is predicted to result in an amino acid substitution from aspartic acid (D) at position p.375 to asparagine (N) (p.D375N). This nucleotide change has not previously been identified in ~750 individuals we have sequenced, or in 172 ethnic matched female controls. The change occurs in exon 1 of PCDH19 that codes for the extracellular cadherin repeat domain 4 (EC4).
Pedigree of the additional USA EFMR family. Displaying the characteristic inheritance pattern of EFMR with affected females and unaffected obligate transmitting males. The sequence chromatogram illustrates the mutation identified in this pedigree with the arrow indicating the mutated base. Individual’s genotype is stated in purple where it has been determined by sequencing. Teal arrows indicated three females sequenced initially. The various shading of individuals describes their phenotype.

Conservation analysis of the affected residue in various protocadherins and PCDH19 from various species reveals that p.D375 is invariant across all protocadherins analysed (figure 4.4). The p.D375N mutation affects the first aspartic acid (D) residue in the calcium binding DXD motif of EC repeat domain 4 (figure 4.5).
Figure 4.4 Conservation of the p.D375 residue affected by the c.1123G>A mutation

Clustal-W alignment of human protocadherins and orthologs of PCDH19 from various species. The amino acid affected by the p.D375N mutation is illustrated in the teal box. Members of the δ1 and δ2-protocadherin subgroup are indicated by the brackets. Amino acids shaded in black are those that do not match PCDH19.

Figure 4.5 Functional consequence of the p.D375N mutation

Clustal-W alignment of the six EC repeat domains of PCDH19. The amino acid mutated in this family is demonstrated in the red circle and indicated by a vertical arrow. Amino acids shaded in grey are those that do not match the consensus sequence. The teal box highlights the DXD calcium-binding motif. Horizontal arrows above the alignment indicated the positioning of the βB and βC strands as identified in Pcdhα4 (Morishita et al., 2006).
4.7.2 Comparison of mutations identified in *PCDH19* in EFMR and PCDH19-DS

Comparison of all known *PCDH19* mutations revealed that both missense and protein truncation mutations have been identified in EFMR and PCDH19-DS patients. Five different protein truncation and 4 different missense mutations have been identified in both EFMR and the same number of each class of mutation for PCDH19-DS patients.

Next we investigated the positioning of the missense mutations. To date 8 different missense mutations have been identified in *PCDH19* (figure 4.6) in nine probands/families. Four different mutations were identified in EFMR patients; p.S276P, p.D375N, p.V441E and p.N557K. Another four different mutations were identified in PCDH19-DS patients; p.D121N, p.E199Q, p.N340S (identified in two unrelated affected individuals) and p.L543P. All of the mutations identified so far occur in exon 1 of *PCDH19*. The four EFMR mutations occur in the EC repeat domains EC3, 4(x2) and 5, while the PCDH19-DS mutations occur in EC1, 2, 3 and 5. So far the same mutation has not been observed in both EFMR and PCDH19-DS patients.
Figure 4.6 Positioning of the missense mutations identified in \textit{PCDH19}

Protein schematic of PCDH19, demonstrating the position of missense mutations identified in EFMR and PCDH19-DS patients. Mutations identified in EFMR patients are illustrated above the protein schematic (Dibbens et al., 2008, Hynes et al., 2009), and mutations identified in PCDH19-DS patients (Depienne et al., 2009a) are illustrated below the protein schematic.

Given that all mutations affect the EC repeat domains we assessed if there was any clustering of the mutations from each disorder to different domains of the EC repeats. Alignment of the six EC repeats of \textit{PCDH19} with the well characterised EC1 domain of \textit{Pcdhα4} (figure 4.7) revealed three out of the 8 missense changes affects regions involved in the formation of the hydrophobic cluster. One of these mutations was identified in a patient with PCDH19-DS while the remaining two were identified in EFMR. Five missense mutations affect regions involved in calcium binding, three of these were identified in PCDH19-DS syndrome and two in EFMR.
Figure 4.7 Positioning of mutations identified in PCDH19 across the EC repeat domains

Clustal W alignment of the six EC repeat domains of PCDH19 with EC domain 1 of mouse and human PCDHα4. Amino acids shaded in grey are those that do not match the consensus sequence. Amino acids mutated in EFMR and PCDH19-DS patients are illustrated in a red circle. The mutation annotation corresponding to each red circle is stated on the left-hand side of the alignment and in line with the mutated amino acid. A star next to the mutation annotation indicates mutations identified in EFMR, those mutations without stars have been identified in PCDH19-DS. Amino acids illustrated in the purple boxes have been identified in PCDHα4 as important hydrophobic cluster residues. Amino acids illustrated in teal boxes are those involved in calcium binding. The horizontal arrows above the alignment indicated the positioning of the β strands as identified in Pcdhα4 (Morishita et al., 2006).

We also assessed the positioning of these mutations in the context of the 3D protein structure of the EC repeat domain. Through the alignment of the six EC repeat domains of PCDH19 with the Pcdhα4 we identified the residues of Pcdhα4 that correspond to those mutated in PCDH19 (figure 4.8). Seven out of 8 missense mutations occur in the N-terminal or C-terminal end of the cadherin repeat. From three N-terminal mutations, two occur in EFMR and one in PCDH19-DS, and from four C-terminal mutations, three were identified in PCDH19-DS and one in EFMR (figure 4.8). Amino acids involved in the formation of the
hydrophobic cluster and calcium binding also cluster to the N-terminal and C-terminal ends of the EC repeat (figure 4.8).

Figure 4.8 3D positioning of mutations in PCDH19 in the EC repeat domain

A A Greek key topography of Pcdho4. A red circle indicates residues equivalent to those mutated in PCDH19. A star beside the mutated residue indicates mutations identified in EFMR patients. Amino acids in a purple box have been identified as hydrophobic cluster residues. A teal box indicates amino acids involved in calcium binding. B A representative ribbon diagram of the NMR structure of Pcdho4. The arrows demonstrate the approximate positioning of amino acids equivalent to those mutated in PCDH19. A star indicates mutations identified in EFMR patients. The purple bracket at the N-terminal end of the EC repeat indicates the region of the protein that contains the hydrophobic cluster residues. The teal bracket at the C-terminal end of the EC repeat indicates the region of the protein that contains the calcium-binding amino acids. Images A and B were taken from (Morishita et al., 2006).
4.7.3 Overall conservation and mutation positions in *PCDH19*

Whilst there was no apparent disease specific pattern to the positioning of mutations in *PCDH19* across EFMR and PCDH19-DS the mutations in both disorders did cluster towards the N-terminal and C-terminal ends of the EC repeat domains (figures 4.7 & 4.8). These regions contain amino acids involved in calcium binding and the formation of the hydrophobic cluster (figures 4.7 & 4.8). This indicates that these regions are important for correct functioning of protocadherins. This led us to investigate the level of amino acid conservation across the EC repeat to determine if amino acids mutated in PCDH19 are generally more highly conserved and functionally significant.

The level of conservation of each amino acid of PCDH19 was assessed relative to the remainder of the δ-protocadherin sub-family using the program Scorecons (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl). A score of 1 is the maximum score attainable and indicates that a residue is 100% conserved across all of the δ-protocadherins assessed. Figure 4.9 illustrates the outcome of this analysis. Each line on the graph represents the average conservation score obtained from all of the amino acids present at each position in the EC domain alignment (illustrated below the graph). The average conservation score represents the level of conservation of an amino acid across the six EC repeat domains of *PCDH19* as well as across the entire δ-protocadherin sub-family. Results from this analysis illustrate once again the high level of conservation of the calcium binding residues. There is an average conservation score of 0.80 across the 12 calcium binding residues. This is well above the average for the entire EC domain of 0.54. Amino acids mutated in EFMR and PCDH19-DS patients are also highly conserved with an average conservation score of 0.77 across the 8 missense mutations. The previously identified
hydrophobic cluster residues are well conserved with an average conservation score of 0.72; however, they are not as well conserved as the calcium binding residues.

**Figure 4.9 Conservation analysis of the EC repeats in PCDH19**
The graph shows the average conservation score for amino acids in the EC repeat domains of PCDH19 as obtained from the Scorecons program, assessing the conservation of PCDH19 amino acids with respect to the remainder of the δ-protocadherin sub family. Each line of the graph represents the average conservation score of the amino acids present at each position in the aligned EC repeat domains. Teal lines on the graph, and the teal boxes in the alignment indicate calcium-binding residues. Purple lines and the purple boxes indicate hydrophobic cluster residues. The line corresponding to amino acid position 31 is coloured half teal and half purple as it has been identified as a hydrophobic and calcium binding amino acid. Navy lines represent the remainder of residues. Yellow stars indicate the positioning of mutations identified in PCDH19. The positioning of mutations is also described below the alignment. The two brackets on the graph indicate two regions of highly conserved residues, which are of interest. The clustal-W alignment of the six EC repeat domains of PCDH19 below the graph demonstrates the amino acids used to calculate the mean conservation score for each position. Amino acids shaded in black are those that match the consensus sequence. The predicted positioning of the seven β-strands (βA-βG) is indicated by horizontal arrows above the alignment. The navy β-strands are involved in the formation of one β-sheet, the teal β-strands are involved in the formation of the second β-sheet (Morishita et al., 2006).
In addition there are two regions of highly conserved amino acids in PCDH19 that are situated in close vicinity to the hydrophobic cluster residues (figure 4.9). Clustal-W alignment of these two regions (figure 4.10) demonstrates that the amino acids in these regions are conserved across the EC2-EC6 domains, but not in the EC1 domain. The amino acids involved in the formation of the hydrophobic cluster in the EC1 repeat are not conserved as hydrophobic residues across the remaining EC repeat domains. The first of the highly conserved regions contains the calcium binding DXD motif, which is also only conserved, across EC domains 2-6.

![Figure 4.10 Additional highly conserved regions of PCDH19 EC repeats](image)

This figure takes a closer look at the two regions identified in the conservation graph of PCDH19 EC repeat domain amino acids. The top panel illustrates a clustal-W alignment of the six EC repeat domains of *PCDH19* with the EC1 repeat domain of human PCDHα4. Amino acids shaded in black are those that match the consensus sequence. Teal boxes indicate calcium-binding domains. Purple shading indicates the highly conserved residues of these domains. Purple boxes indicate residues identified in PCDHα4 as hydrophobic cluster residues. Purple shading indicates regions of high conservation across EC repeat domains 2-6. The positions of mutations identified in *PCDH19* are illustrated below the alignment. The bottom panel contains zoomed in views of the two regions of the two additional highly conserved regions of PCDH19 EC repeats. Purple boxes illustrate the hydrophobic cluster residues identified in PCDHα4. Navy shading illustrates residues conserved as hydrophobic amino acids. Teal shading indicates the conserved DXD motif involved in calcium binding. A bracket indicates regions of highly conserved non-hydrophobic residues.
4.8 Discussion

4.8.1 Large USA family

We have identified a new PCDH19 mutation in a large family. The mutation is predicted to disrupt binding of calcium to the PCDH19 protein. This is a typical EFMR family, with affected females and unaffected transmitting males (Juberg and Hellman, 1971, Ryan et al., 1997). This pedigree does however contain two obligate carrier females (II-6 and II-8, figure 4.3) who have initially been classified as unaffected. We have confirmed the presence of the PCDH19 mutation in these two unaffected females (II-6 and II-8, figure 4.3). The absence of typical clinical presentations of EFMR in these females may be a result of incomplete penetrance or assessment at the time of the evaluation, e.g. these females could have had seizures early in life that went un-noticed or the family is in denial due to possible stigmatisms associated with epilepsy. The investigations to further track the ancestry of this mutation in this family are ongoing.

The aspartic acid residue at position p.375, which is affected by the c.1123 G>A mutation in this family, p.D375, is invariant across all protocadherins we investigated. p.D375 is predicted based on the solution structure of Pcdhα4 (Morishita et al., 2006) to be positioned in the loop between the βB and βC strands. This loop has been identified in cadherins to contain residues important for calcium binding and stabilisation of the cadherin ectodomain conformation (Tamura et al., 1998). In protocadherins the highly conserved DXD motif has been suggested to perform this adhesive function (Kirov et al., 2003, Wu and Maniatis, 2000). The p.D375N mutation affects the first D amino acid in the DXD motif. The substitution of an asparagine (N) at this position we predicted to be detrimental to the calcium binding function of this motif. Aspartic acid residues are present in all of the calcium binding regions identified in the EC repeats of protocadherins. We predict that this mutation, like all other
missense mutations previously identified in PCDH19, will result in a functional knock out of PCDH19.

4.8.2 Comparison of mutations identified in \textit{PCDH19} in EFMR and PCDH19-DS

The identification of \textit{PCDH19} mutations of the same predicted functional effect in two rather distinct patient groups (Depienne et al., 2009a, Dibbens et al., 2008) might superficially appear rather puzzling. The rationale behind this work was to see whether there are subtle differences in the spectrum and type of mutations identified in these cohorts, EFMR and PCDH19-DS, or the differences can be explained, for example, by ascertainment bias. The EFMR families identified from our studies (Dibbens et al., 2008, Scheffer et al., 2008) were identified primarily based on their clinical presentation and inheritance pattern. Therefore there might have been an ascertainment bias towards milder mutations in families in which (at least some) affected females can reproduce and care for their children. Conversely the PCDH19-DS patients have been ascertained by investigating patients with a rather more severe clinical phenotype, which were negative for \textit{SCN1A} testing and most of whom had no family history. Out of the 11 PCDH19-DS families only 1 had multiple generations of affected individuals with one instance of an affected female having children. However, ascertainment bias might not be the only explanation for the difference in the severity of the phenotype in these two disorders. The differences in the type or position of the mutations can also be a contributing factor.

We have determined that there was no difference in the type of mutation in \textit{PCDH19}, which would account for the phenotype differences between EFMR and PCDH19-DS. This is not
surprising given our findings, that there is no obvious difference in the phenotype of EFMR patients with protein truncation and missense mutations (Appendix A, Dibbens et al., 2008).

Our analysis suggests that the positioning of the missense mutation in \textit{PCDH19} cannot account for the differences in the clinical presentation between EFMR and PCDH19-DS (Depienne et al., 2009a, Dibbens et al., 2008, Juberg and Hellman, 1971, Ryan et al., 1997, Scheffer et al., 2008). There is no obvious difference in the position of \textit{PCDH19} mutations either across exon 1 or within the EC repeat domains. There is no apparent genotype-phenotype correlation emerging in individuals with mutations in \textit{PCDH19} (Depienne et al., 2009a, Dibbens et al., 2008, Ryan et al., 1997, Scheffer et al., 2008, Juberg and Hellman, 1971).

\section*{4.8.3 Overall conservation and mutation positions in \textit{PCDH19}}

Missense mutations identified in \textit{PCDH19} in EFMR and PCDH19-DS patients affect highly conserved and functionally important amino acids. Extending our previous suggestion that missense mutations, identified in EFMR families, result in functional knockout of \textit{PCDH19} (Appendix A, Dibbens et al., 2008), I now include missense mutations identified in PCDH19-DS patients (Depienne et al., 2009a).

The analysis of the conservation questions the function of the amino acids involved in the hydrophobic cluster of protocadherins. Morishita \textit{et al} identified 8 amino acids in Pcdhα4 as important residue for the formation of the hydrophobic cluster (Morishita et al., 2006). Identification of the hydrophobic cluster residues was based on the previous work, identifying a deep hydrophobic pocket in the EC1 domain of a number of classical cadherins (Patel et al., 2006). The pocket is important in homophilic binding of classical cadherins to one another (Patel et al., 2006). The amino acids in PCDH19 that correspond to the hydrophobic cluster...
are highly conserved across the EC repeats and throughout the δ-protocadherin sub-family, in terms of their conservation score. However a number of these residues are not actually conserved as hydrophobic amino acids across the EC repeats. This would suggest that these residues are important for correct functioning of protocadherins but they have an alternative function. This notion is supported by protein structure analysis of protocadherins. Whilst the hydrophobic region is important in classical cadherins it does not serve the same function in protocadherins (reviewed in Morishita and Yagi, 2007), with the size of the hydrophobic pocket being significantly reduced in protocadheins (Morishita et al., 2006). The absence of a hydrophobic pocket agrees with the findings that whilst protocadherins are involved in high specificity, homophilic, Ca\textsuperscript{2+} dependent cell-cell adhesion (Obata et al., 1995, Hirano et al., 1999, Phillips et al., 2003, Bradley et al., 1998), the cell-cell adhesion interactions are weaker between protocadherins than between classical cadherins (Hirano et al., 1999, Obata et al., 1995, Redies et al., 2005, Sago et al., 1995, Sano et al., 1993, Yamagata et al., 1999, Yoshida, 2003).

Whilst the hydrophobic clusters are not conserved in PCDH19 there are two regions of highly conserved amino acids in close vicinity to the hydrophobic amino acids. Three of the missenese mutations identified in PCDH19 affect amino acids in conserved regions indicating that these regions are important. The first of the conserved regions contains the DXD calcium-binding domain (Kirov et al., 2003, Wu and Maniatis, 2000) suggesting a role of this domain in calcium binding.

The DXD motif was not identified as a calcium binding domain by Morishita et al (Morishita et al., 2006). This calcium binding motif differs from those identified in Pcdha4 (Morishita et al., 2006), as it is situated in the βB-βC loop at the N-terminal of the EC domain. All of the calcium binding residues identified by Morishita et al are located in the C-terminal loops of
the EC repeats (as illustrated in figure 4.6) (Morishita et al., 2006). The positioning of the
DXD motif suggests that both the N-terminal and C-terminal loops of the EC repeats are
involved in calcium binding. Given that the six EC repeat domains adjoin one another, with
the C-terminal end of one EC domain attaching to the N-terminal end of the next EC repeat it
is not surprising that the N-terminal end is also important for binding calcium.

The conservation of the N-terminal and C-terminal calcium-binding motifs supports the
notion that both regions work in combination to bind calcium. The C and N-terminal calcium-
binding motifs are only conserved where there is a N-terminal and a C-terminal end adjoined
one another. The DXD motif in the N-terminal end is invariant across the EC domains 2 – 6,
but is not conserved in the EC1 repeat domain (Kirov et al., 2003, Wu and Maniatis, 2000).
Conversely, the C-terminal calcium binding residues are highly conserved across EC1-EC5
but are not conserved in the EC6 domain. The explanation as to why the DXD motif was not
identified by Morishita et al (Morishita et al., 2006) can be attributed to the fact that they only
investigated the EC1 domain of Pedhα4, where the N-terminal domain is not involved in
calcium binding.

Our analysis of the conservation of the EC repeats revealed a second highly conserved cluster
of amino acids that is situated in close proximity to the second cluster of hydrophobic amino
acids. This cluster of amino acids is only conserved in EC domains 2-6 and is predicted to be
located close to and within the βF-βG loop at the N-terminal end of EC. This cluster does not
contain a DXD motif however similarities in the amino acid make up, positioning and the
high level of conservation of these amino acids suggests that this region could also be
important in calcium binding.
The localisation of 7 out of the 8 missense mutations identified in PCDH19 to the N- and C-terminal ends of the EC repeat further emphasises the importance of both of these regions. Disruption to the formation of this calcium binding conformation in any of the C-terminal to N-terminal combinations involved in calcium bind is sufficient to disrupt correct functioning of PCDH19 and subsequently result in clinical presentations of EFMR or PCDH19-DS. The missense mutation that does not localise to the N- or C-terminal ends of the EC repeat domain was identified in an EFMR family (Dibbens et al., 2008). This mutated residue is situated within the βG strand in close proximity to a calcium binding domain (figure 4.7 & 4.8). The introduction of a Glutamic acid in this position we believe will disrupt the formation of the βG strand and the overall formation of the β sheet, thereby disrupting the formation of the C-terminal calcium-binding pocket.

These results also raises the question of what phenotype (if any) do mutations in the less conserved regions of the EC repeat domain cause? The lack of population-based variation in the extra cellular domain of PCDH19 (Appendix B, Hynes et al., 2009) suggests that this entire region is important for the correct functioning of PCDH19. However mutations in these less conserved regions have not been identified in any of the cohorts screened. This may imply that additional genetic conditions remain to be associated with PCDH19 mutations in the less conserved regions.

4.9 Conclusion

We have identified an additional large EFMR family with a mutation in the extracellular domain of PCDH19. We have also shown that there is no apparent difference in the type or positioning of mutations in PCDH19 that could account the differences in the severity of
phenotype caused by mutations in PCDH19 in PCDH19-DS and EFMR patients, thus, there is no genotype-phenotype relationship in PCDH19 mutations. Analysis of the additional missense mutations identified in PCDH19 in the PCDH19-DS patients revealed that they are likely to result in functional knock out of PCDH19, this has previously identified in the EFMR patients (Appendix A and B, Dibbens et al., 2008, Hynes et al., 2009). The positioning of these mutations in combination with conservation analysis of amino acids in the EC repeat domains has revealed a likely role for the N-terminal and C-terminal ends in calcium binding and correct functioning of PCDH19.
4.10 Contributions

4.10.1.1 Section 1

With respect to the publication in the Journal of Medical Genetics I was involved in sending all of the samples to Sanger for sequence analysis and co-ordinating and compiling the results obtained from all of the sequencing performed. I performed the assessment of the likely functional consequence of all variants and mutations identified in this follow up screening. I also performed the conservation analysis of \textit{PCDH19} and was responsible for the generation of all figures for this paper. I wrote the first draft of this publication and coordinated successive drafts. I. Scheffer (Epilepsy Research Centre and Department of Medicine, University of Melbourne) primarily wrote the clinical description of affected females, with assistance from the referring clinician D. Gill (TY Nelson Department of Neurology, The Children’s Hospital Westmead, Sydney).

P. Tarpey, R. Smith and M. Stratton (Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK). performed the \textit{PCDH19} sequencing. M. Bayly (Department of Genetic Medicine, Woman’s and Children’s Hospital) also performed some \textit{PCDH19} sequencing.

\textbf{Patients and samples for \textit{PCDH19} screening were contributed by;}

\textbf{Epilepsy +/- ID cohort} – I. Scheffer, S Berkovic (Epilepsy Research Centre and Department of Medicine, University of Melbourne) and D. Gill (TY Nelson Department of Neurology, The Children’s Hospital Westmead, Sydney) referred the affected females for \textit{PCDH19} screening and provided clinical information. S. Turner and T. Desai (Epilepsy Research Centre and Department of Medicine, University of Melbourne) co-ordinated assembly of the Epilepsy cohort. G. Turner (GOLD Service, Hunter Genetics, Waratah, New South Wales), E.
Hann (School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide) and Z. Al Raisi also referred patients for screening.

**Rett syndrome cohort** – J. Christodulu (Western Sydney Genetics Program, The Children’s Hospital at Westmead University of Sydney) and H. Leonard (Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia) provided patient material for the Rett syndrome cohort.

**Autistic cohort** – Sample from the Autism Genetic Resource Exchange were selected by J. Gecz (SA Pathology at the Women’s and Children’s Hospital, Adelaide, Australia). N. Brown (Department of Paediatrics, University of Melbourne, Australia) also contributed some patient material for part of the autistic cohort.

**4.10.1.2 Section 2**

I performed the work presented in section 2, although G. Seiboth (SA Pathology at the Women’s and Children’s Hospital, Adelaide, Australia) was involved under my supervision in screening the large USA family as part of her laboratory placement.

The additional large USA family was referred by C. Schwartz (Greenwood Genetic Centre, Greenwood USA). Our laboratory performed the initial mutation discovery, with the help of G. Seiboth. The segregation testing of the rest of the family as well as additional screening of the ethnic matched controls was performed in C. Schwartz’s laboratory (Greenwood Genetic Centre, Greenwood USA).
Chapter 5 – Microarray expression profiling of EFMR subjects

5.1 INTRODUCTION

5.2 DEFINITIONS OF SPECIFIC TERMS USED IN THIS CHAPTER

5.3 RESULTS

5.3.1 EFMR FEMALE VS CONTROL FEMALE ANALYSIS

5.3.2 Transmitting males with mutations in PCDH19 vs control males

5.3.3 Analysis of EFMR females vs all controls (males and females)

5.3.4 Analysis of control males vs control females

5.4 DISCUSSION

5.4.1 Functional annotation and enriched cluster analysis of the EFMR vs control female analysis

5.4.2 Analysis of the transmitting males vs control males.

5.4.3 Mutations in PCDH19 results in a male-like expression pattern

5.4.4 Estrogen and epilepsy

5.4.5 Estrogen defects associated with disease

5.4.6 PCDH19 /Estrogen receptor (ER) tissue specific interaction

5.4.7 Estrogen and the male brain

5.4.8 Sexual dimorphism and disease

5.5 CONCLUSIONS

5.6 CONTRIBUTIONS
5.1 Introduction

In light of the identification of mutations in *PCDH19* as causative for EFMR we wanted to enhance understanding of the molecular function of *PCDH19* and how mutations in *PCDH19* cause epilepsy and ID. Prior to commencement of this investigation the only published information about *PCDH19* included a publication describing the identification of *PCDH19* as a member of the δ-protocadherin subfamily (Wolverton and Lalande, 2001) and a second publication describing the expression pattern of *Pcdh19* during mouse embryonic development (Gaitan and Bouchard, 2006). We decided to perform microarray expression profiling on skin fibroblast cells obtained from male and female subjects with mutations in *PCDH19*. The aim of this analysis was to use the mutations identified in *PCDH19* to explore what other genes and pathways were disrupted as a result of *PCDH19* mutations and how these relate to epilepsy and ID. The inclusion of transmitting males with mutations in *PCDH19* was performed to explore if any insight could be gained into a possible compensatory mechanism/gene(s) involved in rescuing transmitting males from epilepsy and ID.

5.2 Definitions of specific terms used in this chapter

Principal component analysis (PCA) – Is a statistical method used to express data in such a way that highlights their similarities and differences. It is used to describe the structure of high dimensional data by reducing its dimensionality. PCA involves linear transformation of the data that converts *n* original variables to *n* new variables which are ordered by the amount of variance explained, are uncorrelated and explain all of the variation in the data. The outcome of this analysis is a principal axis rotation of the original data. PCA analysis is useful in analysis of microarray expression data as this method is capable of handling the large
amounts of data generated and is capable of identifying broad patterns of expression alteration (Hilsenbeck et al., 1999).

**Differentially expressed** – refers to genes which have a significantly different gene expression levels between two groups of subjects being assessed, for example, between affected EFMR females and control females.

**Factor** – Is a variable that causes or influences other variables. For the microarray analysis the primary factor being interrogated was status (as described below). Some investigations also involved the use of a subject’s gender as the factor under interrogation.

**Status** – In the context of this analysis status refers to the presence or absence of a *PCDH19* mutation. Four different annotations of status were involved in this analysis: affected EFMR female with a mutation in *PCDH19*, control female with no *PCDH19* mutation, transmitting male with a mutation in *PCDH19* and control male with no *PCDH19* mutation.

**DAVID analysis** – The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to ascertain the biological meaning of the microarray output. DAVID is a bioinformatics resource that extracts biological features/meaning associated with large gene lists (Huang da et al., 2009).

**Functional annotation**– Genes submitted to DAVID, as part of a gene list being analysed, are each assigned “terms” which relate to the classification of a gene with respect to the gene ontology, protein domains and functional categories assigned to a gene. Annotation of genes in DAVID involves integration of information obtained from more than 40 publicly available annotation categories (Huang da et al., 2009, Sherman et al., 2007).
**Enrichment analysis** – One of the major functions of DAVID analysis is the use of enrichment analysis. Enrichment analysis uses the biological annotation (“terms”) associated with each gene to highlight the most overrepresented biological annotation in a submitted gene list. This aids identification of the biological processes that are most significant to the gene list under investigation.

**Functional annotation clustering** – Functional annotation clustering is a means of viewing the outcome of the enrichment analysis. The functional annotation cluster analysis employs a number of statistical algorithms to identify and measure relationship between genes with enriched terms. The outcome of this analysis is the ascertainment of groups of genes with associated annotation terms called enriched annotation clusters. The theory behind this analysis is that the more common genes annotations that are shared, the higher chance they will be grouped together.

**Enriched annotation cluster / enrichment score** – An enriched annotation cluster is the outcome generated from the process of functional annotation clustering. An enriched cluster contains a group of genes that have associated functional annotation terms. The enrichment score of a cluster is the geometric mean (in -log scale) of the P-values of each annotation term in an annotation cluster. These clusters are referred in the text as annotation clusters.

**Fold enrichment** – Is a measure of the magnitude of enrichment of a particular “term” within a submitted gene list when compared to the assigned background.

**EFMR** – throughout this chapter refers to affected EFMR females with a mutation in *PCDH19*. 

Chapter 5- 194
5.3 Results

5.3.1 EFMR female vs control female analysis

To identify the molecular consequences of mutations in *PCDH19*, microarray expression profiling was performed on EFMR female (n=6) and unaffected female (n=3) skin fibroblast cell lines. Principal component analysis (PCA) of the overall global expression profile of the samples analysed revealed that the control female samples clustered together and away from the EFMR females (figure 5.1). This clustering was rather modest with overlap occurring between the two groups.

![Figure 5.1 Principal component analysis of the EFMR vs control female array data](image)

These graphs illustrate the results obtained from the PCA analysis of the EFMR vs control female analysis. Ellipsoids have been added in the right-hand graph to aid visualisation of the clusters. Red circles and ellipsoid illustrate EFMR females and blue circles and ellipsoid represent control females. Graphs were generated using Partek V6.3.
Source of variation analysis identified that status (where status is defined as affected or unaffected with respect to EFMR is the factor being interrogated) accounted for more of the variation than error (figure 5.2). This analysis was performed on the full one way analysis of variance (ANOVA) results prior to any P-value or fold change cut-off being applied. Source of variation analysis of the P<0.01 and P<0.05 gene lists generated from one-way ANOVA of status, demonstrated that status contributed highly to the variation in expression. These results gave us confidence in the array data obtained.

**Figure 5.2 Source of variation results for the EFMR vs control female analysis**

These graphs represent the results obtained from source of variation analysis of the 1way ANOVA analysis performed on the EFMR vs control female data. Red bars illustrate the source of variation present in the factor status which is the factor being interrogated in the ANOVA analysis. Blue bars illustrate the source of variation attributed to error. The left hand graph represents the results obtained from the full ANOVA analysis. The middle graph was generated from the P<0.01 gene list, and the right hand graph from the P<0.05 gene list. Graphs were generated using Partek V6.3.

Dot-plot representation of the expression levels of *PCDH19* demonstrated that there is no significant difference in the expression level of *PCDH19* between the EFMR and control females (figure 5.3). This was expected based on our previous RT-qPCR analysis of *PCDH19* expression in these individuals (see section 3.3.6.1.1).
Figure 5.3 *PCDH19* expression in the EFMR and control female samples

Dot-plot representation of the expression level of *PCDH19* present in the EFMR and control females. Red circles and box and whisker graph represent the results for the affected EFMR females. Blue circles and box represent the results for the control female samples. Graphs were generated using Partek V6.3.

One-way ANOVA analysis of status identified 169 differentially expressed genes with a P-value <0.01. Of these 169 genes 75 were down regulated and 94 genes were up regulated in the EFMR females. The largest degree of down regulation was -12.12 fold and the largest up-regulation of 9.9 fold. Removing genes with an expression fold change difference of less than 1.5 left 110 differentially expressed genes (Appendix table C1). Fifty were up regulated and 60 down regulated genes. Increasing the P-value to <0.05 identified 753 differentially expressed genes. Removing genes with an expression fold change difference of less than 1.5 left 383 differentially expressed genes (Appendix table C1). Two hundred and six were up regulated and 177 down regulated.

The fold change cut off of +/- 1.5 was chosen as a conservative cut off. We know from our analysis of the *PCDH19* mutations that they are all severe mutations which we predict will
result in complete loss of function of PCDH19 (Appendix A, Dibbens et al., 2008). However the EFMR subjects do still have one functioning copy of the gene (Appendix A, Dibbens et al., 2008). One of the most important considerations for determining the success of DAVID analysis (described below) is the quality of the gene list submitted. With the suggestion of using a fold change cut off of ≥2 (Huang da et al., 2009). Given the heterozygous nature of the mutations we used the more conservative cut off of ≥1.5 fold change.

Functional annotation clustering analysis of the P<0.01 gene list performed using DAVID under the high stringency classification identified 12 clusters of genes with an enrichment score of over 1.3. The enrichment score is calculated from the geometric mean of all of the enrichment P-values for each annotation term associated with the gene members in the group. The average P-value is then minus log transformed to obtain the relative enrichment score. An enrichment score of above 1.3 represents a P-value below 0.05 (Huang da et al., 2009).

Out of the 12 enriched annotation clusters 7 were discounted on the basis of having the majority of functions in the enrichment cluster with a fold enrichment score of above 4. Fold enrichment scores of greater than 1.5 are considered interesting. However, this comes with a major caveat when analyzing small numbers of genes. The fold enrichment score measures the magnitude of enrichment present in a submitted list compared to the background gene list used (Affymetrix – HuEx-1_0-st-v2). It is suggested that with low numbers of genes (<100) a fold enrichment score of above 3 should be treated with caution as the reliability of fold enrichment scores decreases with small gene lists (Huang da et al., 2009). Given that the gene list being analysed contained just over 100 genes we set an upper cut off for the fold enrichment score of 4. This left five potentially interesting enriched annotation clusters (table 5.1).
<table>
<thead>
<tr>
<th>Cluster #</th>
<th>Enrichment score</th>
<th>Functions</th>
<th>P-value</th>
<th>Fold enrichment</th>
<th>Benjamini correction</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.28</td>
<td>Microtubule cytoskeleton</td>
<td>2.5E-05</td>
<td>5</td>
<td>1.1E-02</td>
<td>KIF23, PLK1, PRC1, TOP2A, DLG7, CDC6, TTK, KIF2C, LMNB1, FAM110B, CDC42EP3, KIF4A, GAS7, SPAG5, CCNB2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cytoskeletal part</td>
<td>9.6E-04</td>
<td>3.1</td>
<td>1.5E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cytoskeleton</td>
<td>6.2E-03</td>
<td>2.2</td>
<td>5.9E-01</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.12</td>
<td>Adenyl ribonucleotide binding</td>
<td>3.5E-03</td>
<td>2.2</td>
<td>9.7E-01</td>
<td>KIF23, RAB43, PLK1, FAM83D, ABCA6, ABCA9, TOP2A, GPR133, DAPK2, CDC6, TTK, KIF2C, SACS, MCM5, KIF4A, MAP3K8, RFC3, MET, SAR1A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Purine ribonucleotide binding</td>
<td>4.7E-03</td>
<td>2</td>
<td>9.7E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ribonucleotide binding Adenyl</td>
<td>4.7E-03</td>
<td>2</td>
<td>9.7E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATP-binding</td>
<td>5.8E-03</td>
<td>2.1</td>
<td>9.4E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATP binding</td>
<td>7.5E-03</td>
<td>1.9</td>
<td>9.5E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Purine nucleotide binding</td>
<td>7.6E-03</td>
<td>2.1</td>
<td>9.4E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nucleotide-binding</td>
<td>8.9E-03</td>
<td>2.2</td>
<td>7.4E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nucleotide binding</td>
<td>1.3E-02</td>
<td>2</td>
<td>7.7E-01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.67</td>
<td>Pyrophosphatase activity</td>
<td>1.7E-02</td>
<td>2.7</td>
<td>9.8E-01</td>
<td>RAB43, MCM5, ABCA6, ABCA9, TOP2A, RFC3, ENPP1, CDC6, SAR1A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides</td>
<td>1.8E-02</td>
<td>2.7</td>
<td>9.8E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hydrolase activity, acting on acid anhydrides</td>
<td>1.8E-02</td>
<td>2.7</td>
<td>9.8E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nucleoside-triphosphatase activity</td>
<td>3.7E-02</td>
<td>2.5</td>
<td>1.0E+00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.66</td>
<td>Generation of neurons</td>
<td>8.8E-03</td>
<td>3.9</td>
<td>8.4E-01</td>
<td>LAMB1, ITGA3, NR2F2, NTF3, EN1, ITGA6, MET, EFNA5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neurogenesis</td>
<td>1.3E-03</td>
<td>3.6</td>
<td>9.0E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cell migration</td>
<td>2.4E-02</td>
<td>3.6</td>
<td>9.6E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cell motility</td>
<td>4.3E-02</td>
<td>2.7</td>
<td>9.9E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Localization of cell</td>
<td>4.3E-02</td>
<td>2.7</td>
<td>9.9E-01</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.61</td>
<td>System development</td>
<td>9.2E-03</td>
<td>1.8</td>
<td>8.5E-01</td>
<td>LAMB1, DDIT3, CDON, ITGA3, FZD6, TNFAIP2, PPL, ITGA6, ENPP1, KLF4, MDK, CDC42EP3, PDGFC, SECTMI, NR2F2, NTF3, EN1, GAS7, HDAC9, MET, BCL3, EFNA5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organ development</td>
<td>1.1E-02</td>
<td>2</td>
<td>8.6E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anatomical structure development</td>
<td>3.3E-02</td>
<td>1.6</td>
<td>9.8E-01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multicellular organismal development</td>
<td>1.1E-01</td>
<td>1.4</td>
<td>1.0E+00</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.1 Functional annotation analysis of P<0.01 EFMR vs control female analysis

Enriched annotation clusters identified by DAVID functional annotation cluster analysis of the EFMR vs control female P<0.01 genes list. This was obtained using the high stringency classification in DAVID after removal of enriched clusters where multiple functions have a fold enrichment score above 4. Duplication of a function arises from the use of multiple databases of gene function / interaction including SP_PIR_KEYWORDS and GOTERM_BP_ALL. The enriched annotation cluster of particular interest is in bold.
Of particular interest with regard to the EFMR phenotype was the identification of an enriched annotation cluster of genes involved in neuron function and localization. This annotation cluster (cluster 4, table 5.1) had an enrichment score of 1.66 (P-value of 0.022) and comprised 5 functions: generation of neurons, neurogenesis, cell migration, cell motility and localization of cell. Eight genes from the submitted list of significant genes were involved in these functions; \textit{LAMB1, ITGA3, NR2F2, NTF3, EN1, ITGA6, MET} and \textit{EFNA5} (detailed in tables 5.1 and 5.2). 2D view demonstrated the positive and negative associations between the functions and genes identified (figure 5.4). None of these functions held up statistically to Benjamini-Hochberg correction (Benjamini correction), this is a step-down procedure for calculating the false discovery rate (FDR). Repeating this analysis using the less stringent medium classification identified additional functions and genes involved in this enriched annotation cluster. An additional 18 functions were identified in this analysis as outlined in table 5.3, and seven additional genes; \textit{CDON, FZD6, MDK, CDC42EP3, PDGFC, GAS7} and \textit{PPFIBP1} (figure 5.5 and table 5.2). The function, cell localisation and evidence of brain expression were investigated for all of these genes in order to assess their likely relevance to the EFMR phenotype and \textit{PCDH19} function (table 5.4). The majority of these proteins are located in the extracellular matrix or cell membrane. All of the genes investigated had evidence of brain expression as stated in the gene information databases associated with the DAVID program (http://david.abcc.ncifcrf.gov/home.jsp).
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>gene_assignment</th>
<th>P-value (Status)</th>
<th>Fold-Change</th>
<th>Related genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFNA5</td>
<td>Ephrin-A5</td>
<td>0.0002</td>
<td>-2.91</td>
<td>NTF3, GAS7, LAMB1</td>
</tr>
<tr>
<td>MET</td>
<td>Met proto-oncogene (hepatocyte growth factor receptor)</td>
<td>0.0003</td>
<td>3.90</td>
<td>TTK, MAP3K8, DAPK2, PLK1</td>
</tr>
<tr>
<td>LAMB1</td>
<td>Laminin, beta 1</td>
<td>0.0009</td>
<td>2.19</td>
<td>None</td>
</tr>
<tr>
<td>NTF3</td>
<td>Neurotrophin 3</td>
<td>0.0014</td>
<td>1.92</td>
<td>DDIT3</td>
</tr>
<tr>
<td>NR2F2</td>
<td>Nuclear receptor subfamily 2, group F, member 2</td>
<td>0.0027</td>
<td>1.93</td>
<td>TCF19, KLF4, LMCD1, UHRF1, TTK, MSC</td>
</tr>
<tr>
<td>ITGA6</td>
<td>Integrin, alpha 6</td>
<td>0.0030</td>
<td>6.75</td>
<td>ITGA3</td>
</tr>
<tr>
<td>ITGA3</td>
<td>Integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3)</td>
<td>0.0042</td>
<td>4.15</td>
<td>ITGA6, IL10</td>
</tr>
<tr>
<td>EN1</td>
<td>Engrailed homeobox 1</td>
<td>0.0063</td>
<td>-1.83</td>
<td>None</td>
</tr>
<tr>
<td>GAS7</td>
<td>Growth arrest-specific 7</td>
<td>0.0004</td>
<td>-1.95</td>
<td>None</td>
</tr>
<tr>
<td>PDGFC</td>
<td>Platelet derived growth factor C</td>
<td>0.0031</td>
<td>3.89</td>
<td>None</td>
</tr>
<tr>
<td>CDON</td>
<td>Cdon homolog (mouse)</td>
<td>0.0035</td>
<td>-5.46</td>
<td>None</td>
</tr>
<tr>
<td>FZD6</td>
<td>Frizzled homolog 6 (Drosophila)</td>
<td>0.0059</td>
<td>1.72</td>
<td>GPR133, OXTR, IL10RB, KREMEN1</td>
</tr>
<tr>
<td>CDC42EP3</td>
<td>CDC42 effector protein (Rho GTPase binding) 3</td>
<td>0.0073</td>
<td>1.74</td>
<td>None</td>
</tr>
<tr>
<td>MDK</td>
<td>Midkine (neurite growth-promoting factor 2)</td>
<td>0.0079</td>
<td>1.68</td>
<td>None</td>
</tr>
<tr>
<td>PPFIBP1</td>
<td>PTPRF interacting protein, binding protein 1 (liprin beta-1)</td>
<td>0.0093</td>
<td>1.88</td>
<td>None</td>
</tr>
</tbody>
</table>

Table 5.2 Details of the genes identified in the functional annotation analysis (from table 5.1) with additional functions and genes involved in the enriched annotation cluster pulled out using the less stringent analysis from the EFMR vs control female analysis

List of genes identified in the neuron and cell movement enriched annotation cluster. Genes above the line were identified in the P<0.01 gene list under high stringency whilst those below the line were added as a result of the analysis using the less stringent medium classification setting. Fold change = EFMR female compared to control female. Related genes were obtained from DAVID. P-values and fold change data was obtained from the microarray results.
Figure 5.4 Matrix representations of the genes and biological functions present in the enriched annotation cluster #4 from table 5.1

Matrix representation of the enriched annotation cluster of interest identified by high stringency functional annotation cluster analysis performed on the P<0.01 gene list generated from the EFMR vs control female analysis. The matrix represents the biological function and genes present in the enriched annotation cluster. Teal squares represent where a corresponding gene-function association has been reported in the literature. Purple squares represent where a corresponding gene-function association has not been reported yet, as stipulated by the gene function databases associated with the DAVID program (http://david.abcc.ncifcrf.gov/home.jsp).
<table>
<thead>
<tr>
<th>Functions</th>
<th># of genes</th>
<th>P-value</th>
<th>Fold enrichment</th>
<th>Benjamini correction</th>
<th>Genes Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation of neurons</td>
<td>7</td>
<td>8.8E-03</td>
<td>3.9</td>
<td>8.4E-01</td>
<td>LAMB1, CDON, ITGA3, FZD6</td>
</tr>
<tr>
<td>Neurogenesis</td>
<td>7</td>
<td>1.3E-03</td>
<td>3.6</td>
<td>9.0E-01</td>
<td>ITGA6, CDC42EP</td>
</tr>
<tr>
<td>Nervous system development</td>
<td>11</td>
<td>2.1E-02</td>
<td>2.3</td>
<td>9.5E-01</td>
<td>MDK, NR2F2, GAS7</td>
</tr>
<tr>
<td>Cell migration</td>
<td>6</td>
<td>2.4E-02</td>
<td>3.6</td>
<td>9.6E-01</td>
<td>EN1, EFNA5</td>
</tr>
<tr>
<td>Neuron migration</td>
<td>3</td>
<td>3.6E-02</td>
<td>10</td>
<td>9.9E-01</td>
<td>EN1, EFNA5</td>
</tr>
<tr>
<td>Localization of cell</td>
<td>7</td>
<td>4.3E-02</td>
<td>2.7</td>
<td>9.9E-01</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell motility</td>
<td>7</td>
<td>4.3E-02</td>
<td>2.7</td>
<td>9.9E-01</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Focal adhesion</td>
<td>5</td>
<td>6.8E-02</td>
<td>3.1</td>
<td>1.0E+00</td>
<td>NTF3, GAS7, MET, PPFIBP1, EFNA5</td>
</tr>
<tr>
<td>Brain development</td>
<td>4</td>
<td>7.4E-02</td>
<td>4.1</td>
<td>1.0E+00</td>
<td>PPFIBP1, EFNA5</td>
</tr>
<tr>
<td>Central nervous system development</td>
<td>5</td>
<td>7.5E-02</td>
<td>3.1</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Basolateral plasma membrane</td>
<td>3</td>
<td>1.0E-01</td>
<td>5.5</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Small cell lung cancer</td>
<td>3</td>
<td>1.5E-01</td>
<td>4.2</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>ECM-receptor interaction</td>
<td>3</td>
<td>1.6E-01</td>
<td>4.1</td>
<td>9.9E-01</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Neuron development</td>
<td>3</td>
<td>3.3E-01</td>
<td>2.5</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell morphogenesis</td>
<td>5</td>
<td>3.5E-01</td>
<td>1.6</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cellular structure morphogenesis</td>
<td>5</td>
<td>3.5E-01</td>
<td>1.6</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell projection morphogenesis</td>
<td>3</td>
<td>4.2E-01</td>
<td>2.1</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell part morphogenesis</td>
<td>3</td>
<td>4.2E-01</td>
<td>2.1</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell projection organization and biogenesis</td>
<td>3</td>
<td>4.2E-01</td>
<td>2.1</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Neuron differentiation</td>
<td>3</td>
<td>4.6E-01</td>
<td>1.9</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>6</td>
<td>4.8E-01</td>
<td>1.3</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Biological adhesion</td>
<td>6</td>
<td>4.8E-01</td>
<td>1.3</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>3</td>
<td>6.4E-01</td>
<td>1.4</td>
<td>1.0E+00</td>
<td>EFNA5</td>
</tr>
</tbody>
</table>

Table 5.3 Functions and genes identified in the enriched annotation cluster #4 of table 5.1. Identified under medium stringency in the EFMR vs control female analysis.

These results were obtained from DAVID functional annotation cluster analysis.
Figure 5.5 Matrix representations of the genes and biological functions present in the enriched annotation cluster of #4 of table 5.1

Matrix representation of the enriched annotation cluster of interest identified by medium stringency functional annotation cluster analysis performed on the P<0.01 gene list generated from the EFMR vs control female analysis. The matrix represents the biological function and genes present in the enriched annotation cluster. Teal squares represent where a corresponding gene-function association has been reported in the literature. Purple squares represent where a corresponding gene-function association has not been reported yet, as stipulated by the gene function databases associated with the DAVID program (http://david.abcc.ncifcrf.gov/home.jsp).
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Function (PIR_Summary - DAVID)</th>
<th>Cell localisation</th>
<th>Brain expression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EFNA5</strong></td>
<td>May function actively to stimulate axon fasciculation. Induces compartmentalized signalling within a caveolae-like membrane microdomain when bound to the extracellular domain of its cognate receptor. This signalling event requires the activity of the Fyn tyrosine kinase</td>
<td>Cell membrane; Lipid-anchor, GPI-anchor. Note=Compartmentalized in discrete caveolae-like membrane microdomains</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>MET</strong></td>
<td>Receptor for hepatocyte growth factor and scatter factor. Has a tyrosine-protein kinase activity</td>
<td>Cell surface</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>GAS7</strong></td>
<td>May play a role in promoting maturation and morphological differentiation of cerebellar neurons</td>
<td>Cytoplasm</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>LAMB1</strong></td>
<td>Binding to cells via a high affinity receptor, laminin is thought to mediate the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components</td>
<td>Secreted, extracellular space, extracellular matrix, basement membrane</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>NTF3</strong></td>
<td>Seems to promotes the survival of visceral and proprioceptive sensory neurons</td>
<td>Secreted</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>NR2F2</strong></td>
<td>Regulation of the apolipoprotein A-I gene transcription. Binds to DNA site A</td>
<td>Nucleus</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>ITGA6</strong></td>
<td>Integrin alpha-6/beta-1 is a receptor for laminin on platelets. Integrin alpha-6/beta-4 is a receptor for laminin in epithelial cells and it plays a critical structural role in the hemidesmosome</td>
<td>Membrane; Single-pass type I membrane protein</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>PDGFC</strong></td>
<td>Not available</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>CDON</strong></td>
<td>Component of a cell-surface receptor complex that mediates cell-cell interactions between muscle precursor cells. Promotes differentiation of myogenic cells (Hilschmann et al.)</td>
<td>Cell surface</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>ITGA3</strong></td>
<td>Integrin alpha-3/beta-1 is a receptor for fibronectin, laminin, collagen, epiligrin, thrombospondin and CSPG4. Alpha-3/beta-1 may mediate with LGALS3 the stimulation by CSPG4 of endothelial cells migration.</td>
<td>Membrane; Single-pass type I membrane protein</td>
<td>Yes, both isoforms A and B</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Function (PIR_Summary - DAVID)</td>
<td>Cell localisation</td>
<td>Brain expression</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>FZD6</strong></td>
<td>Receptor for Wnt proteins. Most of frizzled receptors are coupled to the beta-catenin canonical signalling pathway, which leads to the activation of disheveled proteins, inhibition of GSK-3 kinase, nuclear accumulation of beta-catenin and activation of Wnt target genes. A second signalling pathway involving PKC and calcium fluxes has been seen for some family members, but it is not yet clear if it represents a distinct pathway or if it can be integrated in the canonical pathway, as PKC seems to be required for Wnt-mediated inactivation of GSK-3 kinase. Both pathways seem to involve interactions with G-proteins. May be involved in transduction and intercellular transmission of polarity information during tissue morphogenesis and/or in differentiated tissues</td>
<td>Transmembrane</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>EN1</strong></td>
<td></td>
<td>Nucleus</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>CDC42EP3</strong></td>
<td>Probably involved in the organization of the actin cytoskeleton. May act downstream of CDC42 to induce actin filament assembly leading to cell shape changes. Induces pseudopodia formation in fibroblasts</td>
<td>Intracytoplasmic membrane; Peripheral membrane protein</td>
<td>Weak in brain</td>
</tr>
<tr>
<td><strong>MDK</strong></td>
<td>Not available</td>
<td>Secreted</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>PFIFBP1</strong></td>
<td>Not available</td>
<td>Secreted</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Table 5.4 Gene information of interesting genes identified in the enriched annotation cluster #4 from table 5.1 in the EFMR vs control female analysis**

Gene information was extracted from the databases associated DAVID (http://david.abcc.ncifcrf.gov/home.jsp).
Functional annotation and enriched annotation analysis in general has higher statistical power resulting in a higher sensitivity (more significant P-values) to slightly enriched terms as well as to more specific terms with larger gene lists (Huang da et al., 2009). Given that the P<0.01 gene list only contained 110 genes we decided to also analyse the P<0.05 gene list of 383 genes to try and gain statistical power. This analysis was used to support the findings from the P<0.01 list, as well as to identify additional genes and biological functions of interest.

Functional annotation cluster analysis of the P<0.05 gene list identified three enriched annotation clusters of interest with respect to EFMR (table 5.5). The enriched annotation clusters identified consisted of genes involved with cell/biological adhesion, extracellular matrix, and again, an enriched annotation cluster involving neuron function and localization. This analysis also identified an additional 16 genes involved in the neuron function and localization annotation cluster than were identified in the P<0.01 gene list analysis. The additional genes identified were \textit{NTN1}, \textit{WNT2}, \textit{TWIST1}, \textit{NR2F2}, \textit{SLIT3}, \textit{FEZ1}, \textit{ITGA1}, \textit{ITGA11}, \textit{CSPG4}, \textit{STAT3}, \textit{HMMR}, \textit{PALM}, \textit{EPHA2}, \textit{EMX2}, \textit{CEBPB}, \textit{RACGAP1} and \textit{CDK6} (table 5.6 and figure 5.6).
### Table 5.5 Functional annotation analysis of P<0.05 EFMR vs control female analysis

Enriched annotation clusters of interest identified by DAVID functional annotation analysis of the EFMR vs control female P<0.05 genes list. Duplication of a function arises from the use of multiple databases of gene function / interaction including SP_PIR_KEYWORDS and GOTERM_BP_ALL.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>gene_assignment - DAVID</th>
<th>P-value (Status)</th>
<th>Fold-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>RACGAP1</td>
<td>Rac gtpase activating protein 1</td>
<td>0.0101</td>
<td>1.97</td>
</tr>
<tr>
<td>STAT3</td>
<td>Signal transducer and activator of transcription 3 (acute-phase response factor)</td>
<td>0.0110</td>
<td>-1.69</td>
</tr>
<tr>
<td>HMMR</td>
<td>Hyaluronan-mediated motility receptor (rhamm)</td>
<td>0.0119</td>
<td>1.66</td>
</tr>
<tr>
<td>FEZ1</td>
<td>Fasciculation and elongation protein zeta 1 (zygin i)</td>
<td>0.0146</td>
<td>-1.58</td>
</tr>
<tr>
<td>CEBPB</td>
<td>Ccaat/enhancer binding protein (c/ebp), beta</td>
<td>0.0149</td>
<td>-1.61</td>
</tr>
<tr>
<td>CSPG4</td>
<td>Chondroitin sulfate proteoglycan 4 (melanoma-associated)</td>
<td>0.0167</td>
<td>2.42</td>
</tr>
<tr>
<td>NTN1</td>
<td>Netrin 1</td>
<td>0.0179</td>
<td>-1.66</td>
</tr>
<tr>
<td>CDK6</td>
<td>Cyclin-dependent kinase 6</td>
<td>0.0221</td>
<td>2.39</td>
</tr>
<tr>
<td>ITGA1</td>
<td>Integrin, alpha 1</td>
<td>0.0262</td>
<td>2.51</td>
</tr>
<tr>
<td>PALM</td>
<td>Paralemmin</td>
<td>0.0263</td>
<td>-1.66</td>
</tr>
<tr>
<td>ITGA11</td>
<td>Integrin, alpha 11</td>
<td>0.0272</td>
<td>2.02</td>
</tr>
<tr>
<td>TWIST1</td>
<td>Twist homolog 1 (acrocephalosyndactyly 3; saethre- chotzen syndrome) (drosophila)</td>
<td>0.0309</td>
<td>-1.72</td>
</tr>
<tr>
<td>SLIT3</td>
<td>Slit homolog 3 (drosophila)</td>
<td>0.0326</td>
<td>-1.84</td>
</tr>
<tr>
<td>EPHA2</td>
<td>Eph receptor a2</td>
<td>0.0353</td>
<td>1.50</td>
</tr>
<tr>
<td>WNT2</td>
<td>Wingless-type mmtv integration site family member 2</td>
<td>0.0444</td>
<td>-1.66</td>
</tr>
<tr>
<td>EMX2</td>
<td>Empty spiracles homolog 2 (drosophila)</td>
<td>0.0449</td>
<td>-2.20</td>
</tr>
</tbody>
</table>

Table 5.6 Additional genes identified in the functional annotation analysis of interest in the EFMR vs control female analysis using the P<0.05 gene list

Additional genes identified in the neuron and cell movement enriched annotation cluster of interest from the p<0.05 list of differentially expressed genes from the EFMR vs control female analysis. Fold change = EFMR females compared to control females. P-values and fold change data was obtained from the microarray results.
Figure 5.6 Matrix representations of the genes and biological functions present in the enriched annotation cluster of interest

Matrix representation of the enriched annotation cluster of interest identified by medium stringency functional annotation cluster analysis performed on the P<0.05 gene list generated from the EFMR vs control female analysis. The matrix represents the biological function and genes present in the enriched annotation cluster. Teal squares represent where a corresponding gene-function association has been reported in the literature. Purple squares represent where a corresponding gene-function association has not been reported yet, as stipulated by the gene function databases associated with the DAVID program (http://david.abcc.ncifcrf.gov/home.jsp).
Even with the larger gene list none of the individual biological functions indentified in any of the enriched annotation clusters passed Benjamini correction (table 5.5). Whilst these biological functions were no longer statistically significant after statistical corrections they do still appear interesting because analysis of large gene lists comes down to more than a purely statistical solution. Analysis performed by the DAVID program is a powerful tool for analysing large gene lists but it does have its limitations. This analysis only looks for over representation of genes involved in a biological function. It does not take into account the P-value or fold change associated with the genes involved. It also does not consider biological knowledge of the disorder and or gene being analysed. This has the advantage of giving a non-biased assessment of the gene list; however, these factors do need to be considered in the interpretation of the results obtained. Therefore the biological functions and genes identified in these clusters were of biological interest.

Functional annotation analysis also identified four KEGG pathways of interest, which were enriched based on the significant gene list. These were; ECM-receptor interaction, adherens junction, focal adhesion and axon guidance. With P-values of; 0.037, 0.057, 0.095 and 0.16 respectively. None of these pathways remained significant after Benjamini correction with corrected P-values of 0.72, 0.77, 0.81 and 0.84. The KEGG pathways identified are illustrated in figure 5.7-5.10. The specific genes present in these pathways are described in table 5.7.
Figure 5.7 KEGG ECM-receptor interaction pathway

ECM-receptor interaction pathway as illustrated by the KEGG pathway diagram. Genes deregulated in the EFMR vs control female analysis P<0.05 gene list are illustrated by stars. Genes with increased expression in the EFMR females are illustrated in purple. The genes with decreased expression in EFMR females are illustrated in teal.
Figure 5.8 KEGG adherens junction pathway

Adherens junction pathway as illustrated by the KEGG pathway diagram. Genes deregulated in the EFMR vs control female analysis P<0.05 gene list are illustrated by stars. Genes with increased expression in the EFMR females are illustrated in purple. The genes with decreased expression in EFMR females are illustrated in teal.
Figure 5.9 KEGG focal adhesion pathway

Focal adhesion pathway as illustrated by the KEGG pathway diagram. Genes deregulated in the EFMR vs control female analysis P<0.05 gene list are illustrated by stars. Genes with increased expression in the EFMR females are illustrated in purple. The genes with decreased expression in EFMR females are illustrated in teal.
Figure 5.10 KEGG axon guidance pathway

Axon guidance pathway as illustrated by the KEGG pathway diagram. Genes deregulated in the EFMR vs control female analysis P<0.05 gene list are illustrated by stars. Genes with increased expression in the EFMR females are illustrated in purple. The genes with decreased expression in EFMR females are illustrated in teal. CXCR7 was not highlighted in this pathway by the DAVID analysis however, CXCR7 has been identified as an alternative receptor for CXCL12 (Burns et al., 2006) and is the closest relative to CXCR4. So it is included in this pathway, where CXCR4 is identified.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>gene_assignment</th>
<th>P-value (Status)</th>
<th>Fold-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAMB1</td>
<td>laminin, beta 1</td>
<td>0.00088</td>
<td>2.19</td>
</tr>
<tr>
<td>HMMR</td>
<td>hyaluronan-mediated motility receptor (RHAMM)</td>
<td>0.012</td>
<td>1.66</td>
</tr>
<tr>
<td>ITGA1</td>
<td>integrin, alpha 1</td>
<td>0.026</td>
<td>2.51</td>
</tr>
<tr>
<td>ITGA3</td>
<td>integrin, alpha 3</td>
<td>0.0042</td>
<td>4.15</td>
</tr>
<tr>
<td>TNXB</td>
<td>tenascin XB</td>
<td>0.021</td>
<td>-1.87</td>
</tr>
<tr>
<td>ITGA11</td>
<td>integrin, alpha 11</td>
<td>0.027</td>
<td>2.02</td>
</tr>
<tr>
<td>ITGA6</td>
<td>integrin, alpha 6</td>
<td>0.003</td>
<td>6.75</td>
</tr>
<tr>
<td>SNAI1</td>
<td>snail homolog 1</td>
<td>0.046</td>
<td>-1.88</td>
</tr>
<tr>
<td>MET</td>
<td>met proto-oncogene</td>
<td>0.00032</td>
<td>3.90</td>
</tr>
<tr>
<td>TGFBR1</td>
<td>transforming growth factor, beta receptor I</td>
<td>0.01</td>
<td>-1.73</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>transcription factor 7-like 2</td>
<td>0.0039</td>
<td>-1.65</td>
</tr>
<tr>
<td>IGF1R</td>
<td>insulin-like growth factor 1 receptor</td>
<td>0.019</td>
<td>-2.13</td>
</tr>
<tr>
<td>SSX2IP</td>
<td>synovial sarcoma, X breakpoint 2 interacting protein</td>
<td>0.042</td>
<td>1.55</td>
</tr>
<tr>
<td>LAMB1</td>
<td>laminin, beta 1</td>
<td>0.00088</td>
<td>2.19</td>
</tr>
<tr>
<td>PDGFC</td>
<td>platelet derived growth factor C</td>
<td>0.0031</td>
<td>3.89</td>
</tr>
<tr>
<td>ITGA1</td>
<td>integrin, alpha 1</td>
<td>0.026</td>
<td>2.51</td>
</tr>
<tr>
<td>ITGA3</td>
<td>integrin, alpha 3</td>
<td>0.0042</td>
<td>4.15</td>
</tr>
<tr>
<td>TNXB</td>
<td>tenascin XB</td>
<td>0.021</td>
<td>-1.87</td>
</tr>
<tr>
<td>ITGA11</td>
<td>integrin, alpha 11</td>
<td>0.027</td>
<td>2.02</td>
</tr>
<tr>
<td>ITGA6</td>
<td>integrin, alpha 6</td>
<td>0.003</td>
<td>6.75</td>
</tr>
<tr>
<td>MET</td>
<td>met proto-oncogene</td>
<td>0.00032</td>
<td>3.90</td>
</tr>
<tr>
<td>IGF1R</td>
<td>insulin-like growth factor 1 receptor</td>
<td>0.019</td>
<td>-2.13</td>
</tr>
<tr>
<td>CCND1</td>
<td>cyclin D1</td>
<td>0.028</td>
<td>1.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>gene_assignment</th>
<th>P-value (Status)</th>
<th>Fold-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>MET</td>
<td>met proto-oncogene (hepatocyte growth factor receptor)</td>
<td>0.00032</td>
<td>3.90</td>
</tr>
<tr>
<td>ABL1</td>
<td>v-abl abelson murine leukemia viral oncogene homolog 1</td>
<td>0.033</td>
<td>-1.57</td>
</tr>
<tr>
<td>SLIT3</td>
<td>slit homolog 3 (drosophila)</td>
<td>0.033</td>
<td>-1.84</td>
</tr>
<tr>
<td>CXCL12</td>
<td>chemokine (c-x-c motif) ligand 12 (stromal cell-derived factor 1)</td>
<td>0.019</td>
<td>-1.66</td>
</tr>
<tr>
<td>EPHA2</td>
<td>eph receptor a2</td>
<td>0.035</td>
<td>1.50</td>
</tr>
<tr>
<td>NTN1</td>
<td>netrin 1</td>
<td>0.018</td>
<td>-3.08</td>
</tr>
<tr>
<td>EFNA5</td>
<td>ephrin-a5</td>
<td>0.00018</td>
<td>-2.91</td>
</tr>
</tbody>
</table>

Table 5.7 Genes from the EFMR vs control female analysis P<0.05 that are involved in the KEEG pathways of interest

List of the genes involved in the KEGG pathways identified of interest in the EFMR vs control females analysis P<0.05. The P-values and fold change were identified in the microarray analysis. Fold-change is EFMR female relative to control female.
Whilst these highlighted pathways did not remain statistically significant after statistical correction, their potential involvement in the disease phenotype is noteworthy. All of these pathways contain deregulated genes, in affected EFMR females, that interact with one another.

Functional annotation analysis of the significantly deregulated genes in EFMR females identified four biological pathways and three enriched annotation clusters of gene functions, which are of potential relevance to EFMR given their connection to PCDH19.

**Real time validation of the EFMR vs control female microarray results**

To validate the microarray data using an independent method, ten genes were selected for RT-qPCR validation. The genes selected were: ITGA3, ITGA6, LAMB1, MET, NTF3, PDGFC, EPHA2, EFNA5, GAS7 and CDON. Nine of these genes were selected from the list of genes with a P<0.01, while EPHA2 was selected from the list of genes with a P<0.05. Figure 5.11 illustrates the individual results obtained from this analysis for each gene. Figure 5.12 demonstrates the combined result for all the genes. Nine out of the ten genes validated with a P<0.05 from the real time analysis (figure 5.12). LAMB1 showed the same trend in the real time results as in the microarray with increased expression in EFMR females, but just missed the 0.05 cut off with a P-value of 0.054. These results verified the array results.
Figure 5.11 Real time validation of the EFMR vs control female analysis

Graphical representation of the RT-qPCR results obtained for each gene investigated. **Purple** bars are control female samples and **teal** bars the EFMR female samples. The y-axis represents the gene expression level standardised to Actin-β. Individual samples are plotted across the x-axis, with the particular *PCDH19* mutation indicated.
Figure 5.12 Real time validation combined results

Graphical representation of average RT-qPCR results obtained for each gene investigated. Purple bars represent the average gene expression from the n=3 control females. Teal bars represent the average gene expression from the n=6 EFMR females. Individual genes are plotted across the x-axis. The error bars represent the standard error. * represents a P-value of <0.05. ** = P<0.01.

5.3.2 Transmitting males with mutations in PCDH19 vs control males

Males with mutations in PCDH19 appear unaffected by EFMR. To search for subclinical effects at the biochemical level we performed a comparison of the expression profile between transmitting males (n=3) and control males (n=3). PCA of the overall global expression profile of the samples analysed revealed that there was no apparent differential clustering between the transmitting males and control males (figure 5.13).
Figure 5.13 PCA of the transmitting male vs control male analysis

This graph illustrates the results obtained from PCA analysis of the transmitting male vs control male analysis. Red circles illustrate transmitting males and blue circles represent control males. Graph generated using Partek V6.3.

1way ANOVA analysis identified 140 differentially expressed genes with a P-value <0.01(Appendix table C2). Of these 140 genes, 50 were down regulated and 90 genes were up regulated in the transmitting males. The largest degree of down regulation was -2.59 fold and the largest up-regulation was 2.23 fold. Removing genes with an expression fold change difference of less than 1.5 left us with a list of only 36 differentially expressed genes (table 5.8). Of the 36 genes 34 had a fold change of less than 2. Increasing the P-value to <0.05 identified 784 genes as differentially expressed. Removing genes with an expression fold change difference of less than 1.5 left 205 differentially expressed genes (Appendix table C2). 151 up-regulated and 54 down regulated.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene name</th>
<th>P-value (Status)</th>
<th>Fold-Change</th>
<th>Function (NCBI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFX1</td>
<td>Nuclear transcription factor, X-box binding 1</td>
<td>0.0001</td>
<td>1.50</td>
<td>Nuclear transcription factor</td>
</tr>
<tr>
<td>N6AMT2</td>
<td>N-6 adenine-specific DNA methyltransferase 2 (putative)</td>
<td>0.0005</td>
<td>1.71</td>
<td>Unknown</td>
</tr>
<tr>
<td>JMJD4</td>
<td>Jumonji domain containing 4</td>
<td>0.0007</td>
<td>-1.50</td>
<td>Unknown</td>
</tr>
<tr>
<td>LOC400590</td>
<td>Hypothetical LOC400590</td>
<td>0.0007</td>
<td>1.68</td>
<td>Unknown</td>
</tr>
<tr>
<td>UQCC</td>
<td>Ubiquinol-cytochrome c reductase complex chaperone, CBP3 ho</td>
<td>0.0009</td>
<td>1.56</td>
<td>Unknown</td>
</tr>
<tr>
<td>TRIM22</td>
<td>Ripartite motif-containing 22</td>
<td>0.0015</td>
<td>2.23</td>
<td>Development and cell growth</td>
</tr>
<tr>
<td>ARHGAP12</td>
<td>Rho GTPase activating protein 12</td>
<td>0.0017</td>
<td>1.57</td>
<td>GTPase-activating protein</td>
</tr>
<tr>
<td>HIST1H2AH</td>
<td>Histone cluster 1, H2ah</td>
<td>0.0020</td>
<td>-1.70</td>
<td>Unknown</td>
</tr>
<tr>
<td>HMG2L1</td>
<td>High-mobility group protein 2-like 1</td>
<td>0.0022</td>
<td>1.55</td>
<td>DNA binding and transcription</td>
</tr>
<tr>
<td>PARN</td>
<td>Poly(A)-specific ribonuclease (deadenylation nuclease)</td>
<td>0.0026</td>
<td>1.50</td>
<td>mRNA degradation</td>
</tr>
<tr>
<td>TMEM136</td>
<td>Transmembrane protein 136</td>
<td>0.0026</td>
<td>1.68</td>
<td>Unknown</td>
</tr>
<tr>
<td>PYGL</td>
<td>Phosphorylase, glycogen; liver (Hers disease, glycogen stor</td>
<td>0.0027</td>
<td>1.52</td>
<td>Glycogen phosphorylase</td>
</tr>
<tr>
<td>ABTB1</td>
<td>Ankyrin repeat and BTB (POZ) domain containing 1</td>
<td>0.0033</td>
<td>-1.62</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>LOC147804</td>
<td>Tropomyosin 3 pseudogene</td>
<td>0.0034</td>
<td>-2.59</td>
<td>Unknown</td>
</tr>
<tr>
<td>STAT3</td>
<td>Signal transducer and activator of transcription 3 (acute-</td>
<td>0.0034</td>
<td>1.54</td>
<td>Signal transducer and activator of transcription</td>
</tr>
<tr>
<td>EPB41L3</td>
<td>Erythrocyte membrane protein band 4.1-like 3</td>
<td>0.0034</td>
<td>1.66</td>
<td>Tumor suppressor</td>
</tr>
<tr>
<td>TOP2B</td>
<td>Topoisomerase (DNA) II beta 180kDa</td>
<td>0.0036</td>
<td>1.52</td>
<td>Alters DNA topology through a double-strand break</td>
</tr>
<tr>
<td>CCDC56</td>
<td>Coiled-coil domain containing 56</td>
<td>0.0038</td>
<td>1.78</td>
<td>Unknown</td>
</tr>
<tr>
<td>CCDC59</td>
<td>Coiled-coil domain containing 59</td>
<td>0.0044</td>
<td>1.70</td>
<td>Unknown</td>
</tr>
<tr>
<td>EPHA2</td>
<td>EPH receptor A2</td>
<td>0.0048</td>
<td>-1.52</td>
<td>Receptor tyrosine kinase</td>
</tr>
<tr>
<td>PIAS1</td>
<td>Protein inhibitor of activated STAT, 1</td>
<td>0.0050</td>
<td>1.52</td>
<td>Protein inhibitors of activated STAT</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Gene name</td>
<td>P-value (Status)</td>
<td>Fold-Change</td>
<td>Function (NCBI)</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>------------------</td>
<td>-------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>UVRAG</td>
<td>UV radiation resistance associated gene</td>
<td>0.0055</td>
<td>1.58</td>
<td>Tumor suppressor candidate</td>
</tr>
<tr>
<td>ESCO1</td>
<td>Establishment of cohesion 1 homolog 1 (S. cerevisiae)</td>
<td>0.0063</td>
<td>1.64</td>
<td>Sister chromatin cohesion</td>
</tr>
<tr>
<td>PATL1</td>
<td>Protein associated with topoisomerase II homolog 1 (yeast)</td>
<td>0.0065</td>
<td>1.74</td>
<td>Translation control</td>
</tr>
<tr>
<td>TRIM5</td>
<td>Tripartite motif-containing 5</td>
<td>0.0072</td>
<td>1.56</td>
<td>Cellular antiretroviral factor</td>
</tr>
<tr>
<td>EWSR1</td>
<td>Ewing sarcoma breakpoint region 1</td>
<td>0.0076</td>
<td>1.58</td>
<td>RNA-binding, transcriptional control</td>
</tr>
<tr>
<td>DENND4C</td>
<td>DENN/MADD domain containing 4C</td>
<td>0.0079</td>
<td>1.52</td>
<td>Unknown</td>
</tr>
<tr>
<td>ZCCHC6</td>
<td>Zinc finger, CCHC domain containing 6</td>
<td>0.0083</td>
<td>1.61</td>
<td>Possibly mRNA regulation</td>
</tr>
<tr>
<td>GFPT2</td>
<td>Glutamine-fructose-6-phosphate transaminase 2</td>
<td>0.0083</td>
<td>1.59</td>
<td>Hexosamine biosynthesis</td>
</tr>
<tr>
<td>SETX</td>
<td>Senataxin</td>
<td>0.0088</td>
<td>1.63</td>
<td>Putative DNA/RNA helicase</td>
</tr>
<tr>
<td>JARID1A</td>
<td>Jumonji, AT rich interactive domain 1A</td>
<td>0.0088</td>
<td>1.50</td>
<td>Demethylase enzyme</td>
</tr>
<tr>
<td>LTA4H</td>
<td>Leukotriene A4 hydrolase</td>
<td>0.0089</td>
<td>1.83</td>
<td>Bifunctional zinc enzyme with the activities of epoxide hydrolase and aminopeptidase</td>
</tr>
<tr>
<td>CENTA1</td>
<td>Centaurin, alpha 1</td>
<td>0.0090</td>
<td>-1.62</td>
<td>Phosphatidylinositol 3,4,5-triphosphate (PIP3)-binding protein</td>
</tr>
<tr>
<td>TIMM9</td>
<td>Translocase of inner mitochondrial membrane 9 homolog (yeast)</td>
<td>0.0091</td>
<td>1.57</td>
<td>Unknown</td>
</tr>
<tr>
<td>ZBTB41</td>
<td>Zinc finger and BTB domain containing 41</td>
<td>0.0093</td>
<td>1.52</td>
<td>Unknown</td>
</tr>
<tr>
<td>ISCU</td>
<td>Iron-sulfur cluster scaffold homolog (E. coli)</td>
<td>0.0096</td>
<td>1.51</td>
<td>Electron transfer, substrate binding/activation, iron/sulfur storage, regulation of gene expression, and enzyme activity</td>
</tr>
</tbody>
</table>

Table 5.8 Significantly altered genes (P<0.01) identified in the Transmitting male vs control male analysis

Genes identified as significantly deregulated between the transmitting males and control males with a P<0.01 and fold change of greater that +/- 1.5. P-values and fold changes were identified in the microarray analysis. Fold change is transmitting male relative to control male.
Functional annotation analysis using DAVID was performed on the P<0.05 gene list due to the low number (36) of genes identified in the P<0.01 list. This analysis failed to identify enrichment of any biological functions or pathways, which reached significance and were of relevance to mutations in *PCDH19*. The 36 genes identified in the P<0.01 gene list were investigated individually. Table 5.8 briefly describes these genes.

Comparison of the list of significantly altered genes at a P-value of <0.01 between the transmitting male vs control male analysis and the EFMR vs control female analysis revealed no overlap in the genes identified. Performing this comparison with the two P<0.05 gene lists identified four genes common to both lists; *AKR1C3, EPHA2, SH3BP5* and *STAT3*. The P-value and fold change identified in the two analysis for these genes is illustrated in table 5.9. These genes are briefly described in table 5.10.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>EFMR vs control female</th>
<th>Transmitting male vs control male</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>AKR1C3</em></td>
<td>Aldo-keto reductase family 1, member C1</td>
<td>0.0088</td>
<td>-5.47</td>
</tr>
<tr>
<td><em>EPHA2</em></td>
<td>EPH receptor A2</td>
<td>0.035</td>
<td>1.50</td>
</tr>
<tr>
<td><em>SH3BP5</em></td>
<td>SH3-domain binding protein 5 (BTK-associated)</td>
<td>0.030</td>
<td>1.94</td>
</tr>
<tr>
<td><em>STAT3</em></td>
<td>Signal transducer and activator of transcription 3</td>
<td>0.011</td>
<td>-1.69</td>
</tr>
</tbody>
</table>

**Table 5.9 Genes identified as significantly deregulated in the EFMR vs control female and the Transmitting male vs control male analysis**

Genes identified in common between the EFMR vs control female analysis and the transmitting male vs control male analysis. Fold changes and P-values were ascertained from the microarray analysis. Fold change for the EFMR vs control female analysis represent EFMR relative to control females. Fold change for the transmitting male vs control male analysis represent transmitting male relative to control males.
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene name</th>
<th>P-value</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKR1C1</td>
<td>aldo-keto reductase family 1, member C1</td>
<td>0.012</td>
<td>-9.19</td>
</tr>
<tr>
<td>AKR1C2</td>
<td>aldo-keto reductase family 1, member C2</td>
<td>0.010</td>
<td>-6.19</td>
</tr>
<tr>
<td>AKR1C3</td>
<td>aldo-keto reductase family 1, member C3</td>
<td>0.0088</td>
<td>-5.47</td>
</tr>
</tbody>
</table>

Table 5.11 Additional aldo-keto reductase family members identified in the EFMR vs control female analysis (P<0.05 list)

P-values and fold changes were ascertained from the microarray analysis. The fold changes represent EFMR relative to control females.
In light of the identification of *STAT3* as a gene of particular interest we interrogated the EFMR vs control female significant gene list for genes involved in Jak-STAT signalling. This identified six genes deregulated in the EFMR females (*P*<0.05); *IL10RB*, *IL7R*, *SOCS3*, *STAT5A*, *IL15RA* and *CCND1*. The P-value and fold change identified for theses genes is illustrated in table 5.12. The Jak-STAT signalling pathway and the genes deregulated are illustrated in figure 5.14.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>gene_assignment</th>
<th>P-value (Status)</th>
<th>Fold-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT3</td>
<td>signal transducer and activator of transcription 3 (acute-phase response factor)</td>
<td>0.011</td>
<td>-1.69</td>
</tr>
<tr>
<td>IL10RB</td>
<td>interleukin 10 receptor, beta</td>
<td>0.0090</td>
<td>-1.75</td>
</tr>
<tr>
<td>IL7R</td>
<td>interleukin 7 receptor</td>
<td>0.047</td>
<td>2.78</td>
</tr>
<tr>
<td>SOCS3</td>
<td>suppressor of cytokine signalling 3</td>
<td>0.041</td>
<td>-2.43</td>
</tr>
<tr>
<td>STAT5A</td>
<td>signal transducer and activator of transcription 5a</td>
<td>0.049</td>
<td>-1.63</td>
</tr>
<tr>
<td>IL15RA</td>
<td>interleukin 15 receptor, alpha</td>
<td>0.045</td>
<td>-1.54</td>
</tr>
<tr>
<td>CCND1</td>
<td>cyclin d1</td>
<td>0.028</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Table 5.12 Additional genes involved in Jak-STAT signalling identified in the EFMR vs control female analysis (*P*<0.05 list)

Additional genes identified in the EFMR vs control female *P*<0.05 gene list that are involved in the KEGG pathway for Jak-STAT signalling. P-values and fold changes were ascertained from the microarray analysis. The fold changes represent EFMR relative to control females.
Figure 5.14 KEGG JAK-STAT signalling pathway

JAK-STAT pathway as illustrated by the KEGG pathway diagram. Genes deregulated in the EFMR vs control female analysis P<0.05 gene list are illustrated by stars. Genes with increased expression in the EFMR females are illustrated in purple. The genes with decreased expression in EFMR females are illustrated in teal.

5.3.3 Analysis of EFMR females vs all controls (males and females)

In an attempt to increase the statistical power of the microarray analysis the EFMR females (n=6) were analysed against all of the controls (n=6, 3 female and 3 male). PCA analysis showed that the controls tended to cluster together but there was definite overlap between the controls and the EFMR samples (figure 5.15).
1way ANOVA analysis identified 138 genes as differentially expressed with a P-value <0.01 (Appendix table C3). Ninety five were down regulated and 43 genes were up-regulated in the in EFMR females. The largest degree of down regulation was -2.87 fold and the largest up-regulation of 5.36 fold. Removing genes with an expression fold change difference of less than 1.5 left a list of only 28 differentially expressed genes. Increasing the P-value to <0.05 identified 970 differentially expressed genes. Removing genes with an expression fold change difference of less than 1.5 left 137 genes (Appendix table C3). Eighty one genes are up regulated and 56 genes are down regulated genes.

Inclusion of the extra three male controls substantially decreased the number of significantly altered genes with a fold change of greater than +/-1.5 identified in EFMR females. Only 6 of the 28 significant genes identified in the EFMR vs all controls analysis were also identified in the original EFMR vs control female analysis. *PDGFC* was the only gene identified as potentially interesting from the DAVID analysis of the EFMR vs female control gene list to still be significantly differentially expressed. Investigation into why the remaining interesting
genes previously identified failed to validate with this analysis revealed differential expression of a number of these genes between the male and female controls. Dot plots of the expression level of a number of these genes (figure 5.16) showed that the expression level in the male controls was more in-line with the EFMR females than the control females.
Figure 5.16 Sex comparison Dot plots from the EFMR vs all control analysis

Dot plots of the expression of 8 representative genes demonstrating the differential levels of expression between the female and male controls demonstrate how the different sex control gene expression levels correlate to expression levels in EFMR females. The purple box and whisker plots represent the unaffected control individuals with males illustrated by a blue circle and females by a red circle. The green box and whisker plots represent the EFMR females who are all represented by red circles. Graphs were generated using Partek V6.3. The box plots show 90% confidence intervals with the whiskers demonstrating the upper and lower 10%.
Subsequent source of variation analysis of the array data revealed that status contributed only slightly more to the overall variation than sex (figure 5.17). Status had a source of variation score of 1.40 and sex had a score of 1.34 relative to error of 1.0.

**Figure 5.17 Source of variation present in the EFMR vs all control analysis**

These two graphs represent the results obtained from source of variation analysis of 1way ANOVA analyses performed on the EFMR vs all control data. Red bars illustrate the source of variation present in the factor being interrogated in the ANOVA analysis. Blue bars illustrate the source of variation attributed to error. The left hand graph represents the results obtained from the full ANOVA analysis using EFMR status as the factor under interrogation. The right hand graph represents the results obtained from the full ANOVA analysis using sex as the factor under interrogation. Graphs were generated using Partek V6.3.
5.3.4 Analysis of control males vs control females

To identify genes that are normally differentially expressed between the sexes we also performed 1-way ANOVA analysis between control males (n=3) and control females (n=3). PCA analysis showed no clustering of the sexes (figure 5.18).

![PCA analysis of control males vs control females](image)

**Figure 5.18 PCA analysis of the control male vs control female analysis**

This graph illustrates the results obtained from PCA analysis of the control male vs control female analysis. Red circles illustrate control females and blue circles represent control males. Graph generated using Partek V6.3.

This analysis identified 167 differentially expressed genes with a P-value <0.01. Of these 67 were down regulated and 100 genes were up-regulated in females when compared to males. The largest degree of down regulation was -74.83 fold (in a Y-chromosome gene) and the largest up regulation was of 11.8 fold. Removing genes with an expression fold change difference of less than 1.5 left a list of 95 differentially expressed genes in females compared with males (Appendix table C4). Of those genes 33 were up regulated and 62 were down regulated. Increasing the P-value to <0.05 identified 741 differentially expressed genes. Removing genes with an expression fold change difference of less than 1.5 left 344
differentially expressed genes (Appendix table C4). With 223 up regulated and 121 down regulated genes.

Comparison between the P<0.01 list identified between the sex controls (n=95 genes) and the P<0.01 gene list identified in the EFMR vs control female analysis (n=110 genes) identified 24 genes present in both lists (table 5.13). All 24 genes had the same direction and similar degree of fold change in the EFMR vs control female and the control male vs control female analyses. This suggests that EFMR females have similar expression level to males for these genes. Extending this comparison further and looking at the P<0.05 gene lists between the sex controls (n=344 genes) and the EFMR vs control female analysis (n=383 genes) identified 114 genes that were present to both lists (Appendix table C5). Out of these genes 113 had the same direction and similar degree of fold change in the EFMR vs control females and the control males vs control female analyses. The only gene to show alternate direction of fold change was RACGAP1.
### Table 5.13 Genes identified as deregulated in both the EFMR vs control female and control male vs control female analysis (P<0.01)

P-values and fold changes were ascertained from the microarray analysis. The fold changes represent EFMR females relative to control females and transmitting males relative to control females.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Name</th>
<th>EFMR vs control females</th>
<th>Male controls vs Female controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKRIC3</td>
<td>Aldo-keto reductase family 1, member C3</td>
<td>0.0088 -5.47</td>
<td>0.0023 -4.53</td>
</tr>
<tr>
<td>CDON</td>
<td>Cdon homolog</td>
<td>0.0035 -5.46</td>
<td>0.0058 -5.46</td>
</tr>
<tr>
<td>C1R</td>
<td>Complement component 1, r subcomponent</td>
<td>0.0064 -3.21</td>
<td>0.0017 -3.02</td>
</tr>
<tr>
<td>CFB</td>
<td>Complement factor B</td>
<td>0.0001 -2.08</td>
<td>0.0077 -2.16</td>
</tr>
<tr>
<td>CXCR7</td>
<td>Chemokine receptor 7</td>
<td>0.0020 -2.49</td>
<td>0.0034 -2.56</td>
</tr>
<tr>
<td>ENPP1</td>
<td>Ectonucleotide pyrophosphatase/phosphodiesterase 1</td>
<td>0.0008 9.90</td>
<td>0.0026 6.59</td>
</tr>
<tr>
<td>FANCB</td>
<td>Fanconi anemia, complementation group B</td>
<td>0.0001 1.36</td>
<td>0.0014 1.32</td>
</tr>
<tr>
<td>GALE</td>
<td>UDP-Galactose-4-Epimerase</td>
<td>0.0017 1.32</td>
<td>0.0010 1.36</td>
</tr>
<tr>
<td>HDAC9</td>
<td>Histone deacetylase 9</td>
<td>0.0012 1.79</td>
<td>0.0024 1.55</td>
</tr>
<tr>
<td>ITGA3</td>
<td>Integrin, alpha 3</td>
<td>0.0042 4.15</td>
<td>0.0048 3.67</td>
</tr>
<tr>
<td>ITGA6</td>
<td>Integrin, alpha 6</td>
<td>0.0030 6.75</td>
<td>0.0097 6.30</td>
</tr>
<tr>
<td>KLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>0.0048 -3.92</td>
<td>0.0002 -4.47</td>
</tr>
<tr>
<td>KREMEN1</td>
<td>Kringle containing transmembrane protein 1</td>
<td>0.0007 -1.64</td>
<td>0.0088 -1.56</td>
</tr>
<tr>
<td>LAMB1</td>
<td>LamininB1</td>
<td>0.0009 2.19</td>
<td>0.0027 1.80</td>
</tr>
<tr>
<td>LMCD1</td>
<td>LIM and cysteine-rich domains 1</td>
<td>0.0013 -2.21</td>
<td>0.0071 -2.35</td>
</tr>
<tr>
<td>LOH3CR2</td>
<td>Loss of heterozygosity, 3, chromosomal region 2, gene A</td>
<td>0.0030 -2.75</td>
<td>0.0093 -3.35</td>
</tr>
<tr>
<td>MET</td>
<td>Met proto-oncogene</td>
<td>0.0003 3.90</td>
<td>0.0076 3.15</td>
</tr>
<tr>
<td>MMP16</td>
<td>Matrix metallopeptidase 16 (membrane-inserted)</td>
<td>0.0078 1.96</td>
<td>0.0070 2.02</td>
</tr>
<tr>
<td>NTF3</td>
<td>Neurotrophin 3</td>
<td>0.0014 1.92</td>
<td>0.0088 2.17</td>
</tr>
<tr>
<td>RRM2</td>
<td>Ribonucleotide reductase, M2 subunit</td>
<td>0.0097 3.71</td>
<td>0.0017 3.24</td>
</tr>
<tr>
<td>TMTC1</td>
<td>Transmembrane and tetratricopeptide repeat containing 1</td>
<td>0.0007 -11.62</td>
<td>0.0010 -11.80</td>
</tr>
<tr>
<td>TNFAIP8</td>
<td>Tumor necrosis factor, alpha-induced protein 8</td>
<td>0.0003 -1.97</td>
<td>0.0006 -2.02</td>
</tr>
<tr>
<td>TP53I11</td>
<td>Tumor protein p53 inducible protein 11</td>
<td>0.0092 -2.20</td>
<td>0.0065 -2.10</td>
</tr>
</tbody>
</table>

Chapter 5-234
Ten genes from the list of interesting genes identified in the EFMR vs control female analysis, which demonstrated the sexual dimorphic expression trend were selected for RT-qPCR validation. The genes selected were EFNA5, GAS7, CDON, CXCR7, LAMB1, MET, NTF3, ITGA3, ITGA6 and EPHA2. Eight of these genes were present in the list of 24 genes identified by both analyses. EPHA2 was identified in the P<0.01 list of differentially expressed genes between control males vs control females and the P<0.05 list from the EFMR vs control females analysis. GAS7 was identified in the P<0.01 list of differentially expressed genes between EFMR vs control females and in the P<0.05 list from the control males vs control females analysis.

Seven out of the ten genes validated as differentially expressed between the male and female controls with a P-value from the real time analysis of <0.05. Figure 5.19 shows the real time results for each gene analysed and figure 5.20 shows the combined average results. The remaining three genes CXCR7, LAMB1 and ITGA3 all showed a trend towards differential expression between the sexes; however, they did not meet the P-value cut off (figure 5.20). Statistical analysis of the differences in expression levels between EFMR females and control males identified that none of the 10 genes showed a significant difference (P<0.05) in gene expression. The only gene to come close was EFNA5 with a P-value of 0.061. The remainder of the genes had P-values between 0.30 and 0.90, demonstrating that the EFMR females have similar levels of expression of these genes as the control males. Further analysis of the PCA results obtained from the EFMR vs all controls analysis (figure 5.15) showed a trend towards closer clustering between the control males and EFMR females than between the control females and EFMR females.
Figure 5.19 Real time validation of sexually dimorphic gene expression

Graphical representation of the RT-qPCR results obtained for each gene investigated. Navy bars are control male samples, pink bars are control females, purple bars EFMR females and teal bars the transmitting males. The y-axis represents the gene expression level standardised to Actin-β. Individual samples are plotted across the x-axis, the n numbers indicate the number of RNA samples used in obtaining the gene expression level.
Figure 5.20 Combined real time validation of the sex specific differences in gene expression

Graphical representation of the average RT-qPCR results obtained for each gene investigated. Navy bars represent the average expression level of the male controls (n=3) pink bars are control females (n=3), purple bars are EFMR females (n=6) and teal bars transmitting males (n=3). Individual genes are plotted across the x-axis. The error bars represent the standard error. No error bars are shown for transmitting males as the RNA from the three samples being analysed were combined together in order to fit the samples on a real time plate, as a result no standard error could be calculated. * represents a P-value of <0.05. ** = P<0.01, the P-values are for the gene expression differences between the control male and control female samples.
Given the evidence suggesting that the EFMR females have a similar expression to males in a number of genes we reassessed the functional annotation cluster analysis of the P<0.05 EFMR gene list searching for hormone related functions. An enriched annotation cluster of genes, which function as hormone receptors, was identified (table 5.14). This cluster involved four genes PPARG, NR2F2, NR4A1 and ZFP36L1. The P-value and fold change identified for these genes is illustrated in table 5.15. These genes are briefly described in table 5.16. The function, cell localisation evidence of brain expression and related genes were investigated for all of these genes in order to assess their likely relevance to the EFMR phenotype and PCDH19 function (table 5.16). The enrichment score for this cluster was quite low at only 0.57 (P-value 0.3). However the sex specific nature of EFMR in combination with the identification of a number of AKR1C genes, which are involved in catalysing a number of hormones including the sex steroid hormones estrogen and testosterone, in the EFMR female gene list means this group of genes is also of interest.
<table>
<thead>
<tr>
<th>Functions</th>
<th># of genes</th>
<th>P-value</th>
<th>Fold enrichment</th>
<th>Benjamini correction</th>
<th>Genes Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zinc finger, nuclear hormone receptor-type</td>
<td>3</td>
<td>2.20E-01</td>
<td>3.4</td>
<td>1.00E+00</td>
<td>PPARG, NR2F2, NR4A1, ZFP36L1</td>
</tr>
<tr>
<td>Steroid hormone receptor</td>
<td>3</td>
<td>2.20E-01</td>
<td>3.4</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>Nuclear hormone receptor, ligand-binding, core</td>
<td>3</td>
<td>2.30E-01</td>
<td>3.3</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>Nuclear hormone receptor, ligand-binding</td>
<td>3</td>
<td>2.30E-01</td>
<td>3.3</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>Zinc finger region:NR C4-type</td>
<td>3</td>
<td>2.30E-01</td>
<td>3.3</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>DNA-binding region:Nuclear receptor</td>
<td>3</td>
<td>2.60E-01</td>
<td>3</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>ZnF_C4</td>
<td>3</td>
<td>2.70E-01</td>
<td>3</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>HOLI</td>
<td>3</td>
<td>3.00E-01</td>
<td>2.7</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>Steroid hormone receptor activity</td>
<td>3</td>
<td>3.60E-01</td>
<td>2.4</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>Ligand-dependent nuclear receptor activity</td>
<td>4</td>
<td>6.00E-01</td>
<td>1.3</td>
<td>1.00E+00</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.14 Enriched annotation cluster of hormone response genes identified in the EFMR vs control female analysis (P<0.05)

Genes and functions identified in the enriched annotation cluster involving hormone function as identified in the functional cluster analysis of the P<0.05 list of genes deregulated in the EFMR vs control female analysis.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>P-value</th>
<th>Fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR4A1</td>
<td>nuclear receptor subfamily 4, group a, member 1</td>
<td>0.036</td>
<td>-1.91</td>
</tr>
<tr>
<td>NR2F2</td>
<td>nuclear receptor subfamily 2, group f, member 2</td>
<td>0.0027</td>
<td>1.93</td>
</tr>
<tr>
<td>PPARG</td>
<td>peroxisome proliferative activated receptor, gamma</td>
<td>0.011</td>
<td>-1.68</td>
</tr>
<tr>
<td>ZFP36L1</td>
<td>zinc finger protein 36, c3h type-like 1</td>
<td>0.035</td>
<td>-2.1</td>
</tr>
</tbody>
</table>

Table 5.15 Details of the genes identified in the hormone response enriched annotation cluster that are deregulated in the EFMR vs control female analysis (P<0.05)

P-values and fold changes were ascertained from the microarray analysis. The fold changes represent EFMR relative to control females.
NR4A1: Orphan nuclear receptor. This gene encodes a member of the steroid-thyroid hormone-retinoid receptor super family. The encoded protein acts as a nuclear transcription factor. Translocation of the protein from the nucleus to mitochondria induces apoptosis.

NR2F2: Regulation of the apolipoprotein A-I gene transcription. Binds to DNA site A. Belongs to the nuclear hormone receptor family.

PPARG: This gene encodes a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors. It belongs to the nuclear hormone receptor family. PPARs form heterodimers with retinoid X receptors (RXRs) and these heterodimers regulate transcription of various genes.

ZFP36L1: Probable regulatory protein involved in regulating the response to growth factors.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Function</th>
<th>Localisation</th>
<th>Brain expression</th>
<th>Related genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR4A1</td>
<td>Orphan nuclear receptor. This gene encodes a member of the steroid-thyroid hormone-retinoid receptor super family. The encoded protein acts as a nuclear transcription factor. Translocation of the protein from the nucleus to mitochondria induces apoptosis.</td>
<td>Nucleus</td>
<td>Yes</td>
<td>PPARG, NR2F2</td>
</tr>
<tr>
<td>NR2F2</td>
<td>Regulation of the apolipoprotein A-I gene transcription. Binds to DNA site A. Belongs to the nuclear hormone receptor family</td>
<td>Nucleus</td>
<td>Yes</td>
<td>PPARG, NR4A1, ZFP36L1</td>
</tr>
<tr>
<td>PPARG</td>
<td>This gene encodes a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors. It belongs to the nuclear hormone receptor family. PPARs form heterodimers with retinoid X receptors (RXRs) and these heterodimers regulate transcription of various genes.</td>
<td>Nucleus</td>
<td>Yes</td>
<td>NR4A1, NR2F2</td>
</tr>
<tr>
<td>ZFP36L1</td>
<td>Probable regulatory protein involved in regulating the response to growth factors.</td>
<td>Nucleus</td>
<td>Unknown</td>
<td>NR2F2, PPARG</td>
</tr>
</tbody>
</table>

Table 5.16 Information about the genes identified in the hormone response annotation cluster that are deregulated in the EFMR vs control female analysis (P<0.05)

Gene information was extracted from the databases associated with the DAVID program.

Investigations into how males are protected from EFMR revealed that there are only minimal gene expression differences present between transmitting males with a mutation in PCDH19 and control males. Inclusion of the male controls with the EFMR vs control analysis unexpectedly revealed EFMR females to have an expression pattern similar to control males. This lead to the identification of a number of genes deregulated in EFMR females, which are involved with sex steroid function.
5.4 Discussion

Microarray expression profiling was performed on skin fibroblast cell lines obtained from EFMR subjects and controls to explore pathways and biological functions disrupted as a result of mutations in *PCDH19*. This was performed with the aim of increasing knowledge about *PCDH19* function and the mechanism involved in EFMR. Of particular interest was to explore what it is about males that saves them from EFMR. To assess these aims we included EFMR females, transmitting males with *PCDH19* mutations, and sex and age matched controls in the analysis. As a result we performed a number of different analyses of the microarray data in order to identify biologically relevant information. We performed four-principle analyses; EFMR females vs control females, transmitting males vs control males, EFMR females vs all controls (male and female) and control males vs control females. The last comparison was performed as a result of information obtained from the EFMR females vs all controls analysis indicating a gender specific basis to the results obtained.

There are two main outcomes of the analyses carried out. Firstly *PCDH19* mutations result in deregulation of a number of genes involved in cell adhesion and axon guidance. Deregulation of cell adhesion genes suggests that *PCDH19* mutations may disrupt cell adhesion in the context of axonal guidance. Secondly, *PCDH19* mutations result in altered levels of gene expression of a number of genes involved in estrogen and potentially testosterone (mosaic deletion of *PCDH19* in a male) response. Overall the results of the analysis into the altered sex specific expression profile identified a number of genes deregulated where expression is modulated by estrogen. These findings lead us to hypothesis that mutations in *PCDH19* disrupt the ability of some cells to respond to estrogen stimulation potentially through disruption to functioning of the estrogen receptor α (*ERα*).
Overall caveats to the microarray expression profiling assay

Firstly there are a number of caveats in the assay performed that need to be taken into account. The expression profiling could only be carried out on low numbers of samples due to subjects availability and consent. Skin fibroblast samples were required for this analysis as lymphoblastoid cell lines did not have sufficient expression of PCDH19 nor were they an adherent cell line. Skin biopsy required to obtain a skin fibroblast sample is an invasive process which can result in a permanent scar. This affects ability to gain consent to obtain these cells. The cost of the microarray itself is also a limiting factor for the number of individuals that can be analysed. The microarray experiment proceeded once we had the 15 samples with at least 3 subjects in every test group. The analysis would determine if the sample was sufficient to give us an indication of the biological effects of mutation in PCDH19. The sample size was considered in the analysis process as we restricted the analysis to the most significant genes with low P-values (P<0.05) and with a robust fold change of greater than plus or minus 1.5 fold.

A further caveat was that we were not working with neural cells which if practical would have been more appropriate for the disease being assessed. Affected EFMR females have defects in brain function. Pluripotent stem cells obtained from a nasal biopsy and differentiated down the neuronal pathway would have been preferable; however, this technology was not available to us and even if it was, collecting samples from the remote regions involved would have presented insurmountable challenges. Such circumstances led us to as the best available cell types. PCDH19 expression was visible by PCR in fibroblasts and given that they are an adherent cell type PCDH19 may function similarly as in neuronal
tissue. Results obtained from this analysis will ultimately require further validation in a neuronal cell model.

5.4.1 Functional annotation and enriched cluster analysis of the EFMR vs control female analysis

Functional annotation and cluster analysis of the EFMR vs control female array results identified four key pathways and three functional annotation clusters as potentially very relevant to PCDH19 mutations and EFMR. The pathways identified had significant overlap with the functional annotation clusters. Overall these results indicate that PCDH19 mutations affected focal adhesion, extracellular matrix receptor interactions, adherens junctions and axon guidance. The high degree of overlap in the genes identified across these functions and pathways, suggests that these pathways and annotation clusters converge at the point of PCDH19. Implying that PCDH19 functions in the extracellular matrix to assist cell-cell adhesion required for axonal guidance.

A disruption to adherens junctions and focal adhesions correlates well with the known function of PCDH19 in calcium dependent cell-cell adhesion. The adherens junction pathway is predominantly centered on cadherin mediated cell-cell adhesions. Six genes from this pathway were deregulated in the EFMR subjects. Four of these genes (SNAIL1, MET, TGFBR1 and IGF1R) occurred in the region of the pathway that is associated with weak cadherin cell-cell adhesion. Protocadherins are involved in weaker cell-cell adhesions than cadherins (Hirano et al., 1999, Obata et al., 1995, Redies et al., 2005, Sago et al., 1995, Sano et al., 1993, Yamagata et al., 1999, Yoshida, 2003), which correlates well with a defect in a protocadherin. Out of the genes involved in weak cell adhesion, three genes (MET, TGFBR1
and \textit{IGF1R}) are single pass transmembrane proteins and therefore potential interactors of \textit{PCDH19}.

Disruption to the extracellular matrix and extracellular matrix receptor interactions correlates well with the positioning of the \textit{PCDH19} mutations, which occur exclusively in the extracellular domain of the protein (Depienne et al., 2009a, Dibbens et al., 2008, Hynes et al., 2009). The extracellular domain of \textit{PCDH19} is considerably more highly conserved than the cytoplasmic domain (Sano et al., 1993), indicating that this region is of high functional significance. Across the four pathways identified with significance to \textit{PCDH19} mutations, the vast majority of genes deregulated in EFMR females are located either in the extracellular matrix or within the cell membrane. This further reinforces the importance of this region in \textit{PCDH19} function. The extracellular matrix genes identified are also potential interactors of \textit{PCDH19} and protocadherins in general. These genes may also have relevance to cadherin mediated cell-cell adhesion. The extracellular domains of protocadherins and cadherins are highly related, whilst the cytoplasmic domains differ (Alagramam et al., 2001, Angst et al., 2001, Frank and Kemler, 2002, Kohmura et al., 1998, Nollet et al., 2000, Obata et al., 1995, Sano et al., 1993, Suzuki, 1996, Wolverton and Lalande, 2001, Wu and Maniatis, 1999). Conservation of the extracellular domain suggests that the extracellular domains could perform similar functions in cadherins and protocadherins.

Our findings therefore suggest that a major function of \textit{PCDH19} dependent cell-cell adhesion is the regulation of genes involved at the site of cell-cell contacts in axonal guidance. This finding disagrees with the previous suggestion that because protocadherins are involved in weaker cell-cell interactions than cadherins (Alagramam et al., 2001, Angst et al., 2001, Frank and Kemler, 2002, Kohmura et al., 1998, Nollet et al., 2000, Obata et al., 1995, Sano et al., 1993, Suzuki, 1996, Wolverton and Lalande, 2001, Wu and Maniatis, 1999) their main
function may lie in signalling (Redies et al., 2005, Suzuki, 2000). That would entail homophilic binding of protocadherins acting as the initial step for instigating intracellular signaling from cell adhesions.

Whilst advances have been made into genes involved in intracellular signalling from protocadherins, no overall signal pathway has emerged as relevant to multiple protocadherins. The intracellular signalling pathways are still being elucidated on a protocadherin by protocadherin basis. To date, Pcdh7 has been identified to interact with protein phosphatase-1α (PP1α) (Vanhalst et al., 2005, Yoshida et al., 1999) and TAF1/Set (Heggem and Bradley, 2003). PAPC (homolog of PCDH8) interacts with Frizzled 7 receptor (FZD7) activating Rho and protein kinase Ca (Medina et al., 2004, Unterseher et al., 2004). PCDH8 also acts through N-cadherin via TAO2beta and p38 MAP kinases (Yasuda et al., 2007). PCDH24-induced contact inhibition occurs through downregulation of beta-catenin signalling (involving; CD44, PLAUR, Myc, cyclin D1 and Met) (Ose et al., 2009). The protocadherin Fat1 acts through the hippo tumor-suppressor pathway to regulate tissue size (Willecke et al., 2006) (involving; EX, Mer, Mst, WW45, Mob, Lats, YAP, TAZ and TEAD (reviewed in Zhao et al., 2008)).

Out of the genes previously associated with intracellular protocadherin signalling, Met and cyclin D1 were also identified deregulated in the EFMR females. This suggested that PCDH19 could signal through similar pathways as PCDH24. A paralog of FZD7, FZD6 was also up regulated in the EFMR females. Three Rho GTPase associated genes were also deregulated in the EFMR females: ARGAP20 (a Rho GTPase activating protein), CDC42EP2 and CDC42EP3 (Rho GTPase binding proteins). However there were no significant expression changes in protein kinase Ca. FZD6 could therefore interact with PCDH19 activating Rho GTPase dependent signalling, suggesting that PCDH19 may also signal
through a similar pathway to \textit{PCDH8}. The overlap in genes identified with altered expression, as a result of mutations in \textit{PCDH19}, with genes previously associated with protocadherin signalling suggests that there is some overlap in intracellular signalling between certain protocadherins. However there are still significant differences present between each of the protocadherins.

A lack of an overall signal pathway involved in intracellular signalling from different protocadherins supports the proposed role for protocadherins in building a highly organised neuronal network (Hilschmann et al., 2001, Wang et al., 2002a). The combination of the large number of protocadherins expressed in the brain and their apparent differences in cytoplasmic targets demonstrates that they have the adhesive diversity and molecular code for specifying highly organized and coordinated neuronal connections in the brain, as previously suggested (Hagler and Goda, 1998, Kohmura et al., 1998, Uemura, 1998, Wu and Maniatis, 1999). Our results support this theory as we did find some similarities to previously identified signaling pathways; but, there are still differences present.

Potentially the most interesting pathway and functional cluster deregulated in the EFMR females are those associated with axon guidance, neuronal development and neuronal migration. Defects in these functions are particularly relevant to the EFMR phenotype. Disruption of normal cortical architecture has been identified in a frontal lobe surgical specimen obtained from an affected EFMR female. This specimen had dysplastic neurons in the cortex and underlying white matter (Ryan et al., 1997). The individual neurons also had a more varied morphology and lacked uniform orientation. Abnormal clustering of pyramidal cells was also noted (Ryan et al., 1997). These results indicate a possible defect in the processes involved in ensuring the correct positioning of neurons. The microarray results
support this suggestion with altered expression of a number of genes involved in axon guidance.

Altogether, our microarray analysis along with the surgical specimen analysis (Ryan et al., 1997), implicates a role for PCDH19 in ensuring neurons are targeted to their correct destination. We have also identified a number of additional genes whose expression level is altered in response to mutation in PCDH19 thereby implicating a role for these genes in cell-cell adhesion and axon guidance. Additional work will be required to replicate these interactions in neuronal cells.

These results advance understanding of the precise function of PCDH19 and how mutations result in epilepsy and ID. However the biggest and most challenging question remains, why are males with mutations in PCDH19 unaffected by epilepsy and ID? To attempt to gain further insights to this question we analysed the expression profile of transmitting males with mutations in PCDH19.

5.4.2 Analysis of the transmitting males vs control males.

The low number of genes identified as markedly differentially expressed (P<0.05 and fold change greater that +/- 1.5 fold) between the transmitting males and control males suggests that the transmitting males are unaffected by mutations in PCDH19 on a molecular level. Males with mutations in PCDH19 are classified phenotypically as unaffected. Specifically these males are unaffected by epilepsy and ID, the most prominent clinical features of EFMR (Ryan et al., 1997, Scheffer et al., 2008). There is some suggestion of a tendency towards obsessive-compulsive traits in these males (Scheffer et al., 2008); however, this was mainly diagnosed retrospectively and anecdotally in light of the identification of a mutation in
PCDH19 in a related female. Their general lack of overt clinical symptoms was matched by absence of subclinical features through the functional annotation analysis, which failed to detect any relevant biological functions or pathways enriched in the significantly deregulated genes (where a significantly deregulated gene is defined as a gene where the expression level was significantly different, with a P-value less than 0.05, between control males and transmitting males and a fold change in gene expression above 1.5 fold).

The most interesting results to come out of the transmitting male vs control male analysis arose from the comparison of genes identified as differentially expressed between the EFMR vs control female and the transmitting male vs control male analyses. Four genes were identified in common between the two P<0.05 gene lists: AKR1C3, EPHA2, SH3BP5, and STAT3.

SH3BP5 does not appear, based on current literature, to be functionally relevant to EFMR. SH3BP5 is a SH3 domain-binding protein that preferentially associates with BTK, a key regulator of B-cell development (Matsushita et al., 1998). The remaining three genes do appear to be functionally relevant based on their involvement with estrogen and the cadherin super family, as well as the interactions present between these genes.

AKR1C3, aldo-keto reductase family 1 - member C3 is also known as 17Beta-hydroxysteroid dehydrogenases 5. Members of the aldo-keto reductase family, AKR1C1-C4, are involved in the inactivation and formation of male and female sex hormones (Penning et al., 2000). They function by catalysing androgen, estrogen, and progesterone metabolism (Penning et al., 2000). AKR1C1-C3 are expressed in the brain, with AKR1C1 and C2 showing the highest level of expression. AKR1C4 is predominantly a liver specific isoform (Penning et al., 2000). Decreased levels of AKR1C3 inhibit production of testosterone (Nakamura et al., 2009).
Estrogen treatment of human cell lines increases expression of \textit{AKR1C1}, \textit{AKR1C2}, and \textit{AKR1C3} but not \textit{AKR1C4} (Kang and Kim, 2008). Decreased expression of \textit{AKR1C1}, \textit{AKR1C2}, and \textit{AKR1C3} was identified in the EFMR females.

\textit{EPHA2}, ephrin receptor A2, a member of the Eph family of receptor tyrosine kinases is activated by membrane bound ephrin ligands. Eph/ephrin signalling is involved in axon guidance, cell morphogenesis and tissue patterning through regulation of cell-cell adhesion (Pasquale, 2005). \textit{EPHA2} can bind a number of ephrinA ligands but preferentially binds to ephrinA1 and A5 (Gale and Yancopoulos, 1997). \textit{EFNA5} was identified as a gene of interest in the EFMR vs control female analysis. \textit{EphA2} has also previously been associated with a number of cadherins. E-cadherin regulates \textit{EphA2} expression and cellular localisation to cell-cell contacts, where \textit{EphA2} then regulates cell-cell adhesion (Miao and Wang, 2009). \textit{EFNA5-EphA2} instigated signalling is required for recruitment of N-cadherin to cell-cell contacts and allowing interaction with β-catenin (Cooper et al., 2008). \textit{EphA2} has also been identified as a downstream target of estrogen (Zelinski et al., 2002) and may regulate progression of breast cancer. \textit{EphA2} and estrogen levels are often inversely related in breast cancer cell lines (Zelinski et al., 2001, Zelinski et al., 2002). \textit{EphA2} had increased expression in EFMR females.

\textit{STAT3} a signal transducer and activator of transcription, is an acute-phase response factor that is rapidly activated in response to the multiple cytokines and growth factors. \textit{STAT3} functions in a number of systems and disease states including but not limited to lymphohocyte differentiation, prostate cancer, T-cell development and inflammatory bowel disease. Of particular interest is the role of \textit{STAT3} in neurons. Activation of \textit{STAT3} results in upregulation of genes which promote neuroprotection, neuroregeneration and neurodevelopment (reviewed in Dziennis and Alkayed, 2008). STAT activation occurs in the central nervous system in
response to injury, growth factors (including EGF, PDGF, NGF and BDNF), cytokines and hormones (reviewed in Dziennis and Alkayed, 2008).

STAT3 is also involved in estradiol-mediated neuroprotection (Dziennis et al., 2007). Estradiol is one of the three major naturally occurring estrogens in women. Estradiol reduces brain injury by reducing lesion size and neuronal death in an experimental model of cerebral ischemia. The neuroprotective effect of estradiol has been linked to increased STAT3 activation in neurons (Dziennis et al., 2007).

Associations between STAT3 and cadherins have also been identified. Cadherin-cadherin engagement activates STAT3 even in the absence of cell-cell contact (Arulanandam et al., 2009). This activation of STAT3 occurs through the activity of Rac1 and Cdc42, which then functions in a positive feedback loop to further increase Rac1, Cdc42 and Stat3 activity which functions as a potent cell survival signal (Arulanandam et al., 2009). Stat3 has also been identified in ephrin signalling. EphrinB1 interacts with Stat3 in a tyrosine phosphorylation-dependent manner, resulting in phosphorylation and enhanced transcriptional activation of Stat3 (Bong et al., 2007). Recruitment and activation of Stat3 reveals a signal pathway from EphrinB1 to the nucleus (Bong et al., 2007). STAT3 expression was decreased in EFMR females.

STAT5A also had decreased expression in the EFMR females but was unaffected in the transmitting males. STAT5A has been identified as a regulator of sex-specific gene expression in female, but not male, mouse liver. STAT5A functions as a feminising factor through positively regulating expression of female genes and negatively regulating male genes (Clodfelter et al., 2007). Estrogen and progesterone have been identified in the mouse
mammary glands to be critical regulators of expression of STAT5A. STAT5A expression is greatly increased by the presence of estrogen and progesterone (Santos et al., 2008).

Links to the female sex steroid estrogen have been identified in all of these genes. Indicating an alteration in estrogen production and or function could possibly be involved in EFMR. Inverse fold changes in expression level were identified in AKR1C3, EphA2 and STAT3 between the EFMR vs control female and transmitting vs control male array analyses. This fits well with the expression pattern of EFMR, with PCDH19 mutations in females resulting in EFMR whilst transmitting males with mutations in PCDH19 are able to compensate for these defects. The inverse fold change in expression levels implicates a role for these genes in protecting males from EFMR. The correlation between the interactions of estrogen and each of these genes with the direction of fold change identified in EFMR females, suggests that they may have decreased ability to response correctly to estrogen stimulation. Estrogen increases expression of AKR1C; EFMR females have decreased expression of AKR1C. Estrogen and EphA2 have inverse expression levels and EFMR females have elevated EphA2 expression. Estradiol increases expression of STAT3 and STAT5A and EFMR females have decreased expression of STAT3 and STAT5A. The correlation between mutations in PCDH19 and altered expression of a number of genes involved in estrogen response pathways is an intriguing finding which warrants further investigation. Validation in the laboratory through in vivo experiments assessing the molecular response to estrogen of cell line with or without mutations in PCDH19 will be important in confirming this correlation. The laboratory-based validation unfortunately was not achievable in the PhD timeframe.

5.4.3 Mutations in PCDH19 results in a male like expression pattern

The altered expression level of a number of genes involved in estrogen response pathways in EFMR females correlates with the unexpected finding that EFMR females appear on a gene
expression level to be male. Out of the 344 genes identified as differentially expressed between the sex controls 114 (42%) of these genes were also differentially expressed in the EFMR females compared to control females. Out of the list of 114 genes 113 (99%) had the same direction of fold change in the EFMR females and control males when compared to the control females. This indicates that the expression level for these 113 genes in EFMR females is similar to that of control males. This was confirmed by the RT-qPCR, which identified significant (p<0.05) differences in expression between control males and control females for 7 of the genes tested. However no significant difference in expression level was present between the control males and the EFMR females.

In light of the identification of altered expression levels of a number of genes involved with signalling in response to estrogen and the apparent male expression pattern in EFMR females we went back and looked for enrichment of hormone related genes in the EFMR vs control female analysis. This identified four genes of interest; PPARG, NR2F2, NR4A1 and ZFP36L1.

PPARG, peroxisome proliferative activated receptor – gamma, is a member of the nuclear receptor family. PPAR-gamma is involved in number of biological functions and disease states including but not limited to regulation of adipose tissue, lipid oxidation, type II diabetes, cancer, immunology and polycystic ovary syndrome. Studies in to the function of PPARs in the ovary have revealed a role for PPARs in influencing estradiol production (reviewed in Komar, 2005). PPARs can bind to estrogen response elements (EREs) where they act as competitive inhibitors. PPARG can also stimulate ubiquitination of estrogen receptor α, instigating its degradation (reviewed in Komar, 2005). Conversely estrogen receptorβ inhibits PPARG activity (Foryst-Ludwig et al., 2008). Activation of PPARG can
also influence progesterone production in ovarian cells (Komar, 2005). \textit{PPARG} expression was decreased in the EFMR females.

\textit{ZFP36L1}, zinc finger protein 36, c3h type-like 1, is a member of the \textit{TIS11} family of early response genes. Not much is known about this gene. It is believed to be a putative nuclear transcription factor that likely functions in regulating response to growth factors. \textit{ZFP36L1} has been implicated in breast cancer. \textit{ZFP36L1} expression was found to be up regulated by medroxyprogesterone acetate (Compagni et al., Pennanen et al., 2009) – a synthetic progestagen which elicits similar responses to progesterone. The increase in expression identified in breast cancer cell lines was dependent on the presence of estrogen. However estrogen alone did not affect expression levels of \textit{ZFP36L1} in these cell models (Pennanen et al., 2009). \textit{ZFP36L} expression was decreased in the EFMR females.

\textit{NR4A1}, nuclear receptor subfamily 4, group a, member 1, belongs to the steroid-thyroid hormone-retinoid receptor super family. \textit{NR4A1} has been implicated in T-cell apoptosis, brain development, and the hypothalamic-pituitary-adrenal axis and vascular disease. \textit{Nur77}, the mouse equivalent of \textit{NR4A1}, is widely expressed in high levels in the brain. Expression of \textit{Nur77} in the ovary is rapidly reduced by a treatment with a strong estrogenic endocrine disrupter, DES (Inaoka et al., 2008). Luteinizing hormone (Jobling et al.) induces \textit{Nur77} gene expression. Over expression of dominant negative \textit{Nur77} reduced LH-mediated progesterone biosynthesis in mouse Leydig cells (Song et al., 2001). \textit{Nor-1} the mouse equivalent of \textit{NR4A3}, an additional member of the nuclear receptor subfamily 4, is highly homologous to \textit{NR4A2} (Nuclear Receptors Nomenclature Committee, 1999). Inhibition of \textit{Nor-1} expression in mice results in defective axonal growth and region-specific cell death in the hippocampus leading to alterations to hippocampal excitability and increased susceptibility to chemically
induced seizures (Ponnio and Conneely, 2004). Expression of NR4A1 was decreased in the EFMR females.

NR2F2, nuclear receptor subfamily 2, group f, member 2, modulates hormone-induced expression of a number of genes. NR2F2 has been associated with a number of functions in the brain including regulating growth and patterning of the postnatal mouse cerebellum, regulating cell migration in the mammalian basal forebrain and involvement in the caudal migratory stream. NR2F2 binds to the nucleotide sequence located between the TATA box and transcriptional initiation site of the angiotensinogen gene and down-regulates estrogen induced promoter activity (Narayanan et al., 1999). NR2F2 influences the expression of a number of genes involved in sex steroid function. NR2F2 expression decreases estrogen synthesis, progesterone metabolism and testosterone metabolism. Conversely NR2F2 expression increases expression of estrogen receptor (Pereira et al., 2000). NR2F2 can also bind PPAR binding sites and down regulate the expression of various genes that are induced by this family of receptors (Narayanan et al., 1999). The increased expression of NR2F2 in EFMR females fits well with its ability to down-regulates estrogen induced promoter activity.

The function of these four genes adds further evidence to support the notion that EFMR females have altered levels of gene expression of a number of genes involved in estrogen response signalling. This is not the first instance of a member of the cadherin super family to have been linked to estrogen or sex-steroids.

Loss of E-cadherin mediated cell-cell contacts has been determined to reduce the transcriptional efficiency of estrogen receptor alpha (ERα) (Huet et al., 2008). Loss of E-cadherin has been observed in a number of cancerous states where it is associated with tumor differentiation, metastasis and worse patient outcome (Jiang and Mansel, 2000). The loss of
E-cadherin results from estrogen regulation of the expression of transcription factors Snail and Slug. Estrogen stimulation suppresses expression and promoter activity of E-cadherin and increasing expression of Snail (expression of snail homolog 1 was decreased in the EFMR females) and Slug in ovarian cancer cells (Park et al., 2008). Estrogen has been identified in this model as the trigger for metastatic behavior, which occurs exclusively through an ERα-dependant pathway (Park et al., 2008). Activation of ERα induces E-cadherin extracellular shedding and internalization followed by lysosomal degradation (Helguero et al., 2008) and ultimately aids migration of these cells.

N-cadherin has also been associated with sex steroids. N-cadherin is regulated by gonadal steroids in the adult hippocampus (Monks et al., 2001b). Estradiol but not testosterone treatment increased expression and immunoreactivity of N-cadherin in CA1 and CA3 pyramidal cells in the adult hippocampus. Both testosterone and estradiol increased N-cadherin immunoreactivity in the neuropil of the stratum lacunosum-moleculare (Monks et al., 2001b). This suggested that various gonadal steroids are involved in regulation of N-cadherin in a brain region specific manner and that steroid regulation of N-cadherin in the adult hippocampus may be involved in structural re-modeling of hippocampal neurons (Monks et al., 2001b). Analysis in the spinal cord identified that whilst N, E and R-cadherins were all expressed gonadal steroids only regulated N-cadherin expression in these cells (Monks et al., 2001a).

Our microarray results indicate that loss of PCDH19 dependent cell-cell adhesion also results in altered levels of gene expression of a number of genes involved in estrogen and potentially testosterone (mosaic deletion of PCDH19 in a male) response. These results have lead us to formulate the hypothesis that EFMR females have a defect in the ability to respond correctly to estrogen stimulation under certain circumstances. This could occur through reduced
transcriptional activity of ERα. The reduced response to estrogen results in altered expression of a number of estrogen responsive transcription factors (including STAT3, STAT5 and NR2F2). As a result EFMR females appear on a molecular expression level to be male. This is currently only a hypothesis and requires laboratory based experimental confirmation.

5.4.4 Estrogen and epilepsy

The hypothesis that EFMR females have a defect in their ability to respond correctly to estrogen stimulation associates well with the seizure progression of EFMR. Females with a mutation in PCDH19 develop normally through infancy and begin to have seizures at a mean age of 14 months (range 6-36 months) (Scheffer et al., 2008). Seizures then cease at a mean age of 12 years (Scheffer et al., 2008). These ages correlate with the levels of circulating female sex steroids, with seizure onset occurring once the contribution from the mother’s sex-steroids has diminished. Offset of seizures occurs around the age of puberty when there is a surge in estrogen levels (reviewed in Ober et al., 2008).

A role for sex-steroids has previously been suggested for epilepsy. Epidemiological studies of epilepsy identified sex differences in the incidence of epilepsy. The menstrual cycle and menopause influences on epilepsy largely eclipse gender differences in incidence suggesting that sex-steroids play a significant role in epilepsy (Christensen et al., 2005). Seizure incidence increases during low estrogen, low progesterone and the follicular phase when plasma estradiol sharply increases. A decrease in seizure incidence is seen when progesterone levels are high relative to estrogen (reviewed in Vagnerova et al., 2008).
5.4.5 Estrogen defects associated with disease

Functions of estrogen and the ERs are highly associated with a number of human pathologies including breast cancer. There is a parallel between the sexually dimorphic genes disrupted in the EFMR females and breast cancer. Out of the 24 genes identified as significantly deregulated in the EFMR vs control female and the male control vs female control analysis (P<0.01) 13 genes have been associated with breast cancer. It is not surprising that genes deregulated as a result of defects in cell-cell adhesion are involved in cancer. The ability of a cancerous cell to grow and metastasize is highly dependant on manipulation of and escape from cell-cell adhesion.

Clinically breast cancers presents as either ERα+ (60-70% of all breast tumors) or ERα- tumors. Presence or absence of ERα is an important predictor of a patient’s prognosis. The presence of ERα correlates with an overall better prognosis than cancers lacking ERα, which typically have a more aggressive phenotype and a poor prognosis (reviewed in Brinkman and El-Ashry, 2009). ERα+ breast cancers respond to endocrine therapies (for example, tamoxifen), while ERα- cancers are resistant to these therapies. However 25-35% of ERα+ cancers are also resistant to endocrine therapies (Brinkman and El-Ashry, 2009). The high correlation between genes disrupted in EFMR females and genes associated with breast cancer suggests a potential role for PCDH19 in breast cancer. Mutations in PCDH19 appear to disrupt the ability of certain cell’s to respond to estrogen stimulation through defective functioning of ERα. Mutations in PCDH19 could be associated with the 25-35% ERα+ cancers that resist endocrine therapies. Interestingly, over expression of EPHA2 (which we have identified to have an altered expression level in EFMR females) decreases dependency of breast cancers to estrogen and reduces their sensitivity to tamoxifen (Lu et al., 2003). This could be particularly relevant in breast cancers presenting later in life, post menopause, when estrogen levels decrease.
5.4.6 PCDH19 /Estrogen receptor (ER) tissue specific interaction

Given that EFMR presents clinically with only defects in brain function it suggests that the hypothesised PCDH19 ER interaction is important for correct brain development, but is redundant in other tissues. EFMR patients do not present with any major problems in sexual development, affected females are phenotypically female and reproductively fit. These females have not yet been interrogated for more subtle estrogen related defects.

Protocadherins do predominantly function in the brain however the δ-protocadherin subfamily (which includes PCDH19) are more widely expressed throughout the body. The lack of phenotype in additional tissues which express PCDH19 in the EFMR females suggests that other protocadherins or even cadherins are able to compensate for defective PCDH19 functioning in these tissues. However, this compensation is not fully achieved in the brain. Alternatively, the required interaction between PCDH19 and ER could be tissue specific, as has previously been seen for N-cadherin (Monks et al., 2001b).

The apparent tissue specific requirement for the PCDH19 / ER interaction could potentially aid in treating these subjects. If this hypothesis turns out to be correct these results could have significant implications for potential treatment and more importantly prevention of EFMR in individuals conceived with mutations in PCDH19. Selective estrogen receptor modulators (SERM) have been developed for use in treatment of breast cancer (eg tamoxifen, for treatment of ERα+ breast cancers). SERMs are molecules that typically decrease estrogen action in breast cancer, but can still maintain the beneficial effects of estrogen in other tissues (Sengupta and Jordan, 2008). Recently additional SERMs have been designed which function in tissues other than the breast. Bazedoxifene acetate is an oral, nonsteroidal, SERM that acts
an as estrogen agonist in skeletal and lipid metabolism but has no effect on breast tissue (Kung et al., 2009). Clinical trials with bazedoxifene/conegenated estrogens have proven it effective for the treatment of postmenopausal osteoperosis (Kung et al., 2009). Given that EFMR females seizures offset at puberty with the increase in estrogen, an agonist of estrogen could be used to prevent seizure onset in EFMR females. This would have to be a specific agonist to PCDH19 dependent neurons to ensure no harmful side effects. If these girls with PCDH19 mutations were treated through until they reached puberty, then the onset of seizures might be avoided. Absence of seizures may also prevent the development of ID. This approach would be dependent on the hypothesis that EFMR females have a defect in their ability to respond correctly to estrogen stimulation being correct and that the defect is specific to tissues that PCDH19 is functionally important.

5.4.7 Estrogen and the male brain

Identification of a possible defect in certain cells ability to respond to estrogen stimulation in affected EFMR females with mutations in PCDH19 raises the question of, how does a defect in estrogen responsiveness cause epilepsy and ID in the PCDH19 mosaic male? Whilst estrogens are predominantly considered a female specific hormone, estrogen is also present in males, all be it at much lower levels. Estrogen in males is primarily produced as a by-product of the aromatization of testosterone. Brain region specific conversion of testosterone to estrogens is involved in correct formation of brain structures, neuronal organization and behavioral sex differences in males (reviewed in Lombardi et al., 2001). Most sexually dimorphic areas of the brain contain heightened levels of both aromatase cytochrome P450 (CYP19), the enzyme involved in conversion of testosterone to estrogen, and high densities of estrogen receptors (Schwarz and McCarthy, 2008).
An ERβ knockout mouse model provides further evidence implicating a role for estrogen in the male brain. ERβ\(^{+/−}\) mice (male and female) exhibit several morphological abnormalities in the brain (Wang et al., 2001). These include regional neuronal hypocellularity in the brain, severe neuronal deficit in the somatosensory cortex (especially layers II, III, IV and V) and a pronounced proliferation of astroglial cells in the limbic system but not the cortex (Wang et al., 2001). These defects are evident by 2 months of age and become more pronounced as the mice age, with degradation of neuronal cell bodies throughout the brain (Wang et al., 2001). Estrogen is also involved in rapid modulation of synaptic plasticity in hippocampal neurons in male and female rodents (Ogiue-Ikeda et al., 2008).

Aged human males with \(CYP19\) mutations and ER mutations suffer from osteopenia, tall stature and open epiphysis. These defects are corrected by estrogen treatment (Grumbach and Auchus, 1999, Lombardi et al., 2001, MacGillivray et al., 1998). Defects in estrogen response in males have also been associated with glucose intolerance, hyperinsulinemia and lipid abnormalities (MacGillivray et al., 1998, Grumbach and Auchus, 1999, Lombardi et al., 2001). It is interesting to note that whilst a role for estrogens has been identified in the correct functioning of the male brain human patients with \(CYP19\) or ER mutations do not have a defect in brain function. Likewise no defects in brain function have been identified in transmitting EFMR males predicted to have complete loss of function of \(PCDH19\) (Appendix A, Dibbens et al., 2008). Yet mosaic deletion of \(PCDH19\) results in epilepsy and ID in a male (Depienne et al., 2009a), implicating a potential requirement for both disrupted estrogen regulation and cellular mosaicism in epilepsy and ID. Expression profile analysis of the \(PCDH19\) mosaic male will be instructive in determining if a defect in estrogen responsiveness also occurs in mosaic males with mutations in \(PCDH19\).
Continued assessment of the PCDH19 mosaic male through puberty could also aid validation of a defect in estrogen responsiveness. We believe that the puberty driven surge in estrogen levels in females facilitates the offset of epilepsy. Estrogen levels in males also increase during puberty with peak levels occurring 3 years after onset of puberty (Klein et al., 1996). It will be interesting to see if puberty also facilitates offset of epilepsy in the PCDH19 mosaic male, who is currently 7 years old (Depienne et al., 2009a).

5.4.8 Sexual dimorphism and disease

There is an increasing body of evidence in the literature that indicates a significant contribution from an individual's sex towards the onset, progression and severity of most common human diseases (reviewed in Ober et al., 2008). Sex differences have been described for a number of neurological and psychiatric disorders. For example neural tube defects are more common in girls whilst autism and stuttering are more common amongst boys (reviewed in Ober et al., 2008). Approximately 14% of genes expressed in the brain have been identified to have sexually dimorphic expression (Yang et al., 2006). This is believed to be a considerable underestimate of the true number of dimorphic genes in the brain as the whole brain was used in their analysis. Substantial physiological differences occur between the sexes in specific brain regions (Arnold and Burgoyne, 2004) and some of these changes will have been masked in their analysis of the whole brain (Yang et al., 2006). The sexual signature of brain expression is evolutionarily conserved across primates (Reinius et al., 2008) reinforcing the functional significance of these expression differences.

Sexual dimorphic expression of genes is highly tissue specific. Analysis of sexually dimorphic genes across the liver, muscle and adipose tissues revealed only 7.6% of genes were identified as sexually dimorphic in all three tissues, compared to an overlap of 23.5% seen in all active genes (Yang et al., 2006). Therefore it will be important to validate our
findings in a neuronal model. Although due to the brain region specific nature of physiological differences between the sexes (Arnold and Burgoyne, 2004) the model used will need to be closely related to the specific region of the brain disrupted in EFMR females, which is currently unknown.

Whilst sexual dimorphic expression is tissue specific, there is evidence to suggest that at least some of the genes identified in our analysis are also sexually dimorphic in brain. \textit{LAMB1} and \textit{MET} were identified in the 24 genes deregulated in the EFMR \textit{vs} control female and the male control \textit{vs} female control (P<0.01) analysis. Both of these genes have previously been associated with susceptibility to autism (Bonora et al., 2005, Campbell et al., 2008, Campbell et al., 2006, Hutcheson et al., 2004). Autism spectrum disorder shows a prominent sex bias, with a male:female ratio of idiopathic autism estimated at 4-10:1. This ratio increases as the intelligence of the affected individuals increases (Folstein and Rosen-Sheidley, 2001). Association between one of the genetic variants in \textit{LAMB1} and autism susceptibility was heightened in male sibling pair families, suggesting a sex-related effect for this variant (Bonora et al., 2005).

All of the 113 genes (Appendix table C5) identified in both the EFMR \textit{vs} control female and the male control \textit{vs} female control (P<0.01) analyses are therefore good candidate genes for disorders that demonstrate differences between the sexes in the onset, progression and/or severity of disease.

\section*{5.5 Conclusions}

The results identified from the microarray expression profile analysis of mutations in \textit{PCDH19} in EFMR have revealed a possible interaction between \textit{PCDH19} and estrogen
responsive genes. These findings lead us to hypothesis that mutations in \textit{PCDH19} disrupt the ability of some cells to respond to estrogen stimulation. Where mutations in \textit{PCDH19} disrupt a cells ability to respond to estrogen stimulation, this results in deregulation of a number of estrogen responsive genes. Consequently, affected EFMR females appear at the gene expression level to be male. If this hypothesis turns out to be correct these results have significant implications for potential treatment and more so prevention of EFMR in individuals conceived with mutations in \textit{PCDH19}. The identification of a number of genes which exhibit sexual dimorphism could provide valuable insight into the underlying mechanisms responsible for differential disease progression seen in males and females.

\section*{5.6 Contributions}

L. Vandeler (Department of Genetic Medicine, Women’s and Children’s Hospital, Adelaide) established the fibroblast cell lines and was instrumental in obtaining a number of control fibroblast cell lines.

Clinicians involved in obtaining skin punches – Z. Afawi (Department of Neurology, Tel Aviv Sourasky Medical Centre, Israel), T. Lerman-Sagie and D. Lev (Metabolic Neurogenetic Clinic, Wolfson Medical Centre, Israel). I. Scheffer and S. Berkovic (Epilepsy Research Centre and Department of Medicine, University of Melbourne). E. Hann (Department of Genetic Medicine, Women’s and Children’s Hospital).
Chapter 6 – Ephrins / Ephs and their relevance to EFMR

6.1 Introduction 266

6.1.1 Ephrin/Eph in axon guidance 266

6.1.2 Ephrin / Eph in cataracts 269

6.2 Results 271

6.3 Discussion 282

6.3.1 Sexual dimorphic expression of EFNA5 and EPHA2 282

6.3.2 Differential regulation of EFNA5 and EPHA2 283

6.3.3 A role for PCDH19 in EphrinA/EphA bi-directional signalling? 284

6.3.4 Implications of cellular mosaicism in EphrinA/EphA and PCDH19 function 288

6.4 Conclusions 290

6.5 Contributions 290
6.1 Introduction

Ephrin A5 (EFNA5) was identified from the microarray expression profiling of EFMR patients’ skin fibroblasts as a gene of particular interest (chapter 5). EFNA5 was significantly down regulated in the EFMR females. As introduced earlier (chapters 1 and 3), mutations in another member of this gene family, EFNBI, have been identified as causative for CFNS (Twigg et al., 2004, Wieland et al., 2004). CFNS is the only other disorder, currently known, with a similar inheritance pattern to EFMR. The identification of EFNA5 as a significantly down regulated gene and one of its cognate receptors EPHA2 as a significantly up regulated gene further reinforces our hypothesis that the molecular mechanisms underlying these two disorders are likely to be similar (see chapter 3).

6.1.1 Ephrin/Eph in axon guidance

Of relevance to EFMR is the role of Eph-ephrin signalling in axon guidance. The first suggestion that ephrins were involved in axon guidance arose from the findings that gradients of EphA2 and EphA5 exist in the tectum, the target destination for axons originating from the retina (Cheng et al., 1995, Drescher et al., 1995). Subsequently the ephrin family of receptors and ligands has been identified as essential for correct formation of a number of axon pathways including thalamocortical, geniculocortical and hippocamposeptal projections, as well as retinal axon crossing in the optic chiasm (Cang et al., 2005, Cang et al., 2008, Gao et al., 1999, Gao et al., 1998, Gao et al., 1996, Mann et al., 2002, Marcus et al., 2000, Williams et al., 2003, Yue et al., 1999).

Efna5 is involved in the regulation of hippocampal synaptic plasticity in the adult brain, a process that is associated with learning and memory (Gao et al., 1998). Efna5 is believed to function as a repulsive guidance cue for limbic axons, preventing them from invading the...
somatosensory cortex (Gao et al., 1998). This role was further validated by studies performed on *Xenopus laevis* retinal axons (Weinl et al., 2003).

*Efna5* has also been implicated in post-stroke axonal sprouting in the aged cortex (Li and Carmichael, 2006), with increased mRNA and protein levels of *Efna5* in aged brains post stroke, when compared to young adults. One of the major responses to stroke is increased axonal sprouting, which is involved in reconnecting and re-mapping the cortex adjacent to the point of origin of the stroke (Li and Carmichael, 2006). Aged individuals typically have a higher incidence of stroke and a worse prognosis for recovery.

*Efna5* repels axons *in vitro* and has a graded expression in the superior colliculus (SC) the major hindbrain target of retinal ganglion cells. These properties implicated *Efna5* in the formation of topographic maps, a fundamental organisational feature of the nervous system. To test this hypothesis, knockout *Efna5*−/− mice were generated (Frisen et al., 1998). The majority of the *Efna5*−/− mice develop to adulthood, are morphologically intact, and have normal anterior-posterior patterning of the midbrain (Frisen et al., 1998). However within the SC, retinal axons establish and maintain dense arborizations at topographically incorrect sites. These sites correlate to locations of low expression of the related ligand *EphA2* (Frisen et al., 1998). In addition, retinal axons transiently overshoot the SC and extend aberrantly into the inferior colliculus (IC) (Frisen et al., 1998). Mice lacking *Efna5* demonstrate the important role for this ligand in ensuring proper guidance and topographic organisation of retinal axons in the midbrain, (Frisen et al., 1998) primarily through repulsive cues instigated by ephrin/Eph signalling.

Given all of the *in vitro* evidence implicating *Efna5* as a gene involved in the fundamental organisation of neurons in the nervous system, it is surprising that the *Efna5*−/− mice do not
have more severe brain defects. Gene expression studies performed on Efna5−/− mice compared to wild type mice identified that 20% of ephrin/Eph genes expressed in the somatosensory cortex are up regulated in the absence of Efna5 (Peuckert et al., 2008). Up regulation of some of the ephrin/Eph genes was restricted to specific cortical layers, with other layers showing no expression changes of the same ephrin/Ephs. This suggested that a compensatory role of ephrin/Ephs during brain development after deletion of a single gene is somewhat tissue specific (Peuckert et al., 2008).

The redundancy in the ephrin/Eph system has been illustrated also by various other ephrin/Eph mouse models. Ephrin/Eph mutant mice generally exhibit only subtle defects or no observable miss-wiring of neuronal circuits in brain structures (Birgbauer et al., 2000, Feldheim et al., 2000, Mendes et al., 2006). The absence of severe defects in brain function as a consequence of ephrin/Eph gene mutations is typically explained by redundancy in the ephrin/Eph signalling network (Gale et al., 1996, Janis et al., 1999). Mouse models with multiple ephrin/Ephs disrupted result in more severe defects than the corresponding single mutations (Cang et al., 2005, Feldheim et al., 2000, Birgbauer et al., 2000, Mendes et al., 2006).

Interestingly a sub-population of Efna5−/− mice (17%) displays neural tube defects (Holmberg et al., 2000). This sub-population display defects in the dorsal midline of the head with various degrees of penetrance. Mild cases display a hematoma or a small aperture in the dorsal midline of the cranium with discrete protrusion of the brain tissue covered by meninges and skin. More severely-affected mice have completely open cranions with cleft nose and palate, anencephaly, and absence of pituitary gland, and trigeminal ganglia (Frisen et al., 1998). These severely affected mice die immediately after birth (Frisen et al., 1998). These defects resemble anencephaly in humans (Holmberg et al., 2000). As is seen in humans there
is a predominance of females with the more severe neural tube defects (Kallen et al., 1994). Of the Efna5⁻/⁻ mice with neural tube defects 71% were female, suggesting that Efna5⁻/⁻ mice may serve as a useful model for human neural tube defects (Holmberg et al., 2000). This also suggests that gender is a contributing factor to the phenotypic outcome resulting from deletion of Efna5. Additional genetic and environmental factors must also be involved in determining the final phenotypic outcome of these mice. Given that some of the severely affected mice are male, gender is not the only contributor to the neural tube defects.

As introduced in the previous chapter (chapter 5) EPHA2 was also identified in the microarray analysis as significantly up regulated in the EFMR females. EPHA2 preferentially binds EFNA1 and EFNA5 (Gale et al., 1996). Targeted knockout of Epha2 in mice results in substantially elevated susceptibility to skin tumour development (Guo et al., 2006).

### 6.1.2 Ephrin / Eph in cataracts

The interaction between EFNA5 and EPHA2 is of interest in EFMR as this interaction has previously been identified as required for recruitment of N-cadherin to cell-cell contacts (Cooper et al., 2008). Defects in this interaction have been associated with disruption to lens fibre cell packing resulting in cataracts in mice and humans (Cooper et al., 2008, Jun et al., 2009). With 87% of Efna5⁻/⁻ mice developing cataracts by 6 months of age (Cooper et al., 2008), 80% of EphA2⁻/⁻ mice developing cataracts by 12 months of age (Jun et al., 2009). No cataracts were identified in the Efna5⁺/−, Epha2⁺/− or WT mice (Cooper et al., 2008, Jun et al., 2009). Epha2 was identified as the Eph receptor responsible for mediating Efna5 function in the lens, Efna5⁻/⁻ mice had significant down regulation of Epha2 (Cooper et al., 2008). The authors conclude that Efna5 activates Epha2 leading to increased recruitment of β-catenin and localisation of N-cadherin to the membrane (Cooper et al., 2008). Subsequently mutations in
EPHA2 have been identified as causative for cataracts in humans, (Shiels et al., 2008, Zhang et al., 2009, Jun et al., 2009) no mutations have been reported to date in EFNA5.

A number of lines of evidence exist which suggest that EFNA5 and EPHA2 could be important genes in understanding EFMR. Both in terms of understanding the mechanism involved in the unusual inheritance pattern of EFMR and the underlying defective in axonal guidance which leads to the EFMR phenotype. EFMR and CFNS are the only two X-linked disorders to display counterintuitive sex specific phenotypic severity with females more severely affected than males. Cellular mosaicism has also been postulated as the mechanism responsible for both of these disorders. The identification of causative mutations in the X-linked EFNB1 gene in CFNS demonstrates how mosaic expression of cell-cell adhesion genes can differentially affect biological function leading to more severe phenotypes in heterozygous females. The involvement of EFNA5 and EPHA2 in both of these scenarios suggests that these genes maybe contributors to the EFMR phenotype when there expression is disrupted in response to PCDH19 mutations.
6.2 Results

In light of the evidence to support a role for ephrin/Eph signalling in EFMR we investigated the microarray results further for these two genes. Figure 6.1 illustrates the dot-plot results for the expression of these genes across alternative data analysis combinations. *EFNA5* was significantly down regulated 2.91 fold in the EFMR females compared to control females with a P-value of 0.0002. *EPHA2* was significantly up regulated 1.5 fold in the EFMR females compared to control females with a P-value of 0.035.

*EFNA5* and *EPHA2* demonstrated sex specific (sexually dimorphic) expression between the control male and control females. *EFNA5* was down regulated 2.49 fold in the male controls compared to control females with a P-value of 0.0092. *EPHA2* was up regulated 1.71 fold in the control males with compared to control females a P-value of 0.0001. The expression level of both genes was similar in EFMR females and control males.

*EPHA2* was also found to have differential expression between the transmitting males and control males. *EPHA2* expression was down regulated by 1.52 fold in the transmitting males compared to control males with a P-value of 0.0048. This was not seen for *EFNA5*. These expression differences were validated by RT-qPCR, table 6.1 summarises the results obtained.
Figure 6.1 EFNA5 and EPHA2 expression identified in the microarray analysis

Dot plot of the expression of EFNA5 and EPHA2. The left panel demonstrates the different levels of expression between EFMR and control females. Here EFMR females are illustrated by a red circle and box and whiskers. Control females are illustrated by a blue circle and box and whiskers. The middle panel demonstrates how the different sex controls gene expression levels correlates to the expression level of EFMR females. In this panel females are illustrated by a red circle and males with a blue circle. The green box and whiskers represents the affected EFMR females, while the purple box and whiskers represents the unaffected control samples. The right-hand panel demonstrates the expression difference of EPHA2 between transmitting males in red circles and box, and control males in blue circles and box. The box plots show 90% confidence intervals with the whiskers demonstrating the upper and lower 10%.
Comparison | Gene symbol | Expression difference
---|---|---
EFMR females vs Control females | EFNA5 | ↓ in EFMR females
 | EPHA2 | ↑ in EFMR females
Control males vs Control females | EFNA5 | ↓ in Control males
 | EPHA2 | ↑ in Control males
Transmitting males vs Control males | EFNA5 | ↑↑ in Transmitting males but did not reach significance
 | EPHA2 | ↓↓ in Transmitting males but did not reach significance

Table 6.1 Summary of the RT-qPCR validation of the microarray results

Figure 6.2 illustrates the results of RT-qPCR between control females and transmitting females. *EFNA5* was down regulated by 4.77 fold in the EFMR females with a P-value of 0.0098. *EPHA2* was up regulated by 2.50 fold in the EFMR females with a P-value of 0.020.

The fold changes identified in the RT-qPCR were on the whole larger in the real-time validation than what was identified in the micro-array analysis. We believe the RT-qPCR results are more representative of the true fold change differences, as the microarray data undergoes a number of normalisation processes resulting in a more conservative estimation of the fold changes. The increased sensitivity of RT-qPCR also tends to result in less significant P-values being obtained due to the increased variation detected between samples.
Figure 6.2 *EFNA5* and *EPHA2* expression across the control female and EFMR females analysed

Graphical representation of the RT-qPCR results obtained for *EFNA5* and *EPHA2* expression across control females (purple bars) and EFMR females (teal bars). The Y-axis represents gene expression level standardised to Actin-β. Individual samples are plotted across the X-axis, with the particular PCDH19 mutation indicated.

We also performed RT-qPCR comparing the expression between control males, control females and EFMR females. In order to fit this analysis on a 96-well plate RNA from EFMR females with the same mutation was combined. We also added in a combined sample of all of the transmitting male samples for reference. The result of this analysis is illustrated in figure 6.3. *EFNA5* was significantly down regulated in the control males when compared to the control females, with a P-value of 0.00025. *EPHA2* was up regulated in the control males when compared to control females with a P-value of 0.019.
Figure 6.3 EFNA5 and EPHA2 expression comparison across the sexes

Graphical representation of the RT-qPCR results obtained for EFNA5 and EPHA2 expression across control males (navy bars), control females (pink bars), EFMR females (purple bars) and transmitting males (teal bars). The Y-axis represents gene expression level standardised to Actin-β. Individual samples are plotted across the X-axis, with the particular PCDH19 mutation indicated. The n numbers on the x-axis indicate the number of RNA samples used in obtaining the gene expression level.

In addition we tested expression of EFNA5 and EPHA2 between transmitting males and control males to see if we could validate the differences in expression of EPHA2 seen in the microarray (figure 6.4). No significant difference was identified for either gene with P-values of 0.19 and 0.10 for EFNA5 and EPHA2 respectively. There was however a trend towards higher expression of EFNA5 in the transmitting males and decreased expression of EPHA2 however the small number of samples and the relatively small fold-changes meant these differences did not hold up to statistical analysis due to the increased level of variation detected by RT-qPCR.
Figure 6.4 *EFNA5* and *EPHA2* expression comparison across the transmitting males and control males

Graphical representation of the RT-qPCR results obtained for *EFNA5* and *EPHA2* expression across control males (navy bars), transmitting males (teal bars), control females (pink bars) and EFMR females (purple bars). The Y-axis represents gene expression level standardised to Actin-β. Individual samples are plotted across the X-axis, with the particular *PCDH19* mutation indicated. The n numbers on the x-axis indicate the number of RNA samples used in obtaining the gene expression level.

Given the identification of a role for *EFNA5* and *EPHA2* in cataracts we tested the expression of these genes in male and female human lens samples, to see if these genes also displayed sexual dimorphic expression. We had cDNA from 3 male (MC1-3) and 3 female (FC1-3) aged human lens samples (65-74 years old). We assessed expression in aged lenses based on published reports of mutations in *EPHA2* causing age-related cortical cataracts (Shiels et al., 2008, Jun et al., 2009), and that age related cataracts show a female bias (Kanthan et al., 2008). We identified increased expression of both *EFNA5* and *EPHA2* in the female lens samples (figure 6.5). *EFNA5* had a fold change of 2.12 in females when compared to males and a P-value of 0.03. *EPHA2* had a fold increase of 2.67 in females with a P-value of 0.067. Whilst *EPHA2* did not quite reach statistical significance of 0.05, there is a trend towards increased expression in females present. The assessment of additional samples will likely decrease the level of variation detected within each sex, thereby decreasing the P-value.
identified. The expression level of \textit{EPHA2} was dramatically higher than that of \textit{EFNA5} in the lens samples.

![Graphical representation of RT-qPCR results for EFNA5 and EPHA2 expression comparison across male and female human lens samples](image)  

\textbf{Figure 6.5} \textit{EFNA5} and \textit{EPHA2} expression comparison across male and female human lens samples  

Graphical representation of the RT-qPCR results obtained for \textit{EFNA5} and \textit{EPHA2} expression across male (navy bars) and females (pink bars) aged human lens samples. The Y-axis represents gene expression level standardised to Actin-\(\beta\). Individual samples are plotted across the X-axis.

We also assessed the expression of \textit{Epha2}, \textit{Efna5} and \textit{Pcdh19} in a model of neuronal differentiation through embryonic stem cells (ES) (Bouhon et al., 2005, Ying et al., 2003). To investigate relative levels of expression between \textit{Epha2}, \textit{Efna5} and \textit{Pcdh19} during early stages of neuronal differentiation, we assessed expression at three differentiation stages: DS 0, DS 8 and DS 12. At DS 0 the cell population is comprised of ES cells. D8 consists of a mixed population of neurogenic progenitors (nNPs) competent to initiate neuronal production and neurogenic rosettes with neural progenitors organised in neural tube-like rosettes, which are actively engaged in neurogenesis and are characteristic of progenitors in the final phase of commitment to differentiation (Abranches et al., 2009). At DS 12 the culture is predominantly composed of differentiated neurons (Abranches et al., 2009).
*Epha2* was most highly expressed at DS 0 then decreased significantly (figure 6.6), fold change of -6.2, at DS 8 and remained low at the DS 12 stage. *Efna5* had highest expression at DS 8 (figure 6.6). *Efna5* was lowly expressed at DS 0 then increased expression by a fold change of 31 at DS 8 then decreased expression by 1.7 fold at DS 12. *Pcdh19* expression increased across the time points with the highest level of expression at the DS 12 stage (figure 6.6).
Figure 6.6 *Epha2, Efna5* and *Pcdh19* expression across different time points in a model of early neuronal differentiation

Graphical representation of the RT-qPCR results obtained for *Efna5, Epha2* and *Pcdh19* expression across the primary hippocampal neuronal differentiation model. Navy bars demonstrate expression levels at DS 0, teal bars for DS 8 and purple bars for DS 12. The Y-axis indicates gene expression level standardised to Actin-β.
Analysis of the significantly altered (P<0.05) gene list obtained from the EFMR vs control female analysis, using the Ingenuity pathway software, identified additional genes involved in the ephrin/Eph signalling disrupted in EFMR females. A number of these genes also displayed sexual dimorphic expression. Figure 6.7 demonstrates the genes involved in the ephrin/Eph signalling pathway, deregulated genes are also illustrated. Table 6.2 outlines the P-values and fold changes identified for these genes in the microarray analysis.

![Ingenuity – Ephrin receptor signalling pathway](image)

**Figure 6.7 Ingenuity – Ephrin receptor signalling pathway**

Ephrin receptor signalling pathway as illustrated by the Ingenuity pathway diagram. Genes deregulated in the EFMR vs control female analysis P<0.05 gene list are coloured in pink. Genes with increased expression in the EFMR females are illustrated in purple. The genes with decreased expression in EFMR females are illustrated in teal.
<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>EFMR vs control female</th>
<th></th>
<th>Control male vs control female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P-value</td>
<td>Fold change</td>
<td>P-value</td>
</tr>
<tr>
<td>EFNA5</td>
<td>ephrin-A5</td>
<td>0.00018</td>
<td>-2.908</td>
<td>0.0092</td>
</tr>
<tr>
<td>EPHA2</td>
<td>EPH receptor A2</td>
<td>0.035</td>
<td>1.50</td>
<td>0.0001</td>
</tr>
<tr>
<td>ITGA1</td>
<td>integrin, alpha 1</td>
<td>0.026</td>
<td>2.51</td>
<td>0.034</td>
</tr>
<tr>
<td>ITGA3</td>
<td>integrin, alpha 3</td>
<td>0.0042</td>
<td>4.15</td>
<td>0.0048</td>
</tr>
<tr>
<td>ITGA6</td>
<td>integrin, alpha 6</td>
<td>0.0030</td>
<td>6.75</td>
<td>0.0097</td>
</tr>
<tr>
<td>ITGA11</td>
<td>integrin, alpha 11</td>
<td>0.027</td>
<td>2.02</td>
<td>-</td>
</tr>
<tr>
<td>PDGFC</td>
<td>platelet derived growth factor C</td>
<td>0.0031</td>
<td>3.89</td>
<td>-</td>
</tr>
<tr>
<td>CXCL1</td>
<td>chemokine (C-X-C motif) ligand 1</td>
<td>0.020</td>
<td>-3.75</td>
<td>-</td>
</tr>
<tr>
<td>CXCL12</td>
<td>chemokine (C-X-C motif) ligand 12</td>
<td>0.019</td>
<td>-3.08</td>
<td>0.043</td>
</tr>
<tr>
<td>CXCR7</td>
<td>chemokine (C-X-C motif) receptor 7</td>
<td>0.0020</td>
<td>-2.49</td>
<td>0.0034</td>
</tr>
<tr>
<td>ABL1</td>
<td>c-ABL oncogene 1, receptor tyrosine kinase</td>
<td>0.033</td>
<td>-1.57</td>
<td>-</td>
</tr>
<tr>
<td>STAT3</td>
<td>signal transducer and activator of transcription 3</td>
<td>0.011</td>
<td>-1.69</td>
<td>0.0054</td>
</tr>
<tr>
<td>ARHGAP20</td>
<td></td>
<td>0.021</td>
<td>-1.55</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 6.2 Genes from the ingenuity ephrin/Eph-signalling pathway significantly deregulated in the EFMR females

Genes identified in the Ingenuity ephrin/Eph-signalling pathway that are disrupted in EFMR females. The P-value and fold change identified for these genes in the control female vs control male analysis are also shown. Fold changes and P-values were ascertained from the microarray analysis. Fold change for the EFMR/control female analysis represent EFMR relative to control females. Fold change for the control female/control male analysis represent control female relative to control males. A – represents no significant difference identified between the sex controls.

Given the number of genes disrupted in the ephrin/Eph signalling pathway and the sexual dimorphic expression of EFNA5 and EPHA2 we suggest that the ephrin/Eph signalling pathway is perturbed by PCDH19 mutations. The Ephrin/Eph pathway, in addition to PCDH19 might then be a factor in moulding the EFMR phenotype.
6.3 Discussion

6.3.1 Sexual dimorphic expression of EFNA5 and EPHA2

Employing the technology of microarray expression profiling and RT-qPCR we have identified deregulated expression of EFNA5 and EPHA2 in EFMR females. These genes also show sexually dimorphic expression with EFMR females having a similar expression of both of these genes to that seen in the control male samples. As mentioned in chapter 5, EPHA2 expression has been negatively associated with estrogen levels (Zelinski et al., 2001, Zelinski et al., 2002). It is therefore not surprising that the control females have lower expression of EPHA2 than control males. Conversely, females have increased expression of EFNA5. This result is consistent with the observation in Efna5−/− mice that the majority (71%) of the subpopulation of severely affected mice with neural tube defects were female (Frisen et al., 1998, Holmberg et al., 2000). This indicates that being female might increase phenotypic severity of EFNA5 loss of function.

Sexual dimorphic expression of EFNA5 and EPHA2 was maintained in the aged human lens samples. This finding could be particularly relevant to age-related and congenital cataracts where mutations have been identified in EPHA2 (Jun et al., 2009, Shiels et al., 2008, Zhang et al., 2009). The sexual dimorphic expression of EFNA5 and EPHA2 in aged lens suggests that gender could be a predisposing factor for age-related cataracts. Increasing evidence is emerging in the literature demonstrating the role of gender in cataracts. Anterior polar cataracts (APC) have been identified to have sex skewing with a bias towards males. Seventy eight percent of patients with APC were male (Kim and Joo, 2008). For all other cataract sub-types investigated there was a bias towards affected females. Between 56%-75% of patients with various sub-types of cataracts are female (Kim and Joo, 2008). Sexually dimorphic expression in which females express higher levels of EPHA2 and EFNA5 might, at least in
part, be responsible for the female bias seen in various sub-types of cataracts. Indicating a possible role for \textit{EPHA2} and \textit{EFNA5} in sub-types of cataracts with a female bias.

In the lens tissue both \textit{EFNA5} and \textit{EPHA2} had elevated expression in females. The direction of fold change for \textit{EPHA2} in the lens was opposite to that seen in the fibroblast samples. Analysis of sexually dimorphic genes across multiple tissues revealed that only 37\% (27/71) of sexually dimorphic genes showed consistent direction, with a male or female bias, across the three tissues analysed (Yang et al., 2006). Interestingly, in the lens, loss of \textit{EFNA5} results in down regulation of \textit{EPHA2} whereas in the fibroblasts decreased expression of \textit{EFNA5} coincides with increased expression of \textit{EPHA2}. This indicates that the direction of regulation (up or down regulation) between \textit{EPHA2} and \textit{EFNA5} might be tissue dependent.

\subsection{6.3.2 Differential regulation of \textit{EFNA5} and \textit{EPHA2}}

\textit{EFNA5} and \textit{EPHA2} illustrated inverse expression levels across the three time points of neuronal differentiation analysed, such that at DS 0 when \textit{EPHA2} was highly expressed \textit{EFNA5} was very lowly expressed. When \textit{EFNA5} expression increased at DS 8 \textit{EPHA2} expression decreased. This indicated that the negative regulation of \textit{EPHA2} by \textit{EFNA5} as seen in the fibroblasts could also occur in immature neurons. This is further supported by differential regulation occurring between \textit{EFNA5} and \textit{EPHA2} in different tissues.

Differential regulation of eprin/Ephs has been identified in a model of glioma. Over expression of \textit{EFNA1} resulted in down regulation of \textit{EPHA2} (Liu et al., 2007) indicating that \textit{EFNA1} serves as a negative regulator of \textit{EPHA2}. Our results suggest that \textit{EFNA5} may also act as a negative regulator of \textit{EPHA2} in fibroblasts and immature neurons but not in the lens. Decreased expression of \textit{EFNA5}, as a result of mutations in \textit{PCDH19} in EFMR females, leads to increased expression of \textit{EPHA2}. Alternatively \textit{EPHA2} could negatively regulate \textit{EFNA5};
however, given the substantially lower P-value and higher fold change identified for altered
EFNA5 expression in EFMR females we speculate that EFNA5 is the driving factor. This does
however require significantly more investigation. The possibility of tissue specific differential
regulation occurring between EFNA5 and EPHA2 reinforces the need to validate our findings
from the microarray analysis derived from fibroblasts (chapter 5) in a neuronal model that
closely relates to the specific brain region disrupted in EFMR females.

Differential regulation of Ephrin/Ephs has also been associated with E-cadherin (Orsulic and
Kemler, 2000). This was observed in an E-cadherin null-mutant embryonic stem cell model.
In this model transfection of E-cadherin decreased Efnb2 and increased Ephb2 expression
(Orsulic and Kemler, 2000).

6.3.3 A role for PCDH19 in EphrinA/EphA bi-directional signalling?

Microarray expression profile analyses of EFMR females identified additional deregulated
genes from the Ephrin/Eph signalling pathway. These included ITGA’s, CXCR, PDGF,
STAT3 and ABL1.

The intergrins, ITGA1, 3, 6 and 11 were all up regulated in EFMR females. ITGA1, 3 and 6
also displayed sexual dimorphic expression with EFMR females expressing similar levels of
these genes to the control males. Interaction between integrin and ephrin/Eph signalling has
been well established with both signalling systems converging on key intracellular effectors
that control cellular protrusions, neurite formation, cell migration and synaptic plasticity
(Chang et al., 1999, Miao et al., 2005, Miyashita et al., 2004, Motegi et al., 2003, Pasquale,
2005, Prevost et al., 2005, Zou et al., 1999). Investigations into gene expression changes in
the somatosensory cortex of Efna5-/- mice detected up regulation of components of the
integrin pathway (Peuckert et al., 2008) supporting the concept of ephrin/Eph and integrin
convergence in which disruption to one pathway leads to altered regulation of the other in order to attain normal developmental (Peuckert et al., 2008).

Chemokines $CXCR7$, $CXCL1$ and $12$ were all down regulated in EFMR females. $CXCR7$ and $CXCL12$ also displayed sexual dimorphic expression with EFMR females expressing similar levels of these genes to the control males. $CXCR7$ is the most recently identified member of the chemokine ligand / receptor component involved in neuronal migration. $CXCR7$ is an alternative receptor for $SDF-1$, stromal cell-derived factor 1 (also known as $CXCL12$) (Balabanian et al., 2005, Burns et al., 2006). SDF-1 and $CXCL4$ (closest relative to $CXCL7$) are involved in cross talk with the ephrin-B reverse signalling whereby EphrinB reverse signalling leads to inactivation of $CXCR4$ and failure of primary cerebral granule cells to migrate in response to $SDF-1$ (Lu et al., 2001). $EphB2$ and $EphB4$ forward signalling, which is required for the formation of capillary-like vascular structures, functions through activation of $SDF-1$ induced signalling (Salvucci et al., 2006). To date a role for only $CXCR4$ has been identified in cross talk between ephrinB/EphB signalling and G-protein coupled chemoattraction (Salvucci et al., 2006).

Our results implicate a possible role for $CXCR7$ in EphrinA/EphA signalling. For this to occur a transmembrane protein must be involved in binding the GPI anchored $EFNA5$ in the extracellular matrix and transmitting a signal to the cytoplasmic region of $CXCR7$. Cadherin-11 ($CDH11$) has been identified as a potential downstream target of $CXCR7$ (Wang et al., 2008a). Expression of $CDH11$ increased with the altered expression of $CXCR7$ associated with more aggressive tumours through enhanced adhesive and invasive activities (Wang et al., 2008a). It is therefore conceivable that $PCDH19$ could also interact with $CXCR7$ in response to ephrin/Eph signalling.
Platelet derived growth factor C (PDGFC) was up regulated in EFMR females. A number of links have been identified between PDGF and ephrin/Eph signalling. Ephrin-B2 is a downstream effector of PDGF signalling (Semela et al., 2008). EphrinB1 has also been identified as an in vivo substrate for PDGF (Bruckner et al., 1997), indicating crosstalk exists between EphrinB1 signalling and tyrosine kinase activated signalling cascades (Bruckner et al., 1997). EphA activation decreases activation of MAPK by EGF, PDGF and VEGF receptors (Miao et al., 2001). The Ras/MAPK cascade has been identified as a point of convergence between EphA and other RTKs, with the resulting counterbalance between these pathways helping to maintain homeostasis (Miao et al., 2001).

STATs, STAT3 and STAT5A, were down regulated in EFMR females. STAT3 also exhibits sexually dimorphic expression. EphrinB1 interacts with STAT3 in a tyrosine phosphorylation dependent manner leading to enhanced activation of STAT3 transcriptional activity. Enhanced activity of STAT3 is dependent on the tyrosine kinase Jak2, and two tyrosines within the intracellular domain of ephrinB1, which are critical for the association with STAT3 and its activation. This reveals a signalling pathway from ephrinB1 to the nucleus (Bong et al., 2007). Even though EphrinAs do not have a cytoplasmic domain, activation of EphrinA1 and EphA4 is capable of activating Jak2 and STAT3. This identifies Jak/STAT proteins as the novel downstream targets of ephrinA /EphA signalling (Lai et al., 2004).

Additional genes in the Jak/STAT pathway were disrupted in EFMR females. Once again for an interaction between ephrinA and Jak2-STAT3 to occur a transmembrane protein must transduce the signal across the plasma membrane. Cadherin-cadherin interaction activates STAT3 through activation of Rac1 and CDC42 (Arulanandam et al., 2009). Cadherins could be involved in activation of STAT3 in response to ephrin/Eph signalling. Mutations in PCDH19 may disrupt this signalling pathway and expression of STAT3.
ABL1, Abelson muurine leukemia viral oncogene homolog 1, is a proto-oncogene that encodes a cytoplasmic and nuclear protein tyrosine kinase. ABL1 has been implicated in cell differentiation, cell division, cell adhesion and stress response. ABL1 expression was decreased in EFMR females. The ABL family of kinases are a major effector of ephrin-A induced retinal ganglion cell repulsion (Harbott and Nobes, 2005). EphB4 also activates ABL family tyrosine kinases (Noren et al., 2006). EphB4 activation of ABL inhibits breast cancer cell viability and proliferation (Noren et al., 2006).

The genes disrupted in the Ephrin/Eph pathway in the EFMR females implicates a possible role for PCDH19 in ephrinA/EphA forward and reverse signalling. Whereby mutations in PCDH19 inhibit reverse signalling from EFNA5. This leads to increased ITGA expression, decreased CXCR/L expression, decreased expression of STAT3 and decreased expression of ABL1. These expression changes ultimately result in disruption to cell repulsion. This scenario fits well with the identification of abnormal cortical architecture with dysplastic neurons in the cortex and underlying white matter identified in a frontal lobe specimen from an EFMR female (Ryan et al., 1997).

We have as yet been unable to demonstrate a physiological interaction between PCDH19 and EFNA5 or an activation of EFNA5 in response to interaction with PCDH19. This is primarily due to the unavailability of a relevant cell type expressing both genes and proteins to sufficient levels.

One of the major underlying mechanisms of ephrin/Eph bidirectional signalling that remains poorly understood is how GPI linked ephrinA ligands instigate reverse signalling given that they do not have a transmembrane domain. Reverse signalling from ephrinA’s has been
identified in correct path finding of vomeronasal and spinal motor axons and topographic mapping of axons of olfactory neurons and retinal ganglion cells (Cutforth et al., 2003, Knoll et al., 2001, Marquardt et al., 2005, Rashid et al., 2005). In order to reverse signal, ephrinA ligands must associate with transmembrane proteins capable of activating intracellular signalling pathways. EphrinAs localise to caveolae which are microdomains in the membrane (Davy et al., 1999, Simons and Toomre, 2000). Calveola function as platforms to localise signal transduction through the assembly of signalling complexes (Davy et al., 1999, Simons and Toomre, 2000).

p75 neurotrophin receptor (p75\textsuperscript{NTR}) has recently been identified as the first transmembrane protein to interact with EFNA5 and facilitates reverse signalling. This interaction is required for the repulsion of retinal axons by EphA (Lim et al., 2008). Constitutive and retina-specific knockout of p75\textsuperscript{NTR} results in nearly all retinal ganglion cell axons incorrectly mapping anterior to their correct target (Lim et al., 2008). The authors concluded that p75\textsuperscript{NTR} acts as a signalling partner with ephrinAs to mediate the repellent effects of ephrin-A reverse signalling, instigated by binding EphAs (Lim et al., 2008). We speculate that \textit{PCDH19} could be an additional interactor of \textit{EFNA5} and functions in ephrinA reverse signalling whereby mutations in \textit{PCDH19} inhibit \textit{EFNA5} reverse signalling and the repulsive effects on axons instigated by this signal pathway.

### 6.3.4 Implications of cellular mosaicism in ephrinA/EphA and \textit{PCDH19} function

To date two disorders CFNS and EFMR/PCDH19-DS have been associated with a molecular mechanism involving cellular mosaicism resulting in a more severe phenotype in females. The identification of a male affected by PCDH19-DS with a mosaic deletion of \textit{PCDH19}
(Depienne et al., 2009a) and the finding of mutations in PCDH19 only affecting heterozygous females whilst sparing hemizygous males (Chapter 3, Appendix A, Dibbens et al., 2008) implicates a specific role for PCDH19 in situations of cellular mosaicism. Evidence from the EFNB1 mouse demonstrates the effects of cellular mosaicism with a heterogeneous population of EFNB1+ and EFNB1- cells. Cellular mosaicism leads to limb bud and skull vault defects in the heterozygous mice that were not evident in the homozygous or hemizygous knock out mice (Compagni et al., 2003, Davy et al., 2004, Davy et al., 2006). This was recapitulated in humans with EFNB1 mutation and CFNS, whereby heterozygous females are more severely affected than hemizygous males (Grutzner and Gorlin, 1988). The identification of an association between mutations in PCDH19 and disrupted ephrin/Eph signalling suggests that ectopic boundary formation between different cell populations could be involved in EFMR.

Expression profile analysis on the mosaic male with the PCDH19 deletion would be highly instructive for determining disruption or otherwise to ephrin/Eph signalling. The question is whether a heterogeneous population of PCDH19 expressing cells in males is sufficient to cause PCDH19-DS or if the resulting phenotype is also dependent on disruption to ephrin/Eph signalling. Given the required function of ephrin/Ephs in setting up tissue boundaries, it is likely that this signalling system will also be disrupted in the mosaic male. This would lead to incorrect axonal targeting in the brain as incorrect boundaries are formed through the presence of a heterogeneous population of cells with correct ephrin/Eph signalling and defective ephrin/Eph signalling as a result on mutations in PCDH19. Artificial generation of a mosaic population of male PCDH19wt and PCDH19mut fibroblasts cells could also prove informative.
6.4 Conclusions

The similarities present between CFNS and EFMR with regard to the unusual inheritance pattern and the similarities in the mechanism predicted to explain the inheritance pattern in both of these two disorders lead us to expand on the gene expression changes identified in \textit{EFNA5} and \textit{EPHA2} in EFMR females. The identification of a role for both of these genes in cataracts also led us to investigate expression of \textit{EFNA5} and \textit{EPHA2} in human lens samples. Here we found that \textit{EFNA5} and \textit{EPHA2} had sexually dimorphic expression in aged human lens samples with increased expression in females. This increased expression in females could be one of the factors contributing to the overrepresentation of females presenting with cataracts. We also identified differential regulation between \textit{EFNA5} and \textit{EPHA2} occurring in a tissue specific manner.

Further investigations into the microarray data (chapter 5) revealed that mutations in \textit{PCDH19} affect the expression level of a number of genes involved in the ephrin/Eph signalling pathway implicating a potential role for \textit{PCDH19} in ephrin/Eph signalling. These results add to the mounting evidence that suggests similarities exist between EFMR and CFNS at a molecular level.

6.5 Contributions

S. Sharma and J. Craig (Department of Ophthalmology, Flinders University, Adelaide) contributed the lens samples analysed.

L. Jolly (SA Pathology, Department of Genetic Medicine, Women’s and Children’s Hospital, Adelaide) contributed the samples used for analysis of the early neuronal differentiation model.
Chapter 7 – Overall discussion
### 7.1 Genetic architecture of ID and epilepsy

Our understanding of the genetic landscape of inherited disorders has improved dramatically in recent years as a result of the advent of automated high-throughput technologies, which facilitate analysis of large numbers of individuals. Large scale re-sequencing of the X-chromosome used in this investigation to identify the causative gene for EFMR is one example of the use of advanced high-throughput technologies. Technological advances have been instrumental in the discovery of a large number of genes involved in a wide range of disorders, for example in XLID more than 90 genes have been identified as causative (reviewed in Gecz et al., 2009) and in epilepsy more than 25 genes have been identified (reviewed in Reid et al., 2009). Identification of genes involved in these disorders has also enhanced our understanding of the molecular mechanisms and pathways that need to be functioning correctly for normal brain function.

Results of the intensive efforts employed to discover causative genes for XLID have demonstrated just how heterogeneous the underlying genetic causes of XLID are (Freude et al., 2004, Jensen et al., 2005, Kalscheuer et al., 2003a, Ropers and Hamel, 2005, Shoichet et al., 2003, Tarpey et al., 2004, Tarpey et al., 2009). The XLID genes have a wide range of functions including regulation of transcription, signal transduction, regulation of actin cytoskeleton, cell adhesion, ubiquitination, kinase activity, post-translational modification and sodium ion transport (reviewed in Gecz et al., 2009). The vast majority of XLID genes identified only account for a small proportion of affected families, with each XLID gene only contributing to 0.2-0.5% of ID cases (Gecz et al., 2009, de Brouwer et al., 2007, Ropers, 2008). The human brain is our most complex organ, functioning as a highly sophisticated information-processing centre and is sensitive to imbalances in pathways supporting its normal function.
In spite of increased knowledge and ever growing number of genes identified the cause of ID the underlying cause of ID remains unknown for at least 50% of patients (Rauch et al., 2006), at least half of these cases are assumed to have a genetic cause (Battaglia, 2003). The proportion of resolved cases of genetic forms of epilepsy is negligible, for example in common idiopathic generalised epilepsy (IGE), which comprises about 20-30% of epilepsies (Jallon et al., 2001) only 1% of cases are accounted for by microdeletions at 15q13.3 (Helbig et al., 2009).

### 7.2 Genetic architecture of complex disorders

To date the vast majority of success in the identification of causative genes for ID and epilepsy has been predominantly restricted to monogenic forms, such as EFMR, where mutations in a single gene cause the disease phenotype (Appendix A, Dibbens et al., 2008). Success has been achieved in monogenic forms, as the causative genes are generally easier to identify as they are inherited in a Mendelian fashion. The consequences of mutations in monogenic disorders are also generally easier to validate in functional assays as the mutations have large effects on gene function. The large effect of these mutations on gene function also means these mutations are generally highly penetrant and present only in affected individuals and not in controls. The genetic causes of monogenic disorders are generally ascertained through linkage analysis and candidate gene screening in large affected families (Steinlein et al., 1995) or in large numbers of small families with syndromic features and genetic homogeneity (Tsui et al., 1985).

Monogenic forms of ID and epilepsy only account for a small number of affected patients, for example in the 70-80% of genetic forms of IGE only 1-2% of patients are believed to have a
monogenic cause while the vast majority (98-99%) are believed to have an oligo or polygenic cause with complex inheritance (Weber and Lerche, 2008). The genetic causes of the complex forms of ID and epilepsy in general remain elusive in the vast majority of affected patients. The common variants that have been associated with these disorders explain only a very small proportion of affected individuals (Mulley et al., 2005). It is therefore important that the information gained from studies such as the one detailed in this thesis into the genes and pathways involved in specific monogenic disorder are translated over into the polygenic forms of epilepsy and ID. Assessment of these relatively rare monogenic disorders are important for enhancing our understanding of genes and pathways required for correct brain function as these same pathways will likely also be affected in the more prevalent polygenic causes of epilepsy and ID.

Gene identification in complex diseases is an enormously challenging task that requires the use of fundamentally different approaches from those used in monogenic disorders. Complex inheritance involves multiple susceptibility alleles with or without added environmental factors, which in combination, lead to a disease phenotype. In complex disorders the hunt for genetic variants involves searching for variants that are enriched in affected individuals when compared to well matched control cohorts (Hirschhorn and Daly, 2005). Typically the genetic variants associated with complex disorders are insufficient on their own because of small individual effects to cause disease and as a result are not easily amenable to functional validation in a laboratory setting.

A relevant example of the difficulties faced in laboratory based functional validation of potential susceptibility genes in epilepsy is NEDD4-2. NEDD4-2 is a good candidate gene for epilepsy as it regulates cell surface levels of several ion channels, receptors and transporters involved in regulating neuronal excitability. In *Xenopus* oocytes NEDD4-2 regulates the cell
surface levels of voltage-gated sodium channels (Fotia et al., 2004). Three missense changes in highly conserved residues have been identified in 4 out of 80 (5%) families, with photosensitive epilepsies, screened for mutations in NEDD4-2 (Dibbens et al., 2007). Functional tests on these variants in the *Xenopus* oocyte heterologous expression system revealed no differences in their ability to down regulate the voltage-gated channel Na\(_{v}\)1.2 (Dibbens et al., 2007). Dibbens *et al* suggest that the effect caused by the missense variants may be to subtle to be detected by the *in vivo* studies, alternatively the NEDD4-2 variants may cause epilepsy via alternate mechanisms such as axonal guidance or synapse formation (Dibbens et al., 2007). Substantial amounts of work would be required to definitively classify NEDD4-2 as a susceptibility gene for epilepsy.

Genome wide association studies (GWAS) have been employed in a wide range of complex diseases/traits in an attempt to identify common variants implicated in diseases/traits through enrichment in affected patients. These studies have primarily been conducted under the presumption that a small number of common variants each of which has a moderate effect collectively result in disease. This assumption has turned out to be true for only a small number of complex diseases (reviewed in Manolio et al., 2009). Age-related macular degeneration is a very good example of a common disease caused by a small number of variants with 5 loci accounting for 50% of the heritability of this disease (Maller et al., 2006). In Crohn’s disease 32 loci have been associated with the disease; together these loci only explain 20% of the heritability (Barrett et al., 2008). The situation for height is even more complex where 40 loci have been identified and which together account for only 5% of the heritability (Visscher, 2008).

The outcomes from more than 400 GWAS published have demonstrated that the underlying genetics of the majority of complex disorders cannot be explained by a limited number of
common variants of moderate effect but that the underlying genetics is even more complex (reviewed in Manolio et al., 2009). GWAS studies have, however, provided valuable insights into the genetic architecture of complex disorders, with more than 80% of associated variants detected by GWAS falling outside of coding regions, emphasising the importance of non-coding regions in disease (Hindorff et al., 2009) provided that these findings are valid and reproducible. It will be interesting to see if these intronic regions do contain functional SNPs in conserved elements of as yet unknown function, or if they are simply tags for the at risk haplotype which contain coding changes. Deep re-sequencing will be informative in determining the contribution of coding changes by finding functional coding SNPs in linkage disequilibrium with intronic SNPs.

A further area of study that is proving fruitful in enhancing our understanding of the genetics of complex disorders is the study of CNVs. This area of research has significantly advanced in recent years through advances in technology, in particular the introduction of array comparative genomic hybridisation (CGH) (Vissers et al., 2003) and SNP microarrays (Mei et al., 2000, Pollack et al., 1999). Whole genome assessment of CNVs has proven successful in uncovering multiple rare variants in disorders where GWAS has had little success, such as, ID, ASD and schizophrenia (reviewed in Mulley and Dibbens, 2009). It has been estimated that >15% of ID patients carry a large microdeletion or microduplication (reviewed in Mefford and Eichler, 2009).

The whole genome approaches to CNV analysis has also lead to two fundamental changes in thinking regarding the genetics of complex disorders. Firstly, it has demonstrated the advantages of the ‘genotype first’ approach to disorders over the historical ‘phenotype first’ approach (reviewed in Mefford and Eichler, 2009). Secondly, the analysis of CNVs in large
and varied patient cohorts has revealed substantial phenotypic diversity associated with at least a subset of recurrent CNVs (reviewed in Mefford and Eichler, 2009).

The ‘genotype first’ approach to the analysis of complex disorders involves the identification of a common rearrangement followed by assessment of the clinical phenotypes present in individuals with a particular rearrangement (reviewed in Mefford and Eichler, 2009). The genotype first approach has revealed that for at least a proportion of CNVs, there is unprecedented phenotypic heterogeneity associated with a single CNV. This heterogeneity includes no phenotype at all, as part of the spectrum. A lack of phenotype negates their application to prenatal diagnosis. The phenotypic heterogeneity has reiterated the role of modifying factors in complex disorders.

Three new recurrent microdeletions 16p11.2 (Bijlsma et al., 2009, McCarthy et al., 2009, Sebat et al., 2007, Shinawi et al., 2009a), 1q21.1 (Brunetti-Pierri et al., 2008, Christiansen et al., 2004, Consortium, 2008, Mefford et al., 2008, Stefansson et al., 2008) and 15q13.3 (Consortium, 2008, Pagnamenta et al., 2009, Sharp et al., 2006, Sharp et al., 2008, Stefansson et al., 2008, van Bon et al., 2009, Dibbens et al., 2009, Helbig et al., 2009) have been identified as enriched in multiple neuropsychiatric diseases (including ASD, schizophrenia, ID and epilepsy) but have eluded syndromic classification due to the phenotypic diversity associated with the microdeletion. Each of these microdeletions were first identified within clinical disciplines, in a cohort of patients with similar phenotypes; however assessment of CNVs in a wider range of neurocognitive disorders across clinical disciplines revealed microdeletions and microduplications of the same chromosomal region were also enriched in additional disorders.
Microdeletions of chromosome 15q13.3 were initially described in a cohort of patients with ID, dysmorphic features and seizures (Sharp et al., 2006, Sharp et al., 2008) and has since been identified in patients with mild to moderate ID alone, autism patients and is enriched in schizophrenia cohorts (Consortium, 2008, Miller et al., 2009, Pagnamenta et al., 2009, Stefansson et al., 2008, van Bon et al., 2009). Very few of the additional patients identified early on with the 15q13.3 microdeletion suffered from seizures yet this deletion is also enriched in patients with idiopathic generalised epilepsy (Dibbens et al., 2009, Helbig et al., 2009). A smaller recurrent 680 Kb deletion at 15q13.3 has also been identified in four unrelated families with various neurological phenotypes including developmental delay, ID and seizures (Shinawi et al., 2009b). Detailed analysis of the breakpoints revealed identical deletions in individuals with highly variable phenotypes. Collectively these results indicate that rare structural variants are not only associated with complex forms of ID and epilepsy but with complex neurological disorders in general.

An additional factor that is likely to further compound the heterogeneity and variable expressivity seen in many complex diseases is gender. Sex-specific differences in the prevalence, age of onset and/or severity have already been noted for a vast array of disorders. For example the incidence of asthma is higher in boys during childhood years but switches after puberty where there is a higher incidence of new cases in females (Postma, 2007). A similar scenario occurs in cardiovascular disease, which is predominant in men through adulthood but then has higher rates in post-menopause females (Choi and McLaughlin, 2007). Elucidating the underlying mechanisms for these observations is a challenging task; however, we can take advantage of extreme cases of sex-specific monogenic diseases, such as EFMR, to learn more about the mechanisms involved in the interplay of gender and disease susceptibility, which will likely also translate to a better understanding of the genetic architecture of complex disorders.
How can we attempt to fill the gaps?

It is evident that there are major gaps in our understanding of the genetic architecture of complex human disorders. In IGE, the most common predisposing genetic variant identified to date, 15q13.3, only accounts for 1% of IGE patients (Helbig et al., 2009). The minimal heritability of most complex disorders or traits such as height that has so far been attributed to genetic variants clearly demonstrates that we are missing a significant proportion of the heritability in complex disorders (Visscher, 2008).

In order to make significant headway into the genetic landscape of complex disorders we need to understand more about the spectrum of genetic variation present in the human population before we can begin to assess the relevance of certain variants in disease states. Next generation sequencing of large populations of individuals in projects such as the 1000 Genomes project will be instrumental in enhancing our understanding of human genome variation. Extensive analysis of CNVs will also be required in these cohorts as next generation sequencing is incapable of detecting all types of structural variation. One problem with these cohorts however, the lack of phenotypic data associated with the genome variation identified. Caution will need to be exercised when using the genome variation data to support or discount variants identified in certain diseases as causative or benign, as control individuals used in the identification of human variation may actually have the common disease being investigated or more likely some of the susceptibility variations without symptoms.

Next generation sequencing is set to revolutionise human genetics and medicine in general. Targeted and exome sequencing in individuals and families with monogenic forms of ID and epilepsy will likely uncover further genes involved with these disorders, which have so far evaded identification through the traditional linkage and candidate gene approach. Next
generation sequencing will also have significant applications in complex disorders with the ability to detect all nucleotide variation in an individual (Bentley et al., 2008, Wang et al., 2008b, Wheeler et al., 2008) and ultimately in a cohort of individuals with a specific disease. Comparison of the specific variants present in the disease cohort against those present in the human population should reveal novel disease variants or significantly greater numbers of rare variants in the disease group compared with controls (Kryukov et al., 2007). For example, next generation sequencing in combination with genomic locus capture has successfully been used to identify causative mutations in PYCR1 in patients with autosomal recessive cutis laxa (Reversade et al., 2009). Next generation sequencing in combination with exome capture has also been used to identify causative mutations in DHODH in patients with Miller Syndrome (Ng et al., 2009).

Genome wide CNV analysis in extended cohorts of patients and controls should also continue to reveal additional CNVs associated with complex disorders. The impact of CNVs on complex disorders may prove to be a more significant contributor than SNPs as there are more differences between individual humans at the CNV level in terms of the number of nucleotides affected than there are at the level of SNPs (Sharp, 2009).

A further approach, which could be useful for gaining additional insight into the underlying genes involved in complex disorders, is the detection of modifying factors involved in altering the phenotypic severity of individuals with the same genetic mutation. For example, there is a wide range of phenotypic severity seen in females with PCDH19 mutations in terms of their seizures and degree of ID, implying that modifying factors are involved. The factors involved in modifying the seizure and ID phenotype are also likely to be involved as susceptibility alleles in complex forms of epilepsy and ID. A good place to start with determining these modifying factors would be to assess families with a known genetic defect,
such as a mutation in *PCDH19*, where closely related individuals (such as siblings) have substantial differences in the severity of their phenotype. Performing comprehensive sequencing and CNV assessment on closely related individuals (such as siblings) should reveal a number of potential variants or modifying factors that can then be further assessed in more distant relatives (such as cousins) and ultimately in individuals with unexplained epilepsy and ID.

Increased understanding of the genetic variants associated with disease is also likely to facilitate major advances in human health through the use of genetic information to determine the best treatment approach for an individual patient. Epilepsy is a disorder where this work could have major positive impacts on patient outcomes. Assessment of the genetic variants associated with responsiveness to the various anti-epileptic drugs could eliminate a lot of the trial and error currently required to find the most effective drug to manage an individual's seizures. A study of 470 newly diagnosed cases of epilepsy found that only 47% of patients became seizure free on the first antiepileptic drug they were prescribed, while 14% of patients became seizure free during treatment with a second or third drug (Kwan and Brodie, 2000). More than 30% of patients had inadequate control of seizures with drug therapy (Kwan and Brodie, 2000).

The identification of novel disease associated SNPs and CNVs has been streamlined in recent years through advances in technology however we face a potentially major bottleneck in the translation of these discoveries to patients. Will we, as consumers, be satisfied that a specific variant or combination of variants is likely to be disease causing based on statistics alone, rather than from traditional laboratory based methods? At what stage will we be sufficiently convinced of the pathogenicity of a variant to use it as a diagnostic test in patients, or are patients going to be satisfied with the outcome of a test being that they “may” develop a
certain disease? These are just a couple of the fundamental questions that will need to be answered in the coming years as we enter this new era of human genetics which promises to be rewarding but extremely challenging.
Chapter 8 – Conclusions
The identification of *PCDH19* as the gene responsible for EFMR in seven families (Dibbens et al., 2008 + unpublished work) ends 37 years of research that began with the original description of the syndrome in 1971 (Juberg and Hellman, 1971). The role for *PCDH19* in epilepsy and ID has been further validated by the identification of 9 additional mutations in 11 females with sporadic epileptic encephalopathies resembling Dravet syndrome, from a cohort of 45 females (Depienne et al., 2009a). Recent reports of 14 further point mutations and 3 deletions of *PCDH19* identified in 113 female patients with isolated infantile epileptic encephalopathy suggests that *PCDH19* is also an important gene involved in sporadic or isolated epilepsy in females (Depienne et al., 2009b).

The research presented in this thesis addresses the initial aims outlined in the introduction.

1) To perform molecular characterisation of *PCDH19* as the gene responsible for EFMR

Under this aim we identified that *PCDH19* is expressed in a wide range of tissues but is predominantly expressed throughout various regions of the brain. We have also shown that there are two alternate isoforms of *PCDH19* and that both isoforms are expressed in a similar pattern across a range of tissues.

2) To assess the likely protein consequences of the EFMR mutations

Based on protein conservation analysis and previously published work on the NMR structure of a member of the protocadherin family *PCDHα4* we revealed that all of the missense mutations identified in *PCDH19* in EFMR and PCDH19-DS patients affect highly conserved and functionally important residues. We predict that all of the missense mutations identified in *PCDH19* result in complete loss of function of PCDH19 protein. We have also demonstrated that two of the protein truncation mutations in *PCDH19* are subject to NMD and therefore we have predicted that no mutant truncated protein is present in EFMR females.
We suspect that all of the protein truncation mutations indentified in \textit{PCDH19} will also be subject to NMD and degraded.

3) To investigate possible mechanisms responsible for the unusual inheritance pattern seen in EFMR families

Three of the four previously suggested mechanisms to explain the inheritance pattern have been eliminated. We have demonstrated that EFMR females have random X-inactivation ruling out the defect in X-inactivation theory. Our discovery that protein truncation mutations in PCDH19 are subject to NMD eliminates the metabolic interference hypothesis. The finding by Deppiene et al (Depienne et al., 2009a) of an affected male with a mosaic deletion of \textit{PCDH19} eliminates the third theory, that a male specific factor rescue males with mutation in \textit{PCDH19} from epilepsy and ID. Our results confirm the fourth theory, that PCDH19 is not required for correct functioning of the male brain. We have demonstrated that males with truncation mutations in \textit{PCDH19} that are subject to NMD, meaning that these males lack PCDH19 protein, are unaffected by epilepsy and ID.

4) To expand the spectrum of phenotypes caused by mutations in the \textit{PCDH19} gene

We found that mutations in \textit{PCDH19} can be identified in sporadic cases of EFMR. We also determined that whilst there are a number of phenotypic characteristics in common between females with Rett syndrome and EFMR, mutations in \textit{PCDH19} are not a prominent cause of Rett syndrome. We have also demonstrated that whilst ASD is present in a number of EFMR females, we did not detect any mutations in a cohort of females (from families with multiple affected females, but not males) with ASD indicating that mutations in \textit{PCDH19} are not a prominent cause of female associated ASD.

5) To perform microarray expression profiling on samples from EFMR patients in order to explore interacting pathways and formulate hypotheses for future work.

Chapter 8 -305
The microarray expression profiling results have shed light on the genes and pathways that are affected as a result of mutations in *PCDH19* in EFMR females. These results have led us to hypothesise that the ability of certain cells to respond correctly to estrogen is disrupted in EFMR females. We also hypothesised from these results that *PCDH19* may be involved in Ephrin / Eph signalling.

Our results also open the door for a potential role for any of the additional >70 protocadherin family members in epilepsy, ID and normal neuronal function in general. *PCDH9* and *PCDH10* have been associated with autism, however no point mutations have been reported as yet in either gene (Marshall et al., 2008, Morrow et al., 2008). Putative association has also been identified between *PCDH12* and schizophrenia (Gregorio et al., 2009). Further investigation into the precise function of *PCDH19* in the brain will be required in order to explain how mutations in this gene result in epilepsy and ID.

The results of the work of Depienne et al (Depienne et al., 2009a) with the identification of a mosaic deletion of *PCDH19* in a male reinforces the validity of the newly hypothesised mechanism of cellular mosaicism / interference in the disease pathology. This proposes that disease only occurs under situations where there is a mosaic population of *PCDH19*<sup>+</sup>e and *PCDH19*<sup>-</sup>e cells, as occurs in heterozygous females and mosaic males. In line with this we have also identified deregulation of the Ephrin / Eph signalling pathways in response to heterozygous mutations in *PCDH19*. Ephrin / Eph genes have previously been associated with cellular mosaicism / interference as mutations in EFNB1 are responsible for CFNS (Twigg et al., 2004, Wieland et al., 2004). Mosaic expression of EFNB1 leads to incorrect compartments of EFNB1<sup>+</sup>e and EFNB1<sup>-</sup>e cells resulting in the formation of ectopic boundaries between compartments (Twigg et al., 2006, Wieacker and Wieland, 2005, Wieland et al., 2004).
Our results have also lead us to hypothesise a role for *PCDH19* in correct signalling in response to estrogen. The identification of a “male-like” expression profile for more than 100 genes in the EFMR females, combined with the age of onset/offset of epilepsy and the deregulation of a number of genes involved in the formation, action and degradation of estrogen as well as down-stream targets of estrogen, raises the possibility of an interaction between *PCDH19* and estrogen signalling. Sex hormones have previously been associated with epilepsy, with the menstrual cycle and menopause triggering gender specific epilepsy (Christensen et al., 2005, Vagnerova et al., 2008).

Preliminary reports have also suggested a role of estrogen signalling in ID (van Karnebeek et al., 2009). Disruption to the *ERβ* gene through a reciprocal translocation (14q22; 18q21) has been identified in three ID individuals from one family (van Karnebeek et al., 2009). The translocation breakpoint identified in this family occurs between the two alternative promoters of *ERβ* that are responsible for tissue specific regulation of expression of *ERβ*. *ERβ* expression in fibroblasts of the translocationed patients was down regulated when compared to controls. Our hypothesis of a link between *PCDH19* and estrogen signalling is particularly significant as it has the potential to facilitate the development of a therapeutic approach for the prevention of epilepsy in individuals with *PCDH19* mutations.

In conclusion we have identified the causative gene for EFMR and made significant progress toward the understanding of the underlying molecular mechanism involved in the rather unique inheritance pattern of this disorder.