

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**

Open access books available

**122,000**

International authors and editors

**135M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



# An Evolutionary Biology Approach to Understanding Neurological Disorders

Azhari Aziz<sup>1</sup>, Jasmina Ilievska<sup>1,2</sup>,  
Paul R. Fisher<sup>2</sup> and Naomi E. Bishop<sup>1,3</sup>

<sup>1</sup>*Cell Biology and Molecular Pathogenesis Laboratory,*

<sup>2</sup>*Molecular Cell Biology Laboratory,*

*Department of Microbiology,*

<sup>3</sup>*Olga Tennison Autism Research Centre,  
La Trobe University, Melbourne, Victoria,*

*Australia*

## 1. Introduction

Many common human neurological disorders, including epilepsy, Alzheimer's disease, Parkinson's disease, autism spectrum disorders, and schizophrenia show complex heritability and genetics. While studies of single-gene diseases typically provide a more straightforward opportunity to understand the underlying molecular mechanisms of disease, complex diseases are more common and inherently more difficult to study. Nonetheless, researchers have begun to make dramatic inroads into the study of complex human diseases, including many neurological disorders, in the post-human genome sequence era. This is largely due to new technologies and resources that are promoting our understanding of protein structure and function, thereby facilitating the association of disease phenotypes with genetic loci. The online Mendelian inheritance in man (OMIM) database lists those genes implicated in human disease, and this highlights progress made in this field, where around 10% of human genes have a known disease-association (Amberger et al., 2009).

In the first few sections of this paper, we highlight the differences and similarities between simple and complex human genetic disorders, and key methods to study these disorders. We emphasize the key role comparative and evolutionary biology techniques play in increasing our understanding of the pathophysiology of complex human disorders, including in the assessment of the functional traits of gene products implicated in human disease. Several human neurological disorders are used to illustrate the power of this methodology. In the last sections of this paper, the significance and implications of comparative and evolutionary biology data are highlighted using schizophrenia, and autism as specific examples. The surprising recent links between neurological disorders and cancer are discussed in the final section. We conclude that exploration of the evolutionary history of human genes, and comparison of protein structure, helps us understand how and why human neurological disorders originated, influences the choice of appropriate animal

models for human disorders, and informs our interpretation of data from model organisms, including the evaluation of novel therapeutics. We conclude that comparative and evolutionary biology, including techniques facilitating the prediction of protein function, has a major role in facilitating further understanding of human neurological disorders and in the development of therapeutic interventions.

## 2. Classification of human genetic disorders

A genetic disorder is a disease caused by an abnormality, or abnormalities, in genetic material or genome architecture. Genetic disorders are traditionally subdivided into four types: (i) single-gene disorders (often referred to as Mendelian or monogenic diseases), (ii) mitochondrial genome disorders; (iii) chromosomal disorders (where there are gross changes in chromosome structure, such as loss, duplication and/or translocation) diseases, and (iv) multigenic or complex diseases.

Much progress has been made in the last few decades in identifying the molecular cause of many rare genetic diseases (Amberger et al., 2009). Most of these are highly-penetrant traits due to single-gene mutations, and therefore follow classical Mendelian inheritance patterns (Antonarakis & Beckmann, 2006). Good progress has also been made in understanding mitochondrial genome disorders, particularly in the 30 years since the publication of the reference sequence for human mitochondrial DNA (Anderson et al., 1981; Kumar, 2008; Tuppen et al., 2010). Furthermore, a wide variety of chromosomal disorders have been characterized, where defects can be visualised microscopically (Theisen A & Shaffer LG, 2010). These three classes of genetic disorders are individually rare, although chromosomal disorders are being reported slightly more frequently in recent decades, due to factors such as increased parental age (Jones, 2008; Fonseka & Griffin, 2011), and technological advances facilitating detection of smaller deletions and duplications (Berg et al., 2010; Shaffer et al., 2007; Slavotinek, 2008). Despite these broad categories frequently being used to classify human genetic disorders, it must be borne in mind that the phenotypic expression of genetic mutations varies, and this is discussed next, before we focus our attention on complex genetic disorders.

### 2.1 The complexity of single-gene disorders

It is now well-known, even for well-characterized Mendelian genetic disorders, that individuals with a specific mutation can display phenotypic differences. This includes variation in the age of onset, in the severity of disease symptoms, and/or in phenotypic characteristics. Indeed, phenotypic pleiotropy is the rule, rather than the exception, even for single-gene disorders (Nadeau, 2001). The variation in individual phenotype can be affected by environmental factors, allelic variation and/or 'modifier' genes. Modifier genes can affect transcription and levels of gene expression directly, or affect phenotype at the cellular, tissue, or organism-level (Nadeau, 2001). While increasing numbers of human modifier genes are being identified, most of the progress in understanding genetic modifiers is dependent on model organisms, such as mice, where gene targeting experiments can be carried out using inbred strains (Nadeau, 2001, 2003). One particularly relevant class of modifiers are referred to as protective alleles, as their presence prevents disease from occurring (Nadeau, 2001, 2003). These findings provide insights relevant to the development

of novel therapeutics, as therapeutics could be based on mimicking and/or possibly enhancing the effects of these protective alleles. This class of gene also means that the same genetic mutation can lead to a different disease phenotype depending on the genetic background (Lobo, 2008). Therefore, while genetic disorders are frequently classified as monogenic or complex (see below), the distinction between the two types is becoming increasingly blurred.

Other emerging aspects of monogenic disorders overlapping with those of complex disorders, are those due to the multi-functional nature of many genes. This multi-functionality can make it difficult to predict phenotype from genotype. This is illustrated by metabolic genes, where some gene products are referred to as 'moonlighting proteins' due to the multiple phenotypic effects of mutations (Jeffery, 2009; Sriram et al., 2005). Such multi-functionality also contributes to the observation that distinct phenotypes can be associated with different mutations in the same gene. For example, the *LMNA* gene encodes two proteins and is linked to five diseases (Vigouroux & Bonne, 2002), while mutation of the *ERCC2* gene may cause xeroderma pigmentosum (XP), Cockayne syndrome with XP, or trichothiodystrophy, three phenotypically different disorders (Lehmann et al., 2001). In other cases, different mutations in a single gene can cause different diseases via mechanistically different pathways. For example, the *FMR1* gene is considered 'a gene with three faces' (Oostra & Willemsen, 2009). Mutation of *FMR1* is best characterised as the cause of fragile X syndrome mental retardation, which is inherited in an X-linked dominant pattern, and is due to a lack of *FMR1* mRNA and protein expression. However, mutations leading to high levels of *FMR1* mRNA are linked to tremor/ataxia syndrome via mRNA 'toxicity', while the gene is also linked to premature ovarian insufficiency via a third uncharacterized molecular pathway, possibly affecting the production of *FMR1* mRNA isoforms (Oostra & Willemsen, 2009; Tassone et al., 2011).

Another factor complicating the phenotype of Mendelian disorders is the finding that heterozygotes for some recessively inherited Mendelian disorders, whom show no symptoms of the homozygotic phenotype, are at risk of an apparently unrelated disorder (Sidransky, 2006; Sriram et al., 2005). For example, patients who are heterozygous for the gene deficient in Gaucher disease are at an increased risk of neurodegenerative synucleinopathies, such as Parkinson disease (Sidransky, 2006). An additional complication arises in patients who show clinical symptoms consistent with a single-gene defect in a metabolic pathway, but do not have a complete deficiency in any one enzyme, but rather have multiple partial defects. This phenomenon is referred to as synergistic heterozygosity (Vockley et al., 2000).

Finally, while some genetic disorders are largely polygenic and complex in nature, a subset is inherited in a classical Mendelian manner (see Fig. 1). For example, with Alzheimer disease (ALZ) and Parkinson disease (PKD), a subset of the diseases (prefixed by the term 'familial') are inherited in a Mendelian manner. With PKD, around 5% of cases are due to mutations in one of several specific genes with either autosomal dominant or recessive inheritance patterns (Gasser, 2009; Lesage & Brice, 2009; Shulman et al., 2011), but PKD-associated genes with a more modest penetrance are now beginning to be identified (International Parkinson Disease Genomics Consortium, 2011; Liu et al., 2011; Shulman et al., 2011). With ALZ, around 0.1% of cases are inherited in an autosomal dominant manner, while one *APOE* allele, present in 2% of Caucasian populations, has recently been

reclassified from 'risk gene' status to being considered moderately penetrant with semi-dominant inheritance (Blennow et al. 2006; Genin et al., 2011). Nonetheless, ALZ in most patients is influenced by a combination of multiple genetic risk factors and protective alleles (Sherva & Farrer, 2011; Waring & Rosenberg, 2008).

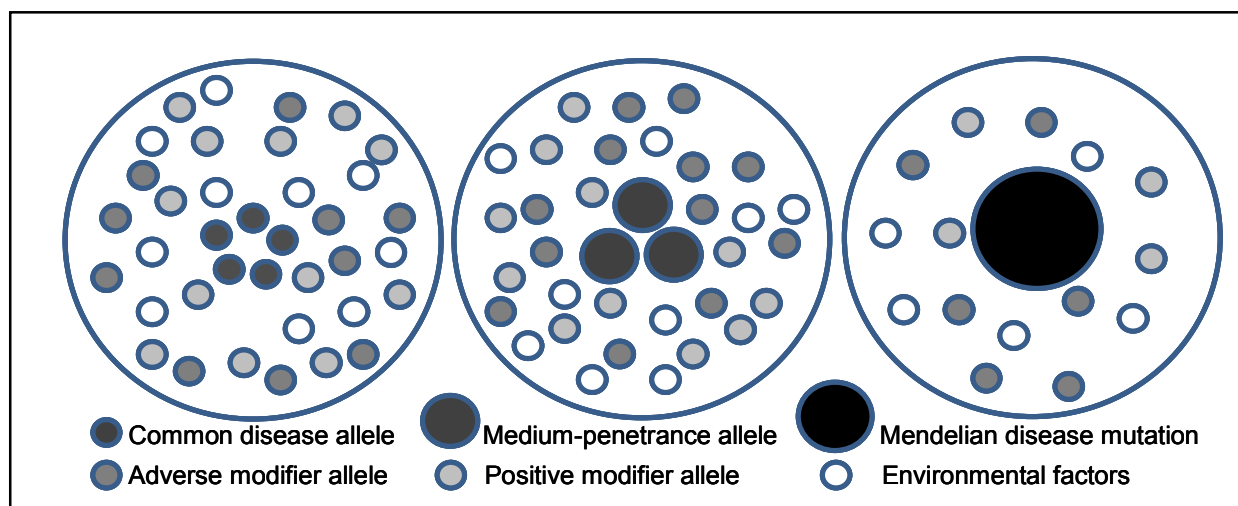


Fig. 1. Multiple genetic and environmental variants in different combinations affect phenotypes. Common variants (left) or rare mutations (right), or alleles of medium penetrance (middle) can all cause human genetic disorders. Penetrance of a disease allele can be affected by so-called modifier alleles.

Therefore, Mendelian disorders have more in common with multi-factorial diseases than originally thought, and both are affected by genetic background and environmental conditions. Furthermore, the rare Mendelian forms of common complex disorders are providing key insights about the pathogenesis of many complex diseases by highlighting cellular pathways perturbed in the disease state (discussed further in Section 4.4) and this is leading to testable hypotheses about disease etiology (Peltonen et al., 2006). Complex genetic disorders are discussed next, emphasizing the importance of evolutionary and comparative biology, while the relevance of these areas of research to multifunctional genes will be discussed further in Section 4.

## 2.2 Complex genetic disorders

While most Mendelian disorders are rare, there are over 7000 such disorders, and so they collectively affect hundreds of millions of people worldwide (Amberger et al., 2009). By contrast, most of the common disorders of children and adults are complex diseases, and a single highly-penetrant gene is not causative of the disease phenotype (see Fig. 1). Indeed, the causes of such disorders are usually heterogeneous, and a combination of effects from more than one gene, combined with non-genetic factors (environment), play a role in disease development (Davey Smith et al., 2005). Such disorders in children include mental retardation, autism spectrum disorders, attention deficit/hyperactivity disorder, and cancer. In adults, common complex disorders include schizophrenia, bipolar disorder, diabetes, coronary heart disease, hypertension, obesity, and cancer. The complex, multigenic, nature of these diseases has made them inherently more difficult to study. However, in the next



section of this paper, we will discuss the key methods used to determine the genetic underpinnings of common complex disorders. Understanding the etiology of these multifactorial diseases is essential for the development of effective means of treatment and/or prevention.

### 3. Studying complex human genetic disorders

Complex disorders often cluster in families without clearly demonstrating Mendelian inheritance patterns. This makes it difficult to determine the genetic versus non-genetic contribution to the disease phenotype, and to calculate the heritable component of the disorder. Below we will discuss methods used to establish the heritability of human complex disorders, generation of the genetic variation that underpins these disorders, and discuss how to establish which genes are responsible for complex human disease.

#### 3.1 Heritability of complex human diseases

Heritability is usually defined as the proportion of total phenotypic variation that can be attributed to genetic variability (Lee et al., 2011; Visscher et al., 2008). While the interaction of environment on phenotype makes heritability difficult to measure accurately in some cases (Ober & Vercelli, 2011), methods of obtaining unbiased estimates of heritability from various types of pedigree data are well established for both continuous phenotypes and complex human disorders (Lee et al., 2011; Visscher et al., 2008). Furthermore, animal models are invaluable in dissecting aspects of genetic and environmental interactions that are more difficult to assess in human studies (Complex Trait Consortium, 2004) and is discussed further in Section 4.5.

For many human diseases, recent data suggest that the heritable component of many has previously been underestimated (Lee et al., 2011). This has been due to limitations of the methodology employed, as well as other factors, such as evidence demonstrating monozygotic twins are less genetically similar than once thought (Zwijnenburg et al., 2010). A further example is that of PKD, which was long considered a non-hereditary disorder (Shulman et al., 2011; Westerlund et al., 2010). Despite extensive efforts to find environmental risk factors for the disease, genetic variants now stand out as the major causative factor (Shulman et al., 2011; Westerlund et al., 2010; Wirdefeldt et al., 2011). This shift of focus away from environmental toxins, towards genetic contributions, is now leading to rapid progress in understanding PKD and in guiding the development of the next generation of therapeutics (Shulman et al., 2011).

While it is clear that genetics underpins the pathophysiology of complex human disorders, the genetic alleles contributing to the disease phenotype are not always inherited. *De novo* mutations are increasingly being implicated in human disease and, by definition, these mutations are not present in the biological parents of the affected individual. Nonetheless, depending on the severity of the phenotype, and on any effects on fitness, these novel mutations may be transmitted to subsequent generations. Indeed, the rise of techniques such as intracytoplasmic sperm injection (ICSI), can facilitate transmission of *de novo* mutations even if they lead to infertility (Jiang et al., 1999). There is a wide variety in *de novo* human germline mutations, and these can include duplications or deletions of various size, as well as alterations in the number of chromosomes (Arnheim & Calabrese, 2009). The frequency of

*de novo* mutations in germ-line cells increases with parental age, and can also be caused by environmental factors such as radiation exposure (Sasaki, 2006), genotoxic chemicals (Phillips & Arlt, 2009), or congenital viral infections (Ansari & Mason, 1977; Fortunato & Spector, 2003; Nusbacher et al., 1967; Vijaya-Lakshmi et al., 1999).

While *de novo* mutations typically occur in gametes, and are often defined as such, new mutations can also occur in the precursors of germ cells, leading to germline mosaicism (Arnheim & Calabrese, 2009), or occur post-fertilization, during embryonic/foetal development and somatic cells (Lupski, 2010). Indeed, new mutations can develop at any time, with cancer being the best-known example of a genetic disorder caused by somatic cell mutagenesis, while Proteus syndrome is a recently-identified disease linked to somatic mosaicism (Lindhurst et al., 2011). While mutations in the genome of somatic cells cannot be passed on to future generations, they may have detrimental effects. While *de novo* mutations are best-studied in cancer (see Section 6.2.3), a contributing role of somatic *de novo* mutations, such as those occurring during brain development, to neurological disorders (see Section 6.1.2) has not been explored.

Mutation frequencies also vary widely across the genome, and often concentrate at certain positions or 'hotspots' (see also Sections 3.2, 6.1.2 & 6.2.3), which have structural and functional features affecting mutagenesis (Ananda et al., 2011; Arnheim & Calabrese, 2009; Carvalho et al., 2010; Rogozin & Pavlov, 2003). For example, CpG context elevates the mutation rate by an order of magnitude (Schmidt et al., 2008), non-B DNA structures induced by palindromic AT-rich repeats facilitate recurrent translocations on chromosomes 11 and 22 at positions 11q23 and 22q11 (Kurahashi et al., 2006), while interspersed repetitive elements such as *Alu*, LINE, long-terminal repeats, and simple tandem repeats are frequently observed at breakpoints in the 9q34.3 subtelomere region (Yatsenko et al., 2009). However, at any given point, multiple mechanisms are acting, making prediction of mutational site and frequency difficult (Arnheim & Calabrese, 2009).

Despite *de novo* mutations most typically being deleterious, rather than neutral or advantageous, their very existence is evidence for ongoing adaptive evolution. Therefore, genetic disorders can be considered 'side-effects' or manifestations of the fundamental mechanisms that provide the genetic variation necessary for evolution to occur. The interaction between the evolutionary past of the human genome and human genetic disease is discussed next.

### 3.2 Human evolution and genetic disorders

Duplicated regions of DNA play a key role in the evolution of novel gene functions (Conant & Wolfe, 2008; Lynch & Conery, 2000; Ohno, 1970), but are also a source of genetic instability, leading to mutations implicated in both rare and common human genetic disorders (Marques-Bonet & Eichler, 2009). It is therefore relevant to explore the origins of human duplicated sequences. Current evidence indicates that many segmental duplications occurred the hominid lineage and, more specifically, in the common ancestor of African great apes (chimpanzee, gorilla, humans) after divergence from the Ponginae or Asian great ape (orangutan) lineage (Bailey & Eichler, 2006; Carvalho et al., 2010; Koszul & Fischer, 2009; Marques-Bonet & Eichler, 2009). A considerable portion of duplicated human sequences have also been found to correspond to expanded gene families, some of which show

signatures of positive selection (Marques-Bonet & Eichler, 2009). Duplications specific to the *Homo sapiens* lineage have also been detected, and include duplications in gene families implicated in neurotransmission, and these may play a role in higher-order brain function in humans (Han et al., 2009).

Different classes of repetitive DNA sequence have been identified (Bao & Eddy, 2002). Some, such as LINES (e.g. L1 family) or SINES (e.g. *Alu* family), are long and short retrotransposable elements, found interspersed throughout the genome. Others are concentrated in certain regions, such as centromeres and telomere-adjacent sequences. These latter regions are also sites of increased genomic instability, which are associated with disease-causing chromosomal breakpoints (Stankiewicz & Lupski, 2002). *Alu* elements have propagated to more than one million copies in primate genomes, and likewise contribute to human genomic diversity (Batzer & Deininger, 2002). Indeed, one in 50 individuals will carry a *de novo* L1 insertion, and one in 20 individuals a *de novo* *Alu* insertion (Collier & Largaespada, 2007). The active nature of many human retrotransposons is therefore linked to disease-causing somatic and germline mutations (Collier & Largaespada, 2007; Wallace et al., 1991; Oldridge et al., 1999; Claverie-Martin et al., 2003). Repeats may also contribute to DNA secondary structures that are more prone to breakage (Yatsenko et al., 2009). One novel aspect of our increased understanding of the role of repetitive DNA sequence in *de novo* mutations, and our ability to detect such sequences, is that this information can now be used to predict rearrangements that will contribute to genomic disorders (Carvalho et al., 2010; Ou et al., 2011; Sharp et al., 2006).

Therefore, while duplicated sequences in primate genomes predispose apes and humans to extensive genetic diversity and biological innovation, the downside is that many *de novo* genomic changes are mediated by recombination events between these duplications. This characteristic of hominids, and *Homo sapiens* in particular, makes humans particularly susceptible to genomic rearrangements. These rearrangements, in turn, then play a major role in human genetic disease pathogenesis (Inoue & Lupski, 2002; Marques-Bonet & Eichler, 2009). The evolutionary history of some specific genomic rearrangements is discussed next.

### 3.2.1 Evolutionary history of specific human disease mutations

Using an evolutionary perspective, we can use comparative genomic analyses to calculate the age of appearance of segmental duplications mediating specific disease-causing mutations. Such analyses have revealed that the segmental duplication flanking the Charcot-Marie-Tooth disease region on chromosome 17 (at position 17p12) has an origin in the hominoid ancestor after the divergence of chimpanzees and humans, those flanking the DiGeorge syndrome region on chromosome 22 (22q11.2) expanded after the divergence of hominoids from Old World monkeys, the duplications flanking the Angelman/Prader-Willi region on chromosome 15 (15q11-q13) began to expand before the divergence of the Old World monkeys, while the Smith-Magenis syndrome segmental duplications (17p11.2) date back to after the divergence of New World monkeys (Marques-Bonet & Eichler, 2009). These, and other similar data, have demonstrated that the predisposing genomic features contributing to many genomic disorders have emerged within the last 25 million years (Marques-Bonet & Eichler, 2009).



Using similar methodology, the evolutionary history of human genetic disorders where *Alu* elements are implicated, have also been dated. For example, *Alu* elements mediating lipoprotein lipase deficiency (*LPL* gene) are found in human, ape, and monkey genomes; those *Alu*-elements implicated in Lesch-Nyhan syndrome mutations (*HPRT* gene) are common to human, chimpanzee, and gorilla; while those implicated in ApoB deficiency are restricted to human and great ape genomes (Martinez et al., 2001). Of note, while *Alu* elements acted as 'selfish DNA' when they first inserted into primate genomes, many have subsequently gained regulatory function, a process known as exaptation (Hasler & Strub, 2006). Therefore, despite contributing to the pathogenesis of human genetic disorders, *Alu* elements are thought to have played a role in the divergence of primates, and to have contributed to the regulatory and developmental complexity in primate lineages (Hasler & Strub, 2006). Therefore, *Alu*-depending human genetic diseases also date to primate lineages. Methods used to predict the different phylogenetic ages of genetic disorders instigated by DNA repeats and duplications can therefore being used to both predict and explain the different susceptibility of various primate species to genetic diseases (Martinez et al., 2001; Marques-Bonet & Eichler, 2009).

### 3.3 Identifying and characterizing genetic determinants of human disease

In the last few decades there has been rapid progress in human disease gene identification, due to recombinant DNA technologies, genome sequencing and analysis methods (Strachan & Read, 2010). With the vast range of resources now available, identification of novel disease genes is currently occurring on a weekly, if not daily, basis. There is no standard procedure for gene identification, however, and identifying the genes responsible for human disease requires information about both gene position and biological function. Functional data is proving a bottleneck for progress in understanding complex diseases in particular and, as outlined below, our understanding of evolutionary biology is of great benefit to studies aimed at identifying and characterizing the genetic determinants of human disease phenotypes.

Genetic linkage- and association-based analyses have been very successful in identifying rare genetic variants with highly penetrant effects, such as those causing Mendelian diseases, and have also been used to investigate the genetics of complex disorders (Altshuler et al., 2008; Jordi, 2000; Ku et al., 2011). More recently, techniques such as genotyping arrays and next-generation DNA sequencing are facilitating the identification of mutations causative of the many as-yet-uncharacterised Mendelian disorders, and of the genetic variation contributing to complex genetic disorders (Kingsley, 2011; Kuhlenbäumer et al., 2011; Roberts et al., 2010).

Two key issues have been emerging from these recent genetic studies. The first is the finding that genome-wide association studies (GWAS) have not been particularly effective in identifying complex genetic disorder risk genes, and the numbers and impact of identified genetic risk factors has been 'disappointing' (Davey Smith et al., 2005; Manolio et al., 2010; Gandhi & Wood, 2010). If sample numbers are sufficiently great, however, GWAS may be better placed to unambiguously identify risk or protective loci for complex diseases (Sullivan, 2001; Wray et al., 2008). Secondly, there are great ongoing difficulties in differentiating disease-causing mutations from rare benign variants (Kuhlenbäumer et al., 2011). These difficulties not only highlight the need for greater 'power' in GWAS, but also

the need for follow-through on genetic findings and the application of many aspects of what is referred to as 'integrative genomics' (Giallourakis et al., 2005). Below, we outline some of the varied ways in which our understanding of evolutionary biology is of vital importance to leveraging information obtained from genetic studies. The resulting information can provide key insights into the biological function of uncharacterised genes, can be used to predict candidate disease genes, to predict detrimental mutations, and can provide valuable information about biological pathways altered in different disease states, which may lead to novel therapeutics.

### 3.3.1 Validating candidate disease genes

GWAS and linkage studies generate large sets of potential disease genes. However, it remains difficult to identify the most likely disease-related genes. Various computational methods for disease gene identification have been described (Oti et al., 2011; Tiffin et al., 2006), and many of these have as their basis data from the field of evolutionary biology. Many software tools apply some of this type of information to genetic datasets, and can be used to determine which genes are the most likely to be involved in the disease in question. The Gene Prioritization Portal website provides an up-to-date summary of web-based candidate disease gene prioritization and prediction tools (Tranchevent et al., 2010). However, most gene prediction tools were designed to study Mendelian disorders, and not for the analysis of complex genetic disorders. The exceptions are the web-based tool, CANDID, specifically designed to prioritize genes implicated in complex human genetic traits (Hutz et al., 2008), and CAESAR, which is not web-based (Gaulton et al., 2007). Tools applicable to complex disorders can be expected to expand over the coming years, due to the large amount of data that will be generated using new genome analysis methods.

A major drawback for the prediction and prioritization of disease genes is that most candidate gene identification tools are reliant on how well-characterized each human gene is, and whether its molecular and cellular function are known. Bear in mind, then, that: (i) over 98% of all gene ontology (GO) annotations are computationally inferred, have not been curated, and are considered by the GO consortium to be potentially unreliable (du Plessis L et al., 2011); (ii) errors in the sequence databases affect at least 1 in 6 sequences (Lagerstrom et al., 2006; Slater & Bishop, 2006; Haitina et al., 2009; Bishop, unpublished data), and this may affect the output from tools such as PROSPECTR (Adie et al., 2005) that use sequence features to rank genes in order of their likelihood of involvement in disease; (iii) while high-throughput protein-protein interaction detection studies have great potential for increasing our understanding of complex genetic disorders, the paucity and unreliability of available data currently limit the power of these approaches (Chen et al., 2008; Chua & Wong, 2008; Kuchaiev et al., 2009); and (iv) many domains in proteins are of unknown function. Tools, such as SUSPECTS (Adie et al., 2006), which rely on detecting shared domains, annotation, and patterns of expression, are clearly limited by the incompleteness and inaccuracy of these data. Therefore the gaps in our knowledge about gene product function are greatly hampering our understanding of complex genetic disease. Computational techniques, many anchored by evolutionary understanding, are helping direct and accelerate our understanding of the biological function of human genes, and the effect of human gene mutation and variation. These techniques are discussed next.

### 3.3.2 Gene age-based candidate gene prioritization

The first systematic study comparing the sequence characteristics of human disease genes (listed in OMIM) with genes not known to be involved in disease, found a subset of sequence-based features to be significantly different between the two sets of genes (Adie et al., 2005). These researchers created a web-based tool called PROSPECTR based on those features, which enriches lists for disease genes (Adie et al., 2005). Relevant DNA characteristics, more common in disease genes, include larger gene length and the presence of a mouse homologue (Adie et al., 2005). However, one limitation of this, and other, disease gene prediction algorithms is that they are developed on the basis of known disease genes, and many disease genes remain unidentified (reviewed by Ropers, 2007).

While PROSPECTR examines whether murid orthologues of a gene exist (Adie et al., 2005), the evolutionary history of human 'disease genes' has subsequently been explored in greater depth. A comprehensive study on the evolutionary 'age' of genes mutated in human diseases compared to those not implicated, revealed that human disease genes are more likely to be 'old' genes (Domazet-Lošo & Tautz, 2008). These so-called 'old genes' are classified on the basis that they have orthologues in urochordates and/or more anciently-diverging lineages, and contrast with 'recent genes' where orthologues are restricted to chordate lineages. The over-representation of human disease genes among old genes is even more pronounced among those genes with tissue-specific expression profiles (Nagaraj et al., 2010). The evolutionary history of Mendelian disease genes, compared to genes implicated in complex disease, has also been examined, and more recently-evolving genes also divided into 'middle-aged' and 'young' gene categories (Cai et al., 2009). This study found that Mendelian disease genes tend to be older than non-disease genes, while complex disease genes are typically middle-aged (Cai et al., 2009). Therefore, despite not being evolutionarily ancient, most complex disease genes originated during the emergence of vertebrates, and are not human- or primate-specific.

## 4. Determining human disease gene function

The function of many human gene products is unknown or very poorly understood, and this greatly hampers progress in all studies on complex human genetic disorders. Below we discuss key methods used to predict and understand gene function, where a thorough understanding of evolutionary biology is imperative.

### 4.1 Sequence-based approaches to predicting gene function

The main method for predicting the function of a gene product in the absence of experimental data is termed 'homology-based transfer' (Friedberg, 2006; Sleator & Walsh, 2010). This approach is based on the detection of significant amino acid sequence similarity to a protein(s) of known function using programmes such as BLAST (Altschul et al., 1997). As sequence similarity suggests a common evolutionary origin, the function of the known protein is then transferred to the query protein. This method is not foolproof, however, and exceptions have been described at both ends of the similarity scale (reviewed by Sleator & Walsh, 2010). Understanding which residues are essential for protein function can be important in evaluating the relevance of similarities detected between proteins, such as those in conserved motifs. Furthermore, there are many proteins where homology-based

prediction cannot be used. Therefore, more recently, non-homology based computational approaches have begun to emerge (reviewed by Sleator & Walsh, 2010). These methods are based on a combination of sequence, structural prediction methods, evolutionary history, biochemical properties, and genetic and genomic knowledge.

#### 4.2 Predicting the effect of mutation on gene function

A major problem in the search for disease-causing mutations is the fact that some of them are difficult to recognize. Tools to evaluate the functional impact of mis-sense mutations is limited by the few solved protein tertiary structures, and on the limitations of software to predict effects of mutations on protein domain function and/or protein conformation, much of which incorporates evolutionary conservation data (reviewed by Ropers, 2007). Complicating matters further, even silent mutations have been found to be pathogenically relevant (Kimchi-Sarfaty et al., 2006; Pagani et al., 2005). Furthermore, non-coding mutations may not be examined or detected and, even if they are, it is currently even more challenging to predict whether these have functional effects. For example, intronic changes may alter the splicing pattern (Richards et al., 2007; Lenski et al., 2007) and promoter mutations may affect gene expression levels (Almeida et al., 2006; Borck et al., 2006). Unlike the situation found with Mendelian disease loci, an increasing proportion of the loci being associated with complex disorders are being found outside protein-coding regions of the genome (Pomerantz et al., 2010). The thousand genomes project will sequence the complete genome sequence of more than 1,000 humans, and will provide valuable information about variants normally present in the human population (Marth et al., 2011).

Genetic mutations implicated in human disease are often mis-sense/nonsense mutations, or involve small duplications/deletions in coding regions. However, recent sequencing of multiple human exomes (coding regions of the genome) suggest that these types of mutation are actually quite common and such changes are frequently benign (Ng et al., 2008; Ng et al., 2009). One way of distinguishing disease-causing mutations, at least in Mendelian disease genes, is because they occur more frequently in evolutionarily well-conserved amino acid residues, than in non-conserved ones, and these changes are expected to have a more severe impact on the function of the resulting protein. By contrast, the distribution of mutations, such as non-synonymous single nucleotide polymorphisms, (nsSNPs or cSNPs) contributing to complex human diseases, are often difficult to distinguish from the distribution for "normal" human variation (Thomas & Kejariwal, 2004). These results indicate that individual SNPs implicated in complex genetic disorders will have more subtle effects on function in isolation. This observation further suggests a disease architecture involving the concerted contribution of multiple genetic loci, each with a small individual effect. Indeed, this explanation is suggested to be the basis of a large proportion of common complex disorders, and is known as the common-variant common-disease (CVCD) model of complex genetic disease (Grady et al., 2003; Visscher et al., 2011). However, as discussed in Section 2.1 above, rare genetic changes can also be causative of a subset of common disorders, and is described as the RVCD (rare-variant common-disease) model. However, there is no cut-off point between these two models (see Fig. 1), and there is a broad variety in both the frequency of disease gene variation and in the penetrance of a given genetic change, which together combine to cause a given disease phenotype (Grady et al., 2003; Visscher et al., 2011).



Computational prediction of the effects of genetic variation utilizes evolutionary conservation of the resultant gene product, in combination with predictions of the changes to the physicochemical properties (Mooney, 2005; Ng & Henikoff, 2006; Ng et al., 2008; Tarpey et al., 2009). Computer-based tools are also used to predict conserved domains and motifs, and can be used to determine whether nsSNPs or other genetic changes are likely to contribute affect protein function. Examples of such databases include PROSITE (Hulo et al., 2008), BLOCKS (Henikoff et al., 2000), and PRINTS (Attwood et al., 2003). Variation in the sequence of orthologues with conserved function can also be used to indicate the amino acid variation possible in a domain, or motif, which is still predicted to maintain some degree of functionality.

Nonetheless, while sequence-based approaches provide a good basis for predicting the function of genes of unknown function, in many cases there may be little or no sequence similarity between an unknown gene product and any characterized gene product. Fortunately, due to evolutionary constraints, there is often still significant structural similarity between an uncharacterized protein and a characterized protein, and this can be a useful indicator of function (Shatsky et al., 2008; Todd et al., 2001; Watson et al., 2005). Therefore, recent developments are aimed at combining both sequence and structural information to increase the likelihood of a functional prediction (Laskowski et al., 2005; Pierri et al., 2010; Skolnick & Brylinski, 2009). However, there is much room for improvement to the currently available approaches for the prediction of protein function, as only ~1% of proteins on the UniProt database have experimentally-supported function, ~65% have some functional annotation, and over one-third are uncharacterized or have no predicted function (Barrell et al., 2009; Erdin et al., 2011; Goldsmith-Fischman & Honig, 2003; Laskowski et al., 2003; Magrane & Consortium, 2011). This may be particularly important for disease genes, as many intrinsically-disordered proteins, lacking stable secondary and tertiary structures, are being found associated with many complex human diseases, including cancer, diabetes, neurodegenerative diseases, and cardiovascular disease (Midic et al., 2009; Uversky, 2009; Uversky et al., 2008, 2009; Wang et al., 2011).

#### **4.3 Evolutionary pedigree and co-expression to leverage functional prediction**

Expression data can also provide information relevant to genetic disorders. For example, it is useful to know whether candidate genes are expressed in the tissue affected by the disease. Furthermore, genes involved in similar cellular processes are also more likely to be co-transcribed, hence the 'guilt-by-association' algorithm (Walker et al., 1999; Oliver, 2000). Not only can 'power' be added to the analysis of co-expression datasets, a better understanding of the evolutionary history of human genes may lead to novel ways to interpret sequence data and predict protein function (Eng et al., 2009; Thornton & DeSalle, 2000). For example, proteins with a similar evolutionary pedigree, and emerging at the same 'point' in the evolutionary history of the Eukaryota, are assumed to have evolved in parallel. This, in turn, indicates they are more likely to have a common function (Eisenberg et al., 2000). The 'guilt-by-association' principle greatly informs our understanding of protein-protein interaction networks, as it is known that most cellular functions are carried out by networks of interacting proteins (Qiu & Noble, 2008). Understanding protein-protein interaction networks is also informing computational approaches aimed at understanding the role of pathways affected in complex genetic diseases, and this is discussed next.



#### 4.4 Network studies to leverage functional prediction

Network-based models incorporating protein-protein interaction (PPI) data are a relatively new way for studying disease-related genes. Nonetheless, this approach has already been proved to be effective for the identification of complex disease genes, including those involved in colon cancer (Nibbe et al., 2009). Therefore, analyzing the functional networks of human genes is providing a key framework for prioritizing candidate disease genes. Such analyses may also lead to the identification of key cellular pathways common to complex diseases that may be amenable to therapeutic intervention.

Based on large- and small-scale PPI studies, preliminary protein-protein interaction networks are being created (e.g. Fig. 2). These can be accessed and viewed using a number of web tools, such as STRING and BioGRID (Han, 2008). Proteins sharing a particular functional category cluster in the same location of PPI networks, and are referred to as functional modules, and placement of proteins in PPI networks can be used to inform protein function prediction studies (Dziembowski & Seraphin, 2004; Yook et al., 2004; Makino & Gojobori, 2006). Understanding networks of PPIs can be used to predict additional genes that, when mutated, may cause the same disease as that associated with mutations in interacting partners (McGary et al., 2010), and also explains why so many different mutated genes can cause the same or similar complex disease (Bill & Geschwind, 2009; Bourgeron, 2009; Crespi et al., 2010; Gilman et al., 2011; Guilmatre et al., 2009).

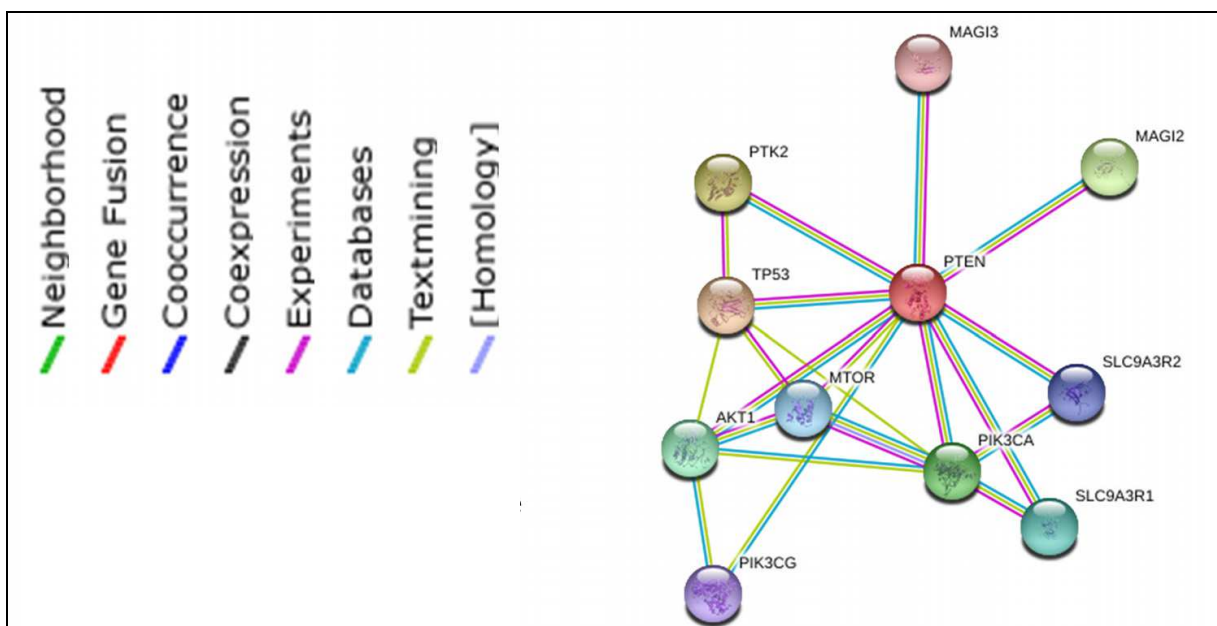


Fig. 2. Protein-protein interaction network example. Multiple lines of evidence from the STRING database (<http://string-db.org/>) demonstrate that the PTEN cancer gene product (red) is a hub-protein.

Relating to the network-based nature of many gene products is the concept that some proteins interact with multiple other proteins, referred to as 'hub' genes (Fig. 2). Multi-functional genes may impact on numerous pathways, while different mutations may cause different diseases (Gillis & Pavlidis, 2011). The encoded hub proteins have multi-functional cellular roles, and are typically annotated with multiple GO categories (Gillis & Pavlidis,

2011). The multi-functional nature of many proteins also contributes to phenotypic pleiotropy, where mutation(s) in a single gene can affect multiple phenotypic traits. Genes leading to pleiotropic phenotypes tend to be more evolutionarily conserved and are more likely to have essential functions (Eisenberg & Levanon, 2003; Feldman et al., 2008; Fraser et al., 2002; Gandhi et al., 2006; Goh et al., 2007; Jeong et al., 2001; Saeed & Deane, 2006). Overall, human disease-causing genes have an intermediate essentiality, being less than that of housekeeping genes, but greater than that of non-disease genes (Feldman I et al., 2008; Goh et al., 2007; Liao & Zhang, 2008; Tu et al., 2006). These findings are, to some extent, commonsense, as mutation of a housekeeping gene would be expected to lead to embryonic lethality.

As gene products that are peripheral in protein-protein interacting networks are known to have a higher evolutionary rate than hub proteins, these data also suggest human disease genes differ in rates of evolutionary change compared to non-disease genes. However, current results on this topic are inconsistent, with some studies indicating disease genes are evolving more slowly than non-disease genes (Blekhman et al., 2008; Tu et al., 2006), and other studies suggesting disease genes are evolving faster than non-disease genes (Huang et al., 2004; Smith & Eyre-Walker, 2003). There are a number of confounding factors contributing to these inconsistencies in establishing the rate of evolution of human disease genes. Comparing proteins based on PPI network interactions, genes encoding proteins that are part of 'modules' tend to be more conserved, evolutionarily old, and ubiquitously expressed. By contrast, genes encoding proteins outside modules are less well-conserved, evolutionarily younger, and enriched with at least some degree of tissue-specific expression (Dezso et al., 2008). Therefore, there appear to be different classes of disease genes, with different evolutionary patterns and different PPI patterns, and this may relate to the evolutionary 'age' of the disease gene (Nagaraj et al., 2010). Characterization the differing traits or classes of disease genes may contribute to our understanding of the different functional 'style' of disease genes. It also appears that Mendelian disease genes and complex disease genes have different evolutionary profiles (see Section 3.2.3) and, of course, these studies are limited by our current knowledge of disease genes, which remains incomplete.

Networks of PPIs are also informing efforts aimed at understanding whether human evolution is heading towards, or away from, susceptibility to a particular disease. One crucial aspect of this discussion, is the presence of alleles that increase the risk of one disorder, while simultaneously decreasing the risk for another. For example, multiple sclerosis and rheumatoid arthritis, or ankylosing spondylitis and multiple sclerosis, are negatively correlated (Sirota et al., 2009). By contrast, other alleles can increase the susceptibility to more than one complex disorder. This leads to multiple sclerosis and autoimmune thyroid disease, or type 1 diabetes and coeliac disease, commonly co-occurring (reviewed by Sirota et al., 2009). While these studies indicate the complexities of disease susceptibility, tools are being developed to leverage PPI data to predict such disease interactions (Chen et al., 2008; Chua & Wong L, 2008).

Overall PPI networks are crucial for the development of testable hypotheses regarding the underlying pathogenic mechanisms of many complex genetic disorders. However, PPI networks are incomplete and include both false-positive and false-negative interactions (Chen et al., 2008; Gandhi et al., 2006; Kuchaiev et al., 2009). Development of a reliable and complete human PPI network will provide an invaluable framework to study the

contribution of multiple genes to complex genetic disorders, and will require both computational and experimental data to achieve accuracy and completeness.

#### 4.5 Role of animal models

Animal research, including animal models of disease, has been responsible, at least in part, for every major medical advance made during the last century (Müller & Grossniklaus, 2010). The reason why model organisms are able to contribute so effectively to our understanding of human diseases lies in the high degree of molecular conservation found between metazoan species, and in the conserved nature of protein-protein, and other, networks (Gandhi et al., 2006). Indeed, even bacteria, plants, protists, and fungi are being exploited to explore differing aspects of biology relevant to human disease (Annesley & Fisher, 2009; Ilievska et al., 2011; McGary et al., 2010; Spradling et al., 2006).

As discussed above, a huge number of susceptibility alleles for a range of complex human genetic disorders have now been identified, but the function of many of these genes is poorly understood. Therefore, while risk-associated loci are being successfully identified (Easton et al., 2007; Hindorff et al., 2009; Jia et al., 2009), these findings are rarely followed up and the contribution of the allele to the molecular basis of disease rarely evaluated (McCarthy & Hirschhorn, 2008). That this failure is leading to a bottleneck in our understanding of complex disorders was highlighted by a recent *Nature Genetics* editorial, which suggested that significant investment in functional characterization of risk loci is needed (Axton, 2010). There are a number of ways to investigate the molecular and cellular function of a gene and its alleles, and these include *in vitro* studies, cell culture systems, and the use of whole animal models. These tools can be used to test hypotheses gained from thorough *in silico* studies. In this section we will discuss the role of animal models, as this is of greatest relevance to the topic of this paper. Although *in vitro* and cell culture studies can be of great benefit, ultimately good animal models provide the best biological models for complex disease.

High-throughput phenotypic screens of RNAi knockdowns in *Caenorhabditis elegans* and *Drosophila melanogaster* often provide the first inkling of the biological function of an uncharacterized human gene (Buckingham et al., 2004). Efforts are also underway to systematically knockout all the genes in the mouse genome to facilitate phenotypic and functional screening (Guan et al., 2010). Animal models also provide a vital system amenable for dissection of the contributions of genetic, environmental and developmental components to the etiology of complex human disorders, and for evaluation of novel therapeutics, which can be achieved in no other way (Complex Trait Consortium, 2004; Iwata et al., 2010).

Many recent advances relevant to studying complex disorders have also been made. This includes the development of a well-defined collection of recombinant inbred mice with different genetic backgrounds (Complex Trait Consortium, 2004). Studies from many organisms indicate that the phenotype of some gene knockouts only becomes apparent upon inactivation of another gene (Barbaric et al., 2007), indicating genetic background can be very important. An alternative, but related, approach to investigating human disease mechanisms, is to study animals that already have a disease-related phenotype of interest. These orthologous phenotypes, or phenologs, can be used to predict novel genes associated

with a disease. This approach has been used to predict genes for angiogenesis, breast cancer, autism spectrum disorder, and Waardenburg syndrome, among others, in many diverse model organisms (Gilby, 2008; McGary et al., 2010; Pearson et al., 2011). Finally, while models were previously limited to studying one variant/gene at a time, efforts are now being made to investigate the cooperative interactions of multiple genes. For example, quintuple knockout mice have been used to study the role of multiple immune system genes in asthma (Dahlin et al., 2011). Therefore, new resources and tools are being developed to study complex diseases more effectively in model organisms.

However, care must be taken in extrapolating data from animal models to the human situation, as no model organism can exactly reproduce the disease of another. It is as important to understand the differences, as it is to highlight the similarities, between the animal model and the human disease. The ability to gain a thorough understanding of the key differences and similarities between species will minimize misinterpretation of data gleaned from model organisms, and lead to the improved use of animal models to both understand and develop treatments for human genetic disorders. The importance in understanding the differences between human and animal gene expression and physiology will be discussed below, using autism spectrum disorder as an example.

## 5. Human evolution and disease susceptibility

Evolutionary analysis has been applied to many aspects of human disease. As discussed in Section 3.3.2 above, one of the earliest findings revealed that Mendelian disease-associated nsSNPs are more frequently found in conserved amino acid positions, and these positions can be conserved even in more distantly-related proteins, while complex disease-associated SNPs are frequently not (Miller & Kumar, 2001; Ng & Henikoff, 2002; Thomas et al., 2003; Thomas & Kejariwal, 2004). Some other aspects of evolutionary biology, increasing our understanding of human genetic disease, are discussed next.

Over the last 100,000 years, humans have adapted to many changes in environment as they moved out of Africa and modified both diet and lifestyle, factors which influence the incidence of common genetic variants by positive selection of those alleles that prove advantageous (Sabeti et al., 2006). This theory is supported by the finding that complex disease-associated gene variants show heterogeneity in allelic frequency among different human populations, leading to a non-homogeneous world-wide distribution of disease alleles (Ioannidis et al., 2004). These findings have implications for GWAS, as disease variants differ between human populations (e.g. between Hispanics and those of African descent, Asian descent or European descent), and will affect the reproducibility of results from genetic studies of complex disease depending on the ethnic mix studied (Marigorta et al., 2011). Therefore, the recent evolutionary history of humankind affects the present global patterns of susceptibility to disease. The classic example of this is the mutation in the *Hemoglobin B* gene, undergoing positive selection in African populations as it promotes malaria resistance, while simultaneously being causative of sickle cell anaemia (Currat et al., 2002; Williams TN, 2006). Another example is a mutation in a regulatory region of the lactase gene (*LCT*) that mediates adult tolerance to lactose. Evidence suggests this particular variant was selected in parts of Europe after the domestication of cattle (reviewed by Sabeti et al., 2006). However, identifying and understanding traits that have been targets of selection is a challenging task. It took forty



years of effort, by a succession of researchers, to unravel the association between malaria and the sickle cell mutation and, even now, there is still work to be done on understanding exactly how the sickle-cell state inhibits malaria pathogenesis (Sabeti et al., 2006).

A recent study examined Mendelian-disease genes and found these genes are under widespread negative ('purifying') selection (Blekhman et al., 2008). By contrast, in this study (Blekhman et al., 2008), genes contributing to an increased risk of complex genetic disease showed little signs of evolutionary conservation, and may be targets of both positive and purifying selection. This latter conclusion was supported by a subsequent study by Corona and colleagues (2010). In their study of genes contributing to seven complex genetic diseases, only genes affecting three diseases showed signs of recent positive selection; those increasing susceptibility to Crohn disease, rheumatoid arthritis, and diabetes (Corona et al., 2010). Of note, alleles decreasing susceptibility to Crohn disease also showed signs of positive selection (Corona et al., 2010). Overall, they found evidence for an evolutionary trajectory towards a decreasing risk of Crohn disease, but an increasing risk of type 1 diabetes.

Therefore, we may need to 'think outside the box' about why some complex disorders occur in the human population, and look not only for the disadvantages of a disease-causing mutation, but also consider unthought-of selective advantages. Furthermore, an allele increasing susceptibility to disease in the modern era, may have increased fitness in an earlier human environment. For example, rheumatoid arthritis susceptibility alleles are thought to enhance resistance to tuberculosis (Mobley, 2004; Rothschild et al., 1992), while the type 1 diabetes risk gene *IFIH1*, helps protect against enterovirus infection (Nejentsev et al., 2009). Therefore, for polygenic disorders, not only can the same polymorphisms contribute to more than one disease, some alleles may increase the risk for one disorder while simultaneously decreasing the risk for another (Sirota et al., 2009). Progress in our understanding of these complexities will be facilitated by an increased understanding of the function of disease genes and risk alleles, and rigorous studies examining natural selection in human disease-risk genes.

## 6. Applications of evolutionary biology to the study of human neurological disorders

Below, how knowledge from evolutionary biology-based analyses are providing us with crucial information impacting greatly on our understanding of neurodevelopmental disorders, and a surprising link to cancer, will be illustrated.

### 6.1 Neurodevelopmental disorders

Both schizophrenia (SCZ) and autism spectrum disorder (ASD) are considered common neurodevelopmental disorders (~1% of the population affected), which are typically diagnosed in early adulthood (SCZ) or childhood (ASD) (Bale et al., 2010; Costa e Silva, 2008; Lewis & Levitt, 2002; Owen et al., 2005). Both ASD and SCZ are behaviourally-based diagnoses, with the diagnostic criteria outlined in DSM-IV (American Psychiatric Association, 1994) and ICD-10 (World Health Organization, 1993). Briefly, SCZ is diagnosed on the presence of a collection of positive symptoms, negative symptoms, and cognitive deficits, while ASD is diagnosed on the basis of a triad of behavioural manifestations: social



deficits, impaired communication skills, together with repetitive behaviours and/or restricted interests (see Fig. 3).

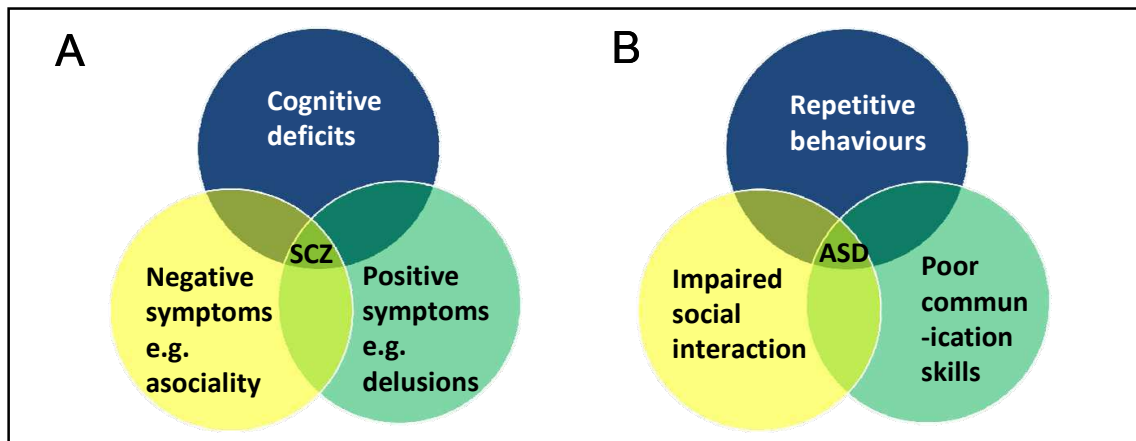


Fig. 3. Diagnostic criteria for (A) schizophrenia and (B) autism spectrum disorder.

Both ASD and SCZ are considered complex genetic disorders, with heritability estimates of around 80% for both (Ronald & Hoekstra, 2011; Sullivan et al., 2003). While some genetic disorders with Mendelian inheritance lead to syndromic forms of ASD (i.e. a phenotype of which ASD is typically one part), and while some alleles of intermediate penetrance appear to contribute, a large proportion of ASD cases fit the CVCD model (Eapen, 2011). As predicted by the CVCD model, parents of children with ASD share a subset of phenotypic traits, without having the 'full' ASD phenotype (Bernier et al., 2011; Robinson et al., 2011). A further 5-10% of ASD is caused by *de novo* mutations (reviewed by Eapen V, 2011), and this is discussed further below. For SCZ, the CVCD model of inheritance also dominates, as individuals with SCZ are less likely to pass on their genes to the next generation (Crow, 2011). Both ASD and SCZ are large, active areas of scientific research, and below we highlight those aspects where evolutionary biology is relevant.

### 6.1.2 Recurrent rearrangements in neurodevelopmental disorders

Recurrent microdeletions at 1q21.1, 15q11.2, 16p11.2, 16p13.11 and 22q11.2 are found in patients with a broad spectrum of neuropsychiatric conditions, including developmental delay, SCZ, psychotic disorder, ASD, or mental retardation (Brunetti-Pierri et al., 2008; Kumar et al., 2008; Sahoo et al., 2011; Stankiewicz & Lupski, 2010). Of note, recurrent rearrangements in synaptic and neuro-developmental genes affect shared biologic pathways contributing to the risk of developing several disorders, including SCZ, ASD, and mental retardation (Guilmatre et al., 2009; Sahoo et al., 2011).

Certain structural features make some chromosomal regions more prone to rearrangements, such as deletion, duplication or translocation (Smith et al., 2010), and this is discussed further in Section 6.2.3 below. For example, the 16p11.2 microdeletion linked to ASD is mediated by segmental duplications (Kumar et al., 2007), while deletions at 15q11-13 occur during meiosis and are caused by a number of repeated DNA elements in this chromosomal region. The 15q duplication syndrome is the result repetitive sequences mediating unequal but homologous recombination and, of note, also causes ASD (Chamberlain & Lalande, 2010). This indicates that duplication or deletion of some genes can cause an overlapping

phenotype. However, this is not a universal phenomenon. Deletion of 16p11.2 leads to a phenotype of ASD with macrocephaly, while duplication leads to SCZ and microcephaly (Brunetti-Pierri et al., 2008; McCarthy et al., 2009; Shinawi et al., 2010). A further example is duplication of 1q21.1, which is implicated in ASD with macrocephaly. Deletion of this region causes SCZ and microcephaly (Dumas & Sikela, 2009; Crespi et al., 2010). This is not restricted to neurodevelopmental disorders, as the most common locus affected in Charcot-Marie-Tooth (CMT) neuropathy, *PMP22*, when deleted causes a different neurological phenotype to that associated with gene gain (Chance, 2006). Duplication of *PMP22* leads to CMT type 1A, while *PMP22* deletion leads to a disease known as hereditary neuropathy with liability to pressure palsies (HNPP). Therefore, CNV gain, versus CNV loss, at identical loci can mediate similar or distinct phenotypes.

### 6.1.3 Phylostratigraphy of neurodevelopmental genes

As might be expected for neurodevelopmental disorders, many ASD- and SCZ-implicated gene products participate in protein-protein interaction networks implicated in neuron function (Bourgeron, 2009; Bill & Geschwind, 2009; Gilman et al., 2011; Sun et al., 2010; Torkamani et al., 2010; Voineagu et al., 2011). These genetic and PPI studies are supported by pathological findings, as structural alterations of dendritic spines are associated with both SCZ and ASD (reviewed by Penzes et al., 2011).

In addition to furthering our understanding of the evolution of human disease, model organisms play a major part in developing an understanding of the etiology of human genetic disorders. As discussed above, model organisms play a key role in characterisation of the normal and abnormal functions of risk genes (Aitman et al., 2011). Therefore, it is important to determine whether the genes implicated in SCZ and ASD are conserved in model organisms. Phylostratigraphy is a term applied to the application of phylogenetic methods to evaluate the evolutionary origin of disease genes and/or the origin of genes contributing to major evolutionary adaptations (Domazet-Loso et al., 2007; Domazet-Loso & Tautz, 2008). Such an approach has not yet been applied to any neurodevelopmental disorder. However, some data are available relating to the conservation of key PPI networks implicated in SCZ and ASD. Of relevance to SCZ and ASD, as well as other neurological disorders, is evidence that the core components of the nervous system and immune system are conserved in vertebrates. Indeed, the core components of the synapse are found in cnidarians, which form primitive nerve networks, and evolved around 680 million years ago (Galliot et al., 2009; Grimmelikhuijzen & Westfall, 1995). Furthermore, many synaptic genes are found in sponges, the oldest-surviving metazoan phyletic lineage, which actually lack synapses (Kosik et al., 2008; Srivastava et al., 2010). These data suggest many of the genes involved in neurological disorders have a more ancient evolutionary origin than previously thought.

A contributing role of the adaptive and innate immune systems, to ASD or SCZ etiology has also been suggested by some studies (Ashwood et al., 2006; Muller et al., 2000; Sun et al., 2010; Voineagu et al., 2011). This role would need to be studied in vertebrate models (possessing adaptive and innate immune systems), as non-vertebrate species only have an innate immune system. Nonetheless, a comparison of phenotypes between vertebrates and non-vertebrate may provide vital information about the relative contributions of immune dysfunction to the ASD or SCZ phenotype. A similar logic has been used to explore novel cell-death pathways, where regulated cell death processes are examined in species lacking classical caspase

enzymes (Degterev & Yuan, 2008; Guisti et al., 2010; Smirlis & Soteriadou, 2011). A thorough, systematic analysis of differences in key disease networks is also required to gain the best insights into the strengths and weaknesses of each specific model organism. As the genomes of all currently used model organisms have been sequenced, accurate network analyses and disease network analyses are now imperative if we are to understand disease evolution and the limitations and/or benefits of various model organisms for disease.

#### 6.1.4 Evolution of common neurodevelopmental disorders

A number of hypotheses regarding the evolution of autistic and schizotypal traits have been proposed, promoting controversy and debate. For example, it has been proposed that the intense focus and repetitive behaviours of ASD may have been beneficial to hunters and gatherers (Reser, 2007). This hypothesis suggests that, subsequent to the ascendance of agriculture and the development of complex community-based lifestyles, these alleles have become increasingly disadvantageous, and the common alleles responsible for ASD traits only remain in the human gene pool because they previously had adaptive function (Reser, 2011). Alleles contributing to SCZ have likewise been hypothesized to have prior adaptive function, and contributed to a fitness advantage in the ancestral human environment, such as providing physiological and behavioural characteristics increasing survival in conditions of nutritional paucity and stress (Reser, 2007). The clinical ASD or SCZ diagnosis would then ensue when individuals with subclinical phenotypes mate (Del Giudice et al., 2010), following the CVCD model.

Potential evolutionary advantages afforded by sub-clinical phenotypes of other psychiatric disorders have likewise been proposed, based on data indicating that a large proportion of these disorders are due to the inheritance of multiple copies of low-risk gene variants, which are present in the parents and siblings of affected individuals (Bernier et al., 2011; Hoffman & Stat, 2010; Robinson et al., 2011). Providing more direct evidence for these hypotheses is difficult. Indeed, the debate over the amount of adaptive evolution occurring in the genome itself is far from resolved (Amos & Bryant, 2011; Eyre-Walker, 2006). Furthermore, little evidence supports adaptive evolution in genes linked to brain development (Voight et al., 2006). Finally, as the genetic and molecular pathways underpinning both ASD and SCZ remain very poorly understood, this adds to these difficulties.

Inverse comorbidity, a concept introduced in Section 5, may also be relevant to ASD and SCZ, and the example of sickle cell anaemia gene selection occurring due to protection of carriers from malaria (Currat et al., 2002; Williams TN, 2006), was provided. Recent data supporting immune system differences in individuals with ASD or SCZ (Cohly & Panja, 2005; Müller & Schwarz, 2010), suggests that other selective pressures may be involved for risk genes for these disorders. Of relevance, there is much stronger evidence for adaptive evolution in genes with known immune-system function (reviewed by Eyre-Walker, 2006). Another possible selective pressure is found in recent studies supporting a lower than expected occurrence of cancer in patients with SCZ (Tabares-Seisdedos et al., 2011), and cancer will be discussed further later in this paper. Understanding the molecular pathways underpinning these disorders will help us clarify these issues further. However, these emerging findings suggest the selective pressure on ASD or SCZ disease alleles may not be related to the behavioural phenotype.

### 6.1.5 Model organisms for neurodevelopmental disease

Many professionals were initially sceptical whether phylogenetically 'lower' species could be successfully used to study the molecular and cellular mechanisms of human brain disorders. Surprisingly, many animal species are emerging as excellent model systems for such disorders, and their tractable nature, plus the ability to control for environment and genetic background, has already led to far-reaching advances in our understanding of many neurological diseases (Chesselet, 2005; Shah et al., 2010; Tayebati, 2006). For example, many recent reviews discuss how animal models are proving their usefulness in improving our understanding of the pathophysiology of SCZ and in the development of novel therapeutic strategies (Arguello et al., 2010; Feifel & Shilling, 2010; Lazar et al., 2011; Powell, 2010; Young et al., 2010). A greater understanding of the genetics underlying SCZ will also inform the development of future animal models. Likewise, for ASD, there are a wide variety of animal models available, which are becoming increasingly well characterized (Patterson, 2011; Tordjman et al., 2007). These include naturally-bred rodents (Gilby, 2008; Pearson et al., 2010) and transgenic mouse models (Minschew & McFadden, 2011; Robertson & Feng, 2011). Some wide-reaching findings have already been made based on results from these model systems. For example, multiple studies indicate neurodevelopmental disorders may be treatable, even in adults (Ey et al., 2011; Silva & Ehninger, 2009). Furthermore, rodent models are replicating some of the co-morbidities associated with an ASD diagnosis, such as immune abnormalities (Heo et al., 2011) and epilepsy (Gilby, 2008; Peñagarikano et al., 2001).

Despite the progress and potential that animal models of disease provide, caution must be exercised with the interpretation of data from such model systems, particularly when considering disorders affecting the central nervous system (CNS). Factors to consider when evaluating animal model data include the developmental trajectories unique to humans. The most dramatic of these is in the timing of maturation and pruning of the CNS during childhood (reviewed by Dean, 2009). Regulation of gene expression differs between animal species, with differences in microRNAs (Berezikov et al., 2006), DNA methylation patterns (Enard et al., 2004) and, as discussed next, mRNA splicing, also being detected.

Understanding the role of alternative isoforms is vitally important, as aberrant gene splicing is emerging as a key contributor to a variety of neurological diseases (Anthony & Gallo, 2010) and cancers (Ward & Cooper, 2010). Alternative splicing is considered a major mechanism underpinning metazoan biological complexity, including the increasingly-complex brain function of metazoa, with neurons having specific systems for regulating mRNA splicing and generating brain-specific isoforms (Ule & Darnell, 2007). Indeed, increasing numbers of splice site mutations are being implicated in the etiology of ASD and SCZ (Glatt et al., 2011). However, differences in splicing occur, even between closely-related species such as humans and chimpanzees (Calarco et al., 2007; Blekhman et al., 2010; Lin et al., 2010). Therefore, conservation of splice variation may be relevant to disease etiology, and different profiles should be assessed in animal models of disease. Current best-practice guidelines for preclinical studies for neurological disease in animal models, have recently been published (Shineman et al., 2011).

### 6.1.6 Evolution of the *Deleted In Autism 1* gene

*DIA1* (*Deleted In Autism 1*) is implicated in an autosomal recessive form of ASD (Morrow et al., 2008; Aziz et al., 2011a). An evolutionary biology-based approach to understanding the



role of this gene in ASD has illustrated the importance of many of the principles outlined above. While *DIA1* is conserved from cnidaria to humans, it is not detected in nematodes, suggesting *C. elegans* is not a suitable model in which to study the cellular role of this gene. Strikingly, a related gene was found in humans using phylogenetic-based analyses, *DIA1R*, which localizes to the X chromosome (Aziz et al., 2011b). *DIA1R* is vertebrate-specific and, as with *DIA1*, is implicated in ASD (Aziz et al., 2011a, 2011b). Of possible relevance to the ASD phenotype, *DIA1R* had been 'lost' in fish of a solitary nature, while those retaining the gene are 'social' schooling fish (Aziz et al., 2011b). Of further relevance to the use of animal models, *DIA1R* was found to be X-inactivated in mouse, but not in humans (Aziz et al., 2011b; Yang et al., 2010), and splicing may be species-specific. Indeed, we have preliminary evidence for brain-specific splicing of *DIA1* that is primate-specific, and not found in other vertebrate lineages (Aziz & Bishop, unpublished data). This type of evolutionary-based evidence facilitates an educated approach to the choice of model organism for functional studies, and highlights issues that may arise from studies in mice or fish.

## 6.2 Cancer

Natural selection and evolution is dependent on genetic variability and the occurrence of new mutations in the germ-line. However, mutation occurs in both germinal and somatic cell lineages and, over a human lifespan, somatic mutations accumulate and may lead to cancer (Greaves, 2007). Therefore, as with other genetic disorders, carcinogenesis is another manifestation of the biological processes on which evolution depends. Cancer is considered a probabilistic disease, and is inevitable in long-lived organisms such as humans, where the lifetime risk is around one in three (Greaves, 2007; Simpson & Camargo, 1998). In most, but not all, cases this is due to multiple genetic changes accumulating in cells (Maffini et al., 2004; Stratton et al., 2009; Touw & Erkeland, 2007).

Genes contributing to cancer are often divided into two broad groups: caretakers and gatekeepers (Kinzler & Vogelstein, 1997; Macleod, 2000; Michor et al., 2004; Russo et al., 2006). Mutations affecting caretaker genes promote neoplasia indirectly, increasing genetic instability and mutation rates, which leads to defects in many genes including gatekeeper genes. Mutations affecting gatekeeper genes directly promote cancer progression, and have roles in cell differentiation, growth and/or death. Gatekeeper genes may be further divided into oncogenes and tumour-suppressor genes. Well-known caretaker genes include *BRCA1*, *BRCA2*, *ATM* and *FANCA*; gatekeeper genes include *TSC1*, *TSC2*, *Rb*, *NF1*, *NF2* and *PTEN* (see Fig. 3); while some genes are multifunctional and can act as gatekeepers and/or caretakers, including *p53* and *ARF* (Dominguez-Brauer et al., 2010; Rubbi & Milner, 2005; Russo et al., 2006).

### 6.2.1 Cancer and neurological disorders

One surprising recent finding is the emerging link between a number of human disorders and cancer. Recent data indicate a lower than expected occurrence of cancer in patients with SCZ, Down syndrome, Parkinson disease, Alzheimer disease and multiple sclerosis (Tabares-Seisdedos et al., 2011). What is striking, is that most of the disorders found to protect against cancer, and which lead to an inverse cancer morbidity in humans, are neurological disorders. This is thought to occur due to the genetic and molecular



connections between cancer and these complex human diseases (Tabares-Seisdedos et al., 2011). Of greater concern, is a higher incidence of certain cancers in patients with ASD (Crespi, 2011). This is likewise due to shared molecular pathways (Crespi, 2011), and ASD also shares genetic connections with SCZ and Alzheimer disease (Crespi et al., 2010; Sokol et al., 2011). Indeed, some lines of evidence suggest that ASD and SCZ involve diametric etiology (Crespi et al., 2010). Exploring these links will have far-reaching implications.

### 6.2.2 Evolution of cancer genes

A link between cancer formation and the evolution of multicellularity was predicted, and this hypothesis was recently explored (Domazet-Loso & Tautz, 2010). Using a phylostratigraphic strategy, Domazet-Loso and Tautz (2010) found different evolutionary origins for gatekeeper, compared to caretaker, human cancer genes. Genes with a caretaker function have an evolutionary origin in the first cellular organisms, and are also found in bacteria and/or archaea (Domazet-Loso & Tautz, 2010). By contrast, only genes with gatekeeper functionality correspond to the origin of metazoa, and are detected in Porifera and/or Cnidaria and subsequently diverging phyla (see Fig. 4). For example, genes encoding the ESCRT (endosomal sorting complex required for transport) proteins, are essential for the downregulation of many cell-surface signalling molecules (reviewed by Ilievska et al., 2011). Not only do many of the genes encoding ESCRT subunits appear to have caretaker functions, with mutations being linked to both cancer and neurological disorders, these genes also have an origin pre-dating the metazoan lineage (reviewed by Ilievska et al., 2011). Model organisms, such as the amoeba, *Dictyostelium discoideum* (Annesley & Fisher, 2009; Williams RS et al., 2006), can therefore be used to study the fundamental molecular roles of these proteins (Blanc et al., 2009; Mattei et al., 2006). Overall, the evolutionary age of cancer genes parallels their role in human cancer etiology (Fig. 4).

### 6.2.3 Heritability of cancer and co-heritability with neurological disease

Most tumour gene mutations can be inherited as well as acquired, although inherited cancer syndromes are quite rare. The cancer syndromes are inherited in a dominant Mendelian manner and are often associated with developmental defects and benign tumours (Knudson, 1971, 1993; Ponder, 2001). Single genes and deletions of multiple genes can both cause syndromes of which cancer is one part. For example, deletions of multiple genes at position 11p13, is clinically associated with WAGR syndrome and patients have a collection of symptoms including mental retardation, aniridia and kidney tumours (Fischbach et al., 2005). Mutations in *TSC1* or *TSC2* cause tuberous sclerosis, a disease characterised by non-malignant hamartoma formation in many organs, with epilepsy, ASD and/or mental retardation being frequent comorbidities (de Vries, 2010), while *PTEN* mutations are identified in human cancers, and also in the germline of patients with hamartoma tumour-related syndromes (PHTSs). In addition to hamartomas, ASD is a common comorbidity in individuals with *PTEN* mutations (Rodríguez-Escudero et al., 2011). Molecular network analyses (e.g. Fig. 2) are being used to establish the molecular pathways leading to the multiple phenotypes caused by mutation a single 'cancer gene' (de Vries, 2010; Rodríguez-Escudero et al., 2011). Genetic and functional studies are also being used, with current data indicating that different mutations in the *PTEN* gene can cause cancer alone, or cancer with ASD (Rodríguez-Escudero et al., 2011).

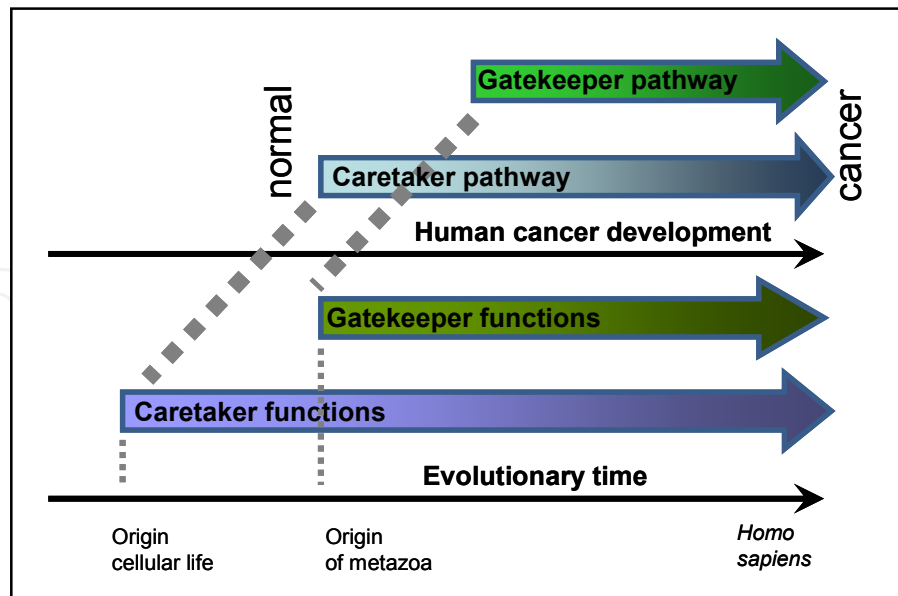


Fig. 4. The evolutionary age of cancer genes parallels their role in carcinogenesis (adapted from Michor & Tautz, 2010).

#### 6.2.4 Recurrent genome rearrangements and cancer

Both somatic-cell chromosome rearrangements and germ-line rearrangements are implicated in both human variation and disease, including the development of cancers (Hanahan & Weinberg, 2000; Kidd et al., 2008; Inoue & Lupski, 2002; Stankiewicz & Lupski, 2002). On an evolutionary timescale, chromosome rearrangements have also played a key role in the divergence of species and differences in chromosomal arrangements between species (Drosophila 12 Genomes Consortium, 2007; Kehrer-Sawatzki & Cooper, 2007; Pevzner & Tesler, 2003; Peng et al., 2006). Furthermore, the evolutionary chromosome breakpoints and chromosome breakpoints found in diseases, including cancer and developmental disorders, overlap (Darai-Ramqvist et al., 2008; Lindsay et al., 2006; Murphy et al., 2005). Therefore, understanding the mechanisms of chromosome breakage and rearrangement is an active field of biological research.

Initially, genomic rearrangements were thought to occur randomly at non-specific, independent sites (Nadeau & Taylor, 1984; Ohno, 1970). However, subsequent studies, including those on cancer breakpoints, found non-random patterns (Cohen et al., 1996; Larkin et al., 2003; Pevzner & Tesler G, 2003; Sankoff et al., 2002), leading to the fragile-site model, where specific regions in the genome are 'hot spots' for genome rearrangements. Since that time, the most highly-fragile spots in the human genome have been mapped (Smith CL et al., 2010). The fragile-site model leads to various testable hypotheses as to the nature of these hot spots, and factors such as DNA sequence and chromosome structure were proposed. Since then, evolutionary breakpoints have been shown, for example, to be gene-rich regions with significantly more segmental duplications and/or repetitive elements than expected, which may facilitate homologous recombination (Bailey et al., 2004; Bulazel et al., 2007; Everts-van der Wind et al., 2004, 2005; Kehrer-Sawatzki & Cooper, 2008; Murphy et al., 2005; Schibler et al., 2006). These regions also encompass genes with higher densities

of copy number variation and SNPs (Larkin et al., 2009). Understanding chromosomal 'hotspots' will not only inform our understanding of evolutionary biology, but such studies will also increase our understanding of the etiology of human genetic disorders caused by chromosomal rearrangements.

## 7. Conclusion

Genetic disorders are an inescapable component of evolution. However, comparative and evolutionary biology methodologies also play an important role in developing and improving our comprehension of many aspects of complex human genetic disorders including schizophrenia and autism spectrum disorder. Thorough *in silico* analyses of disease genes, the encoded proteins, their structure and function, are fundamental tools for evaluating differences and similarities between human genes and genetic pathways, and the comparison of these with equivalents in animal models. Such explorations of the evolutionary conservation of human genes, proteins, protein structures, and resulting cellular networks enables us to: (i) understand how and why human diseases originated; (ii) predict disease genes; (iii) best predict the impact of genetic variation on cellular function; (iv) choose appropriate animal models for human diseases; (v) better interpret data obtained in studies using model organisms; and (vi) evaluate more accurately the validity of therapeutics. Therefore, comparative and evolutionary biology is of major relevance to our understanding of, and in the development of treatments for, many human neurological disorders.

## 8. Acknowledgment

Thanks to all members of the NEB and PRF laboratories for helpful discussions. We apologize to those authors whose work we have been unable to cite directly, due to space limitations. AA was supported by the Malaysian Ministry of Higher Education and the Islamic Science University of Malaysia. JI was supported by an Australian postgraduate award. This work was supported by grants from the Australian Research Council and the Thyne Reid Memorial Trust.

## 9. References

- Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS. (2005). Speeding disease gene discovery by sequence based candidate prioritization. *BMC Bioinformatics*, 6:55.
- Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS. (2006). SUSPECTS: enabling fast and effective prioritization of positional candidates. *Bioinformatics*, 22(6):773-4.
- Aitman TJ, Boone C, Churchill GA, Hengartner MO, Mackay TF, Stemple DL. (2011). The future of model organisms in human disease research. *Nat Rev Genet*, 12(8):575-82.
- Almeida AM, Murakami Y, Layton DM, Hillmen P, Sellick GS, et al. (2006). Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. *Nat Med*, 12(7):846-51.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389-402.
- Altshuler D, Daly MJ, Lander ES. (2008). Genetic mapping in human disease. *Science*, 322(5903):881-8.

- Amberger J, Bocchini CA, Scott AF, Hamosh A. (2009). McKusick's online mendelian inheritance in man (OMIM). *Nucleic Acids Res*, 37:D793-6.
- American Psychiatric Association (1994). *Diagnostic & Statistical Manual of Mental Disorders 4<sup>th</sup> Edition*, American Psychiatric Association, Washington DC, USA.
- Amos W, Bryant C. (2011). Using human demographic history to infer natural selection reveals contrasting patterns on different families of immune genes. *Proc Biol Sci*, 278(1711):1587-94.
- Ananda G, Chiaromonte F, Makova KD. (2011). A genome-wide view of mutation rate covariation using multivariate analyses. *Genome Biol*, 12(3):R27.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806):457-65.
- Annesley SJ, Fisher PR. (2009). *Dictyostelium discoideum* - a model for many reasons. *Mol Cell Biochem*, 329(1-2):73-91.
- Ansari BM, Mason MK. (1977). Chromosomal abnormality in congenital rubella. *Pediatrics*, 59(1):13-5.
- Anthony K, Gallo JM. (2010). Aberrant RNA processing events in neurological disorders. *Brain Res*, 1338:67-77.
- Antonarakis SE, Beckmann JS. (2006). Mendelian disorders deserve more attention. *Nat Rev Genet*, 7(4):277-82.
- Arguello PA, Markx S, Gogos JA, Karayiorgou M. (2010). Development of animal models for schizophrenia. *Dis Model Mech*, 3(1-2):22-6.
- Arnheim N, Calabrese P. (2009). Understanding what determines the frequency and pattern of human germline mutations. *Nat Rev Genet*, 10(7):478-88.
- Ashwood P, Wills S, Van de Water J. (2006). The immune response in autism: a new frontier for autism research. *J Leukoc Biol*, 80(1):1-15.
- Attwood TK, Bradley P, Flower DR, Gaulton A, Maudling N, et al. (2003). PRINTS and its automatic supplement, preprints. *Nucleic Acids Res*, 31(1):400-2.
- Axton M. (2010). Editorial: On beyond GWAS. *Nat Genet*, 42(7):551.
- Aziz A, Harrop SP, Bishop NE. (2011a). DIA1R is an X-linked gene related to Deleted In Autism-1. *PLoS One*, 6(1): e14534.
- Aziz A, Harrop SP, Bishop NE. (2011b). Characterization of the Deleted In Autism 1 protein family: implications for studying cognitive disorders. *PLoS One*, 6(1):e14547.
- Bale TL, Baram TZ, Brown AS, Goldstein JM, Insel TR, et al. (2010). Early life programming and neurodevelopmental disorders. *Biol Psychiatry*, 68(4):314-9.
- Bao Z, Eddy SR. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res*, 12(8):1269-76.
- Bailey JA, Eichler EE. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet*, 7(7):552-64.
- Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE. (2004). Hotspots of mammalian chromosomal evolution. *Genome Biol*, 5(4):R23.
- Barbaric I, Miller G, Dear TN. (2007). Appearances can be deceiving: phenotypes of knockout mice. *Brief Funct Genomic Proteomic*, 6(2):91-103.
- Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, et al. (2009). The GOA database in 2009-an integrated gene ontology annotation resource. *Nucleic Acids Res*, 37:D396-403.
- Batzer MA, Deininger PL. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet*, 3(5):370-9.



- Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, et al. (2006). Diversity of microRNAs in human and chimpanzee brain. *Nat Genet*, 38(12):1375-7.
- Berg JS, Potocki L, Bacino CA. (2010). Common recurrent microduplication syndromes: diagnosis and management in clinical practice. *Am J Med Genet A*, 152A(5):1066-78.
- Bernier R, Gerds J, Munson J, Dawson G, Estes A. (2011). Evidence for broader autism phenotype characteristics in parents from multiple-incidence autism families. *Autism Res*, (Sep 8) in press.
- Bill BR, Geschwind DH. (2009). Genetic advances in autism: heterogeneity and convergence on shared pathways. *Curr Opin Genet Dev*, 19(3):271-8.
- Blanc C, Charette SJ, Mattei S, Aubry L, Smith EW, et al. (2009). Dictyostelium Tom1 participates to an ancestral ESCRT-0 complex. *Traffic*, 10(2):161-71.
- Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, et al. (2008). Natural selection on genes that underlie human disease susceptibility. *Curr Biol*, 18(12):883-9.
- Blekhman R, Marioni JC, Zumbo P, Stephens M, Gilad Y. (2010). Sex-specific and lineage-specific alternative splicing in primates. *Genome Res*, 20(2):180-9.
- Blennow K, de Leon MJ, Zetterberg H. (2006). Alzheimer's disease. *Lancet*, 368(9533):387-403.
- Bourgeron T. (2009). A synaptic trek to autism. *Curr Opin Neurobiol*, 19(2):231-4.
- Borck G, Zarhrate M, Cluzeau C, Bal E, Bonnefont JP, et al. (2006). Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5' untranslated region of the NIPBL gene. *Hum Mutat*, 27(8):731-5.
- Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, et al. (2008). Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet*, 40(12):1466-71.
- Buckingham SD, Esmaili B, Wood M, Sattelle DB. (2004). RNA interference: from model organisms towards therapy for neural and neuromuscular disorders. *Hum Mol Genet*, 13(R2):R275-88.
- Bulazel KV, Ferreri GC, Eldridge MD, O'Neill RJ. (2007). Species-specific shifts in centromere sequence composition are coincident with breakpoint reuse in karyotypically divergent lineages. *Genome Biol*, 8(8):R170.
- Cai JJ, Borenstein E, Chen R, Petrov DA. (2009). Similarly strong purifying selection acts on human disease genes of all evolutionary ages. *Genome Biol Evol*, 1:131-44.
- Calarco JA, Xing Y, Cáceres M, Calarco JP, Xiao X, et al. (2007). Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev*, 21(22):2963-75.
- Carvalho CM, Zhang F, Lupski JR. (2010). Evolution in health and medicine Sackler colloquium: Genomic disorders: a window into human gene and genome evolution. *Proc Natl Acad Sci USA*, 107 (Suppl 1):1765-71.
- Chamberlain SJ, Lalande M. (2010). Neurodevelopmental disorders involving genomic imprinting at human chromosome 15q11-q13. *Neurobiol Dis*, 39(1):13-20.
- Chance PF. (2006). Inherited focal, episodic neuropathies: hereditary neuropathy with liability to pressure palsies and hereditary neuralgic amyotrophy. *Neuromolecular Med*, 8(1-2):159-74.
- Chen PY, Deane CM, Reinert G. (2008). Predicting and validating protein interactions using network structure. *PLoS Comput Biol*, 4(7):e1000118.
- Chesselet MF. (2005). Animal models of neurological disorders. *NeuroRx*, 2(3):395.
- Chua HN, Wong L. (2008). Increasing the reliability of protein interactomes. *Drug Discov Today*, 13(15-16):652-8.



- Claverie-Martin F, González-Acosta H, Flores C, Antón-Gamero M, García-Nieto V. (2003). De novo insertion of an Alu sequence in the coding region of the CLCN5 gene results in Dent's disease. *Hum Genet*, 113(6):480-5.
- Cohen O, Cans C, Cuillel M, Gilardi JL, Roth H, et al. (1996). Cartographic study: breakpoints in 1574 families carrying human reciprocal translocations. *Hum Genet*, 97(5):659-67.
- Cohly HH, Panja A. (2005). Immunological findings in autism. *Int Rev Neurobiol*, 71:317-41.
- Collier LS, Largaespada DA. (2007). Transposable elements and the dynamic somatic genome. *Genome Biol*, 8 (Suppl 1):S5.
- Complex Trait Consortium. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*, 36(11):1133-7.
- Conant GC, Wolfe KH. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*, 9(12):938-50.
- Corona E, Dudley JT, Butte AJ. (2010). Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS One*, 5(8):e12236.
- Costa e Silva JA. (2008). Autism, a brain developmental disorder: some new pathophysiologic and genetics findings. *Metabolism*, 57(Suppl 2):S40-3.
- Crespi B. (2011). Autism and cancer risk. *Autism Res*, 4:301-10.
- Crespi B, Stead P, Elliot M. (2010). Evolution in health and medicine Sackler colloquium: Comparative genomics of autism and schizophrenia. *Proc Natl Acad Sci USA*, 107(Suppl 1):1736-41.
- Crow TJ. (2011). "Just the facts" of schizophrenia in the context of human evolution: commentary. *Schizophr Res*, 129(2-3):205-7.
- Currat M, Trabuchet G, Rees D, Perrin P, Harding RM, et al. (2002). Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) mutation. *Am J Hum Genet*, 70(1):207-23.
- Dahlin JS, Ivarsson MA, Heyman B, Hallgren J. (2011). IgE immune complexes stimulate an increase in lung mast cell progenitors in a mouse model of allergic airway inflammation. *PLoS One*, 6(5):e20261.
- Darai-Ramqvist E, Sandlund A, Müller S, Klein G, Imreh S, et al. (2008). Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. *Genome Res*, 18(3):370-9.
- Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR. (2005). Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet*, 366(9495):1484-98.
- Dean B. (2009). Is schizophrenia the price of human central nervous system complexity? *Aust NZ J Psychiatry*, 43(1): 13-24.
- Degterev A, Yuan J. (2008). Expansion and evolution of cell death programmes. *Nat Rev Mol Cell Biol*, 9(5):378-90.
- Del Giudice M, Angeleri R, Brizio A, Elena MR. (2010). The evolution of autistic-like and schizotypal traits: a sexual selection hypothesis. *Front Psychol*, 1:41.
- de Vries PJ. (2010). Targeted treatments for cognitive and neurodevelopmental disorders in tuberous sclerosis complex. *Neurotherapeutics*, 7(3):275-82.
- Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, et al. (2008). A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol*, 6:49.
- Domazet-Lošo T, Tautz D. (2008). An ancient evolutionary origin of genes associated with human genetic diseases. *Mol Biol Evol*, 25(12):2699-707.

- Domazet-Lošo T, Brajkovi J, Tautz D. (2007). A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet*, 23(11):533-9.
- Dominguez-Brauer C, Brauer PM, Chen YJ, Pimkina J, Raychaudhuri P. (2010). Tumor suppression by ARF: gatekeeper and caretaker. *Cell Cycle*, 9(1):86-9.
- Drosophila 12 Genomes Consortium. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, 450(7167):203-18.
- Dumas L, Sikela JM. (2009). DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb Symp Quant Biol*, 74:375-82.
- du Plessis L, Skunca N, Dessimoz C. (2011). The what, where, how and why of gene ontology-a primer for bioinformaticians. *Brief Bioinform*, 12(6):723-35.
- Dziembowski A, Séraphin B. (2004). Recent developments in the analysis of protein complexes. *FEBS Lett*, 556(1-3):1-6.
- Eapen V. (2011). Genetic basis of autism: is there a way forward? *Curr Opin Psychiatry*, 24(3):226-36.
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447(7148):1087-93.
- Eisenberg E, Levanon EY. (2003). Preferential attachment in the protein network evolution. *Phys Rev Lett*, 91(13):138701.
- Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. (2000). Protein function in the post-genomic era. *Nature*, 405(6788):823-826.
- Enard W, Fassbender A, Model F, Adorján P, Pääbo S, et al. (2004). Differences in DNA methylation patterns between humans and chimpanzees. *Curr Biol*, 14(4):R148-9.
- Eng KH, Bravo HC, Kele S. (2009). A phylogenetic mixture model for the evolution of gene expression. *Mol Biol Evol*, 26(10):2363-72.
- Erdin S, Lisewski AM, Lichtarge O. (2011). Protein function prediction: towards integration of similarity metrics. *Curr Opin Struct Biol*, 21(2):180-8.
- Everts-van der Wind A, Kata SR, Band MR, Rebeiz M, Larkin DM, et al. (2004). A 1463 gene cattle-human comparative map with anchor points defined by human genome sequence coordinates. *Genome Res*, 14(7):1424-37.
- Everts-van der Wind A, Larkin DM, Green CA, Elliott JS, Olmstead CA, et al. (2005). A high-resolution whole-genome cattle-human comparative map reveals details of mammalian chromosome evolution. *Proc Natl Acad Sci USA*, 102(51):18526-31.
- Ey E, Leblond CS, Bourgeron T. (2011). Behavioral profiles of mouse models for autism spectrum disorders. *Autism Res*, 4(1):5-16.
- Eyre-Walker A. (2006). The genomic rate of adaptive evolution. *Trends Ecol Evol*, 21(10):569-75.
- Feifel D, Shilling PD. (2010). Promise and pitfalls of animal models of schizophrenia. *Curr Psychiatry Rep*, 12 (4): 327-334.
- Feldman I, Rzhetsky A, Vitkup D. (2008). Network properties of genes harboring inherited disease mutations. *Proc Natl Acad Sci USA*, 105(11):4323-8.
- Fischbach BV, Trout KL, Lewis J, Luis CA, Sika M. (2005). WAGR syndrome: a clinical review of 54 cases. *Pediatrics*, 116(4):984-8.
- Fonseka KG, Griffin DK. (2011). Is there a paternal age effect for aneuploidy? *Cytogenet Genome Res*, 133(2-4):280-91.

- Fortunato EA, Spector DH. (2003). Viral induction of site-specific chromosome damage. *Rev Med Virol*, 13(1):21-37.
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. (2002). Evolutionary rate in the protein interaction network. *Science*, 296(5568):750-2.
- Friedberg I. (2006). Automated protein function prediction- the genomic challenge. *Brief Bioinform*, 7(3):225-42.
- Galliot B, Quiquand M, Ghila L, de Rosa R, et al. (2009). Origins of neurogenesis, a cnidarian view. *Dev Biol*, 332(1):2-24.
- Gandhi S, Wood NW. (2010). Genome-wide association studies: the key to unlocking neurodegeneration? *Nat Neurosci*, 13(7):789-94.
- Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, et al. (2006). Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet*, 38(3):285-93.
- Gasser T. (2009). Mendelian forms of Parkinson's disease. *Biochim Biophys Acta*, 1792(7):587-96.
- Gaulton KJ, Mohlke KL, Vision TJ. (2007). A computational system to select candidate genes for complex human traits. *Bioinformatics*, 23(9):1132-40.
- Genin E, Hannequin D, Wallon D, Sleegers K, Hiltunen M, et al. (2011). APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol Psychiatry*, 16(9):903-7.
- Giallourakis C, Henson C, Reich M, Xie X, Mootha VK. (2005). Disease gene discovery through integrative genomics. *Annu Rev Genomics Hum Genet*, 6:381-406.
- Gilby KL. (2008). A new rat model for vulnerability to epilepsy and ASDs. *Epilepsia*, 49(Suppl 8):108-10.
- Gillis J, Pavlidis P. (2011). The impact of multifunctional genes on "guilt by association" analysis. *PLoS One*, 6 (2): e17258.
- Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, et al. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron*, 70(5):898-907.
- Giusti C, Luciani MF, Golstein P. (2010). A second signal for autophagic cell death? *Autophagy*, 6(6):823-4.
- Glatt SJ, Cohen OS, Faraone SV, Tsuang MT. (2011). Dysfunctional gene splicing as a potential contributor to neuropsychiatric disorders. *Am J Med Genet B Neuropsychiatr Genet*, 156B (4):382-92.
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. (2007). The human disease network. *Proc Natl Acad Sci USA*, 104(21):8685-90.
- Goldsmith-Fischman S, Honig B. (2003). Structural genomics: computational methods for structure analysis. *Protein Sci*, 12(9):1813-21.
- Grady DL, Chi HC, Ding YC, Smith M, Wang E, et al. (2003). High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. *Mol Psychiatry*, 8(5):536-45.
- Greaves M. (2007). Darwinian medicine: a case for cancer. *Nat Rev Cancer*, 7(3):213-21.
- Grimmelikhuijzen CJ, Westfall JA. (1995). The nervous systems of cnidarians. *EXS*, 72:7-24.
- Guan C, Ye C, Yang X, Gao J. (2010). A review of current large-scale mouse knockout efforts. *Genesis*, 48(2):73-85.
- Guilmatre A, Dubourg C, Mosca AL, Legalic S, Goldenberg A, et al. (2009). Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic

- pathways in schizophrenia, autism, and mental retardation. *Arch Gen Psychiatry*, 66(9):947-56.
- Han JD. (2008). Understanding biological functions through molecular networks. *Cell Res*, 18(2):224-37.
- Han MV, Demuth JP, McGrath CL, Casola C, Hahn MW. (2009). Adaptive evolution of young gene duplicates in mammals. *Genome Res*, 19(5):859-67.
- Haitina T, Fredriksson R, Foord SM, Schiöth HB, Gloriam DE. (2009). The G protein-coupled receptor subset of the dog genome is more similar to that in humans than rodents. *BMC Genomics*, 10:24.
- Hanahan D, Weinberg RA. (2000). The hallmarks of cancer. *Cell*, 100(1):57-70.
- Häsler J, Strub K. (2006). Alu elements as regulators of gene expression. *Nucleic Acids Res*, 34(19):5491-7.
- Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. (2000). Blocks-based methods for detecting protein homology. *Electrophoresis*, 21(9):1700-6.
- Heo Y, Zhang Y, Gao D, Miller VM, Lawrence DA. (2011). Aberrant immune responses in a mouse with behavioral disorders. *PLoS One*, 6(7):e20912.
- Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*, 106(23):9362-9367.
- Hoffman EJ, State MW. (2010). Progress in cytogenetics: implications for child psychopathology. *J Am Acad Child Adolesc Psychiatry*, 49(8):736-751.
- Huang H, Winter EE, Wang H, Weinstock KG, Xing H, et al. (2004). Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol*, 5(7):R47.
- Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuče BA, et al. (2008). The 20 years of PROSITE. *Nucleic Acids Res*, 36:D245-9.
- Hutz JE, Kraja AT, McLeod HL, Province MA. (2008). CANDID: a flexible method for prioritizing candidate genes for complex human traits. *Genet Epidemiol*, 32(8):779-90.
- Ilievska J, Bishop NE, Annesley SJ, Fisher PR. (2011). The roles of ESCRT proteins in healthy cells and in disease, In: *Cell Biology*, S. Najman, (Ed.), InTech, Rijeka, Croatia.
- Inoue K, Lupski JR. (2002). Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet*, 3:199-242.
- International Parkinson Disease Genomics Consortium (2011). Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*, 377(9766):641-9.
- Ioannidis JP, Ntzani EE, Trikalinos TA. (2004). 'Racial' differences in genetic effects for complex diseases. *Nat Genet*, 36(12):1312-8.
- Iwata K, Matsuzaki H, Takei N, Manabe T, Mori N. (2010). Animal models of autism: An epigenetic and environmental viewpoint. *J Central Nervous Syst Dis*, 2:37-44. Retrieved from [www.la-press.com](http://www.la-press.com)
- Jeffery CJ. (2009). Moonlighting proteins- an update. *Mol Biosyst*, 5(4):345-50.
- Jeong H, Mason SP, Barabási AL, Oltvai ZN. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833):41-2.
- Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, et al. (2009). Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet*, 5(8):e1000597.



- Jiang MC, Lien YR, Chen SU, Ko TM, Ho HN, et al. (1999). Transmission of de novo mutations of the deleted in azoospermia genes from a severely oligozoospermic male to a son via intracytoplasmic sperm injection. *Fertil Steril*, 71(6):1029-32.
- Jones KT. (2008). Meiosis in oocytes: predisposition to aneuploidy and its increased incidence with age. *Hum Reprod Update*, 14(2):143-58.
- Jorde LB. (2000). Linkage disequilibrium and the search for complex disease genes. *Genome Res*, 10(10):1435-44.
- Kehrer-Sawatzki H, Cooper DN. (2007). Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum Mutat*, 28(2):99-130.
- Kehrer-Sawatzki H, Cooper DN. (2008). Molecular mechanisms of chromosomal rearrangement during primate evolution. *Chromosome Res*, 16(1):41-56.
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Samps N, et al. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*, 453(7191):56-64.
- Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, et al. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, 315(5811):525-8.
- Kingsley CB. (2011). Identification of causal sequence variants of disease in the next generation sequencing era. *Methods Mol Biol*, 700:37-46.
- Kinzler KW, Vogelstein B. (1997). Cancer-susceptibility genes -- Gatekeepers and caretakers. *Nature*, 386 (6627): 761- 763.
- Knudson AG. (1993). All in the (cancer) family. *Nat Genet*, 5(2):103-4.
- Kosik KS. (2009). Exploring the early origins of the synapse by comparative genomics. *Biol Lett*, 5(1):108-11.
- Kozul R, Fischer G. (2009). A prominent role for segmental duplications in modeling eukaryotic genomes. *C R Biol*, 332(2-3):254-66.
- Ku CS, Naidoo N, Pawitan Y. (2011). Revisiting Mendelian disorders through exome sequencing. *Hum Genet*, 129(4):351-70.
- Kuchaiev O, Wang PT, Nenadic Z, Przulj N. (2009). Structure of brain functional networks. *Conf Proc IEEE Eng Med Biol Soc*, 2009:4166-70.
- Kuchaiev O, Rasajski M, Higham DJ, Przulj N. (2009). Geometric de-noising of protein-protein interaction networks. *PLoS Comput Biol*, 5(8):e1000454.
- Kuhlenbäumer G, Hullmann J, Appenzeller S. (2011). Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat*, 32(2):144-51.
- Kumar D. (2008). Disorders of the genome architecture: a review. *Genomic Med*, 2(3-4):69-76.
- Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C. et al. (2008). Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet*, 17(4):628-38.
- Kurahashi H, Inagaki H, Ohye T, Kogo H, Kato T, Emanuel BS. (2006). Chromosomal translocations mediated by palindromic DNA. *Cell Cycle*, 5(12):1297-303.
- Lagerström MC, Hellström AR, Gloriam DE, Larsson TP, Schiöth HB, et al (2006). The G protein-coupled receptor subset of the chicken genome. *PLoS Comput Biol*, 2(6):e54.
- Larkin DM, Everts-van der Wind A, Rebeiz M, Schweitzer PA, Bachman S, et al. (2003). A cattle-human comparative map built with cattle BAC-ends and human genome sequence. *Genome Res*, 13(8):1966-72.

- Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA. (2009). Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res*, 19(5):770-7.
- Laskowski RA, Watson JD, Thornton JM. (2003). From protein structure to biochemical function? *J Struct Funct Genomics*, 4(2-3):167-77.
- Lazar NL, Neufeld RW, Cain DP. (2011). Contribution of nonprimate animal models in understanding the etiology of schizophrenia. *J Psychiatry Neurosci*, 36(4):E5-29.
- Lee SH, Wray NR, Goddard ME, Visscher PM. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*, 88(3):294-305.
- Lehmann AR. (2001). The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev*, 15(1):15-23.
- Lenski C, Kooy RF, Reyniers E, Loessner D, Wanders RJ, et al. (2007). The reduced expression of the HADH2 protein causes X-linked mental retardation, choreoathetosis, and abnormal behavior. *Am J Hum Genet*, 80(2):372-7.
- Lesage S, Brice A. (2009). Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet*, 18(R1):R48-59.
- Lewis DA, Levitt P. (2002). Schizophrenia as a disorder of neurodevelopment. *Annu Rev Neurosci*, 25:409-32.
- Liao BY, Zhang J. (2008). Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci USA*, 105(19):6987-92.
- Lin L, Shen S, Jiang P, Sato S, Davidson BL, et al. (2010). Evolution of alternative splicing in primate brain transcriptomes. *Hum Mol Genet*, 19(15):2958-73.
- Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, et al. (2011). A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N Engl J Med*, 365(7):611-9.
- Lindsay SJ, Khajavi M, Lupski JR, Hurles ME. (2006). A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am J Hum Genet*, 79(5):890-902.
- Liu X, Cheng R, Verbitsky M, Kisselev S, Browne A, et al. (2011). Genome-Wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC Med Genet*, 12:104.
- Lobo I. (2008) Same genetic mutation, different genetic disease phenotype. *Nature Education*, 1(1), Retrieved from <http://www.nature.com/scitable/topicpage/same-genetic-mutation-different-genetic-disease-phenotype-938>
- Lupski JR. (2010). New mutations and intellectual function. *Nat Genet*, 42(12):1036-8.
- Lynch M, Conery JS. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290 (5494):1151-1155.
- Macleod K (2000). Tumor suppressor genes. *Curr Opin Genet Dev*, 10(1):81-93.
- Maffini MV, Soto AM, Calabro JM, Ucci AA, Sonnenschein C. (2004). The stroma as a crucial target in rat mammary gland carcinogenesis. *J Cell Sci*, 117(Pt 8):1495-502.
- Magrane M, Consortium U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)*, 2011:bar009.
- Makino T, Gojobori T. (2006). The evolutionary rate of a protein is influenced by features of the interacting partners. *Mol Biol Evol*, 23(4):784-9.
- Manolio TA. (2010). Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, 363(2):166-76.

- Marigorta UM, Lao O, Casals F, Calafell F, Morcillo-Suárez C, et al. (2011). Recent human evolution has shaped geographical differences in susceptibility to disease. *BMC Genomics*, 12:55.
- Marques-Bonet T, Girirajan S, Eichler EE. (2009). The origins and impact of primate segmental duplications. *Trends Genet*, 25(10):443-54.
- Marth GT, Yu F, Indap AR, Garimella K, Gravel S, et al. (2011). The functional spectrum of low-frequency coding variation. *Genome Biol*, 12(9):R84.
- Martinez J, Dugaiczky LJ, Zielinski R, Dugaiczky A. (2001). Human genetic disorders, a phylogenetic perspective. *J Mol Biol*, 308(4):587-96.
- Mattei S, Klein G, Satre M, Aubry L. (2006). Trafficking and developmental signaling: Alix at the crossroads. *Eur J Cell Biol*, 85(9-10):925-36.
- McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, et al. (2009). Micro-duplications of 16p11.2 are associated with schizophrenia. *Nat Genet*, 41(11):1223-7.
- McCarthy MI, Hirschhorn JN. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet*, 17(R2):R156-65.
- McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, et al. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci USA*, 107(14):6544-9.
- Michor F, Iwasa Y, Nowak MA. (2004). Dynamics of cancer progression. *Nat Rev Cancer*, 4(3):197-205.
- Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. (2009). Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *BMC Genomics*, 10(Suppl 1):S12.
- Miller MP, Kumar S. (2001). Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet*, 10(21):2319-28.
- Minschew N, McFadden K. Commentary for special issue of autism research on mouse models in ASD: a clinical perspective. *Autism Res*, 4(1):1-4.
- Mobley JL. (2004). Is rheumatoid arthritis a consequence of natural selection for enhanced tuberculosis resistance? *Med Hypotheses*, 62(5):839-43.
- Mooney S. (2005). Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. *Brief Bioinform*, 6(1):44-56.
- Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, et al. (2008). Identifying autism loci and genes by tracing recent shared ancestry. *Science*, 321(5886):218-23.
- Müller B, Grossniklaus U. (2010). Model organisms- a historical perspective. *J Proteomics*, 73(11):2054-63.
- Müller N, Schwarz MJ. (2010). Immune system and schizophrenia. *Curr Immunol Rev*, 6(3):213-220.
- Müller N, Riedel M, Gruber R, Ackenheil M, Schwarz MJ. (2000). The immune system and schizophrenia. An integrative view. *Ann NY Acad Sci*, 917:456-67.
- Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, et al. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science*, 309(5734):613-7.
- Nadeau JH, Taylor BA. (1984). Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci USA*, 81(3):814-8.
- Nadeau JH. (2001). Modifier genes in mice and humans. *Nat Rev Genet*, 2(3):165-74.
- Nadeau JH. (2003). Modifier genes and protective alleles in humans and mice. *Curr Opin Genet Dev*, 13(3):290-5.

- Nagaraj SH, Ingham A, Reverter A. (2010). The interplay between evolution, regulation and tissue specificity in the human hereditary diseasome. *BMC Genomics*, 11(Suppl 4):S23.
- Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. (2009). Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*, 324(5925):387-9.
- Ng PC, Henikoff S. (2006). Predicting the effects of amino acid substitutions on protein function (2006). *Annu Rev Genomics Hum Genet*, 7:61-80.
- Ng PC, Henikoff S. (2002). Accounting for human polymorphisms predicted to affect protein function. *Genome Res*, 12(3):436-46.
- Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, et al. (2008). Genetic variation in an individual human exome. *PLoS Genet*, 4(8):e1000160.
- Nibbe RK, Markowitz S, Myeroff L, Ewing R, Chance MR. (2009). Discovery and scoring of protein interaction subnetworks discriminative of late stage human colon cancer. *Mol Cell Proteomics*, 8(4):827-45.
- Nusbacher J, Hirschhorn K, Cooper LZ. (1967). Chromosomal abnormalities in congenital rubella. *N Engl J Med*, 276(25):1409-13.
- Ober C, Vercelli D. (2011). Gene-environment interactions in human disease: nuisance or opportunity? *Trends Genet*, 27(3):107-15.
- Ohno S. (1970). *Evolution by gene duplication*. Springer Verlag, New York, NY, USA.
- Oldridge M, Zackai EH, McDonald-McGinn DM, Iseki S, Morriss-Kay GM, et al. (1999). De novo Alu-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. *Am J Hum Genet*, 64(2):446-61.
- Oliver S. (2000). Guilt-by-association goes global. *Nature*, 403(6770):601-3.
- Oostra BA, Willemsen R. (2009). FMR1: a gene with three faces. *Biochim Biophys Acta*, 1790(6):467-77.
- Oti M, Ballouz S, Wouters MA. (2011). Web tools for the prioritization of candidate disease genes. *Methods Mol Biol*, 760:189-206.
- Ou Z, Stankiewicz P, Xia Z, Breman AM, Dawson B, et al. (2011). Observation and prediction of recurrent human translocations mediated by NAHR between nonhomologous chromosomes. *Genome Res*, 21(1):33-46.
- Owen MJ, Craddock N, O'Donovan MC. (2005). Schizophrenia: genes at last? *Trends Genet*, 21(9):518-25.
- Pagani F, Raponi M, Baralle FE. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc Natl Acad Sci USA*, 102(18):6368-72.
- Patterson PH. (2011). Modeling autistic features in animals. *Pediatr Res*, 69(5 Pt 2):34R-40R.
- Pearson BL, Pobbe RL, Defensor EB, Oasay L, Bolivar VJ, et al. (2011). Motor and cognitive stereotypies in the BTBR T+tf/J mouse model of autism. *Genes Brain Behav*, 10(2):228-35.
- Peng Q, Pevzner PA, Tesler G. (2006). The fragile breakage versus random breakage models of chromosome evolution. *PLoS Comput Biol*, 2(2):e14.
- Peltonen L, Perola M, Naukkarinen J, Palotie A. (2006). Lessons from studying monogenic disease for common disease. *Hum Mol Genet*, 15(Spec No 1):R67-74.
- Peñagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, et al. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell*, 147(1):235-46.
- Penzes P, Cahill ME, Jones KA, VanLeeuwen JE, Woolfrey KM. (2011). Dendritic spine pathology in neuropsychiatric disorders. *Nat Neurosci*, 14(3):285-93.



- Pevzner P, Tesler G. (2003). Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. *Genome Res*, 13(1):37-45.
- Phillips DH, Arlt VM. (2009). Genotoxicity: damage to DNA and its consequences. *EXS*, 99:87-110.
- Pierri CL, Parisi G, Porcelli V. (2010). Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening. *Biochim Biophys Acta*, 1804(9):1695-712.
- Pocklington AJ, Cumiskey M, Armstrong JD, Grant SG. (2006). The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Mol Syst Biol*, 2:2006.0023.
- Pomerantz MM, Shrestha Y, Flavin RJ, Regan MM, Penney KL, et al. (2010). Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. *PLoS Genet*, 6(11):e1001204.
- Ponder BA. (2001). Cancer genetics. *Nature*, 411(6835):336-41.
- Powell SB. (2010). Models of neurodevelopmental abnormalities in schizophrenia. *Curr Top Behav Neurosci*, 4:435-81.
- Qiu J, Noble WS. (2008). Predicting co-complexed protein pairs from heterogeneous data. *PLoS Comput Biol*, 4(4):e1000054.
- Reser JE. (2011). Conceptualizing the autism spectrum in terms of natural selection and behavioral ecology: the solitary forager hypothesis. *Evol Psychol*, 9(2): 207-238, Retrieved from <http://www.epjournal.net/filestore/EP09207238.pdf>
- Richards AJ, Laidlaw M, Meredith SP, Shankar P, Poulson AV, et al. (2007). Missense and silent mutations in COL2A1 result in Stickler syndrome but via different molecular mechanisms. *Hum Mutat*, 28(6):639.
- Roberts R, Wells GA, Stewart AF, Dandona S, Chen L. (2010). The genome-wide association study- a new era for common polygenic disorders. *J Cardiovasc Transl Res*, 3(3):173-82.
- Robertson HR, Feng G. (2011). Annual Research Review: Transgenic mouse models of childhood-onset psychiatric disorders. *J Child Psychol Psychiatry*, 52(4):442-75.
- Robinson EB, Koenen KC, McCormick MC, Munir K, Hallett V, et al. (2011). A multivariate twin study of autistic traits in 12-year-olds: Testing the fractionable autism triad hypothesis. *Behav Genet*, 42(2):245-55.
- Rodríguez-Escudero I, Oliver MD, Andrés-Pons A, Molina M, Cid VJ, et al. (2011). A comprehensive functional analysis of PTEN mutations: implications in tumor- and autism-related syndromes. *Hum Mol Genet*, 20(21):4132-42.
- Rogozin IB, Pavlov YI. (2003). Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res*, 544(1):65-85.
- Ronald A, Hoekstra RA. (2011). Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am J Med Genet B Neuropsychiatr Genet*, 156B(3):255-74.
- Ropers HH. (2007). New perspectives for the elucidation of genetic disorders. *Am J Hum Genet*, 81(2):199-207.
- Rothschild BM, Woods RJ, Rothschild C, Sebes JI. (1992). Geographic distribution of rheumatoid arthritis in ancient North America: implications for pathogenesis. *Semin Arthritis Rheum*, 22(3):181-7.
- Rubbi C, Milner J. (2005). p53: Gatekeeper, caretaker, or both? In: *25 Years of p53 Research*, P Hainaut & KG Wiman, (Eds.), Springer, London, UK.

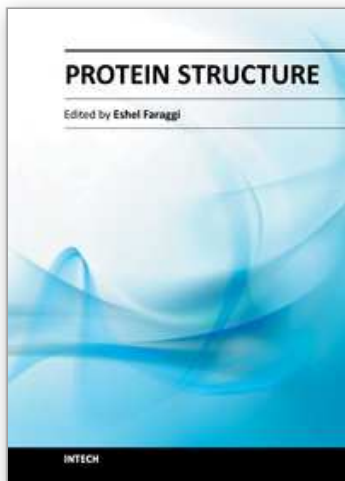
- Russo A, Migliavacca M, Zanna I, Macaluso M, Gebbia N, et al. (2006). Caretakers and Gatekeepers. In: *Encyclopedia of Life Sciences*, John Wiley & Sons, Retrieved from <http://onlinelibrary.wiley.com/doi/10.1038/npg.els.0006048/pdf>
- Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, et al. (2006). Positive natural selection in the human lineage. *Science*, 312(5780):1614-20.
- Saeed R, Deane CM. (2006). Protein protein interactions, evolutionary rate, abundance and age. *BMC Bioinformatics*, 7:128.
- Sahoo T, Theisen A, Rosenfeld JA, Lamb AN, Ravnan JB, et al. (2011). Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. *Genet Med*, 13(10):868-80.
- Sankoff D, Deneault M, Turbis P, Allen C. (2002). Chromosomal distributions of breakpoints in cancer, infertility, and evolution. *Theor Popul Biol*, 61(4):497-501.
- Sasaki MS. (2006). Delayed manifestation and transmission bias of de novo chromosome mutations: their relevance for radiation health effect. *J Radiat Res*, 47(Suppl B):B45-56.
- Schibler L, Roig A, Mahe MF, Laurent P, Hayes H, et al. (2006). High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. *BMC Genomics*, 7:194.
- Schmidt S, Gerasimova A, Kondrashov FA, Adzhubei IA, Kondrashov AS, et al. (2008). Hypermutable non-synonymous sites are under stronger negative selection. *PLoS Genet*, 4(11):e1000281.
- Shaffer LG, Bejjani BA, Torchia B, Kirkpatrick S, Coppinger J, et al. (2007). The identification of microdeletion syndromes and other chromosome abnormalities: cytogenetic methods of the past, new technologies for the future. *Am J Med Genet C Semin Med Genet*, 145C(4):335-45.
- Shah A, Garzon-Muvdi T, Mahajan R, Duenas VJ, Quiñones-Hinojosa A. (2010). Animal models of neurological disease. *Adv Exp Med Biol*, 671:23-40.
- Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, et al. (2006). Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet*, 38(9):1038-42.
- Shatsky M, Nussinov R, Wolfson HJ. (2008). Algorithms for multiple protein structure alignment and structure-derived multiple sequence alignment. *Methods Mol Biol*, 413:125-46.
- Sherva R, Farrer LA. (2011). Power and pitfalls of the genome-wide association study approach to identify genes for Alzheimer's disease. *Curr Psychiatry Rep*, 13(2):138-46.
- Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, et al. (2010). Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet*, 47(5):332-41.
- Shineman DW, Basi GS, Bizon JL, Colton CA, Greenberg BD, et al. (2011). Accelerating drug discovery for Alzheimer's disease: best practices for preclinical animal studies. *Alzheimers Res Ther*, 3(5):28.
- Shulman JM, De Jager PL, Feany MB. (2011). Parkinson's disease: genetics and pathogenesis. *Annu Rev Pathol*, 6:193-222.
- Sidransky E. (2006). Heterozygosity for a Mendelian disorder as a risk factor for complex disease. *Clin Genet*, 70(4):275-82.
- Silva AJ, Ehninger D. (2009). Adult reversal of cognitive phenotypes in neurodevelopmental disorders. *J Neurodev Disord*, 1(2):150-7.

- Simpson AJ, Camargo AA. (1998). Evolution and the inevitability of human cancer. *Semin Cancer Biol*, 8(6):439-45.
- Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. (2009). Autoimmune disease classification by inverse association with SNP alleles. *PLoS Genet*, 5(12):e1000792.
- Skolnick J, Brylinski M. (2009). FINDSITE: a combined evolution/structure-based approach to protein function prediction. *Brief Bioinform*, 10(4):378-91.
- Slater R, Bishop NE. (2006). Genetic structure and evolution of the Vps25 family, a yeast ESCRT-II component. *BMC Evol Biol*, 6:59.
- Slavotinek AM. (2008). Novel microdeletion syndromes detected by chromosome microarrays. *Hum Genet*, 124(1):1-17.
- Sleator RD, Walsh P. (2010). An overview of in silico protein function prediction. *Arch Microbiol*, 192(3):151-5.
- Smith NG, Eyre-Walker A. (2003). Human disease genes: patterns and predictions. *Gene*, 318:169-175.
- Smith CL, Bolton A, Nguyen G. (2010). Genomic and epigenomic instability, fragile sites, schizophrenia and autism. *Curr Genomics*, 11(6):447-69.
- Smirlis D, Soteriadou K. (2011). Trypanosomatid apoptosis: 'apoptosis' without the canonical regulators. *Virulence*, 2(3):253-6.
- Spradling A, Ganetsky B, Hieter P, Johnston M, Olson M, et al. (2006). New roles for model genetic organisms in understanding and treating human disease: report from the 2006 Genetics Society of America meeting. *Genetics*, 172(4):2025-32.
- Sriram G, Martinez JA, McCabe ER, Liao JC, Dipple KM. (2005). Single-gene disorders: what role could moonlighting enzymes play? *Am J Hum Genet*, 76(6):911-24.
- Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, et al. (2010). The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature*, 466(7307):720-6.
- Stankiewicz P, Lupski JR. (2002). Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev*, 12(3):312-9.
- Stankiewicz P, Lupski JR. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med*, 61:437-55.
- Strachan T, Read AP. (2010). *Human Molecular Genetics* (4th ed.), Garland Science, London, UK.
- Stratton MR, Campbell PJ, Futreal PA. (2009). The cancer genome. *Nature*, 458(7239):719-24.
- Sullivan PF, Kendler KS, Neale MC. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*, 60(12):1187-92.
- Sullivan P. (2011). Don't give up on GWAS. *Mol Psychiatry*, 17(1):203.
- Sun J, Jia P, Fanous AH, van den Oord E, Chen X, et al. (2010). Schizophrenia gene networks and pathways and their applications for novel candidate gene selection. *PLoS One*, 5(6):e11351.
- Tabarés-Seisdedos R, Dumont N, Baudot A, Valderas JM, Climent J, et al. (2011). No paradox, no progress: inverse cancer comorbidity in people with other complex diseases. *Lancet Oncol*, 12(6):604-8.
- Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, et al. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet*, 41(5):535-43.
- Tassone F, De Rubeis S, Carosi C, La Fata G, Serpa G, et al. (2011). Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Res*, 39(14):6172-85.

- Tayebati SK. (2006). Animal models of cognitive dysfunction. *Mech Ageing Dev*, 127(2):100-8.
- Theisen A, Schaffer LG. (2010) Disorders caused by chromosome abnormalities. *App Clin Genet* 3, 159-174.
- Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, et al. (2003). PANTHER: a browsable database of gene products. *Nucleic Acids Res*, 31(1):334-41.
- Thomas PD, Kejariwal A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci USA*, 101(43):15398-403.
- Thornton JW, DeSalle R. (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annu Rev Genomics Hum Genet*, 1:41-73.
- Tiffin N, Adie E, Turner F, Brunner HG, van Driel MA. (2006). Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res*, 34(10):3067-81.
- Todd AE, Oregano CA, Thornton JM. (2001). Evolution of function in protein superfamilies. *J Mol Biol*, 307(4):1113-43.
- Tordjman S, Drapier D, Bonnot O, Graignic R, Fortes S, et al. (2007). Animal models relevant to schizophrenia and autism: validity and limitations. *Behav Genet*, 37(1):61-78.
- Torkamani A, Dean B, Schork NJ, Thomas EA. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res*, 20(4):403-12.
- Touw IP, Erkeland SJ. (2007). Retroviral insertion mutagenesis in mice as a comparative oncogenomics tool to identify disease genes in human leukemia. *Mol Ther*, 15(1):13-9.
- Tranchevent LC, Capdevila FB, Nitsch D, De Moor B, De Causmaecker P, et al. (2011). A guide to web tools to prioritize candidate genes. *Brief Bioinform*, 12(1):22-32.
- Tu Z, Wang L, Xu M, Zhou X, Chen T, et al. (2006). Further understanding human disease genes by comparing with housekeeping genes and other genes. *BMC Genomics*, 7:31.
- Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. (2010). Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta*, 1797(2):113-28.
- Ule J, Darnell RB. (2007). Functional and mechanistic insights from genome-wide studies of splicing regulation in the brain. *Adv Exp Med Biol*, 623:148-60.
- Uversky VN. (2009). Intrinsic disorder in proteins associated with neurodegenerative diseases. *Front Biosci*, 14: 5188-5238.
- Uversky VN, Oldfield CJ, Dunker AK. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*, 37:215-46.
- Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, et al. (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, 10(Suppl 1):S7.
- Vigouroux C, Bonne G (2002). Laminopathies: one Gene, two proteins, five diseases, In: *Madame Curie Bioscience Database*. Landes Bioscience, Austin, Texas, USA. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK6151>
- Vijaya-Lakshmi AN, Ramana MV, Vijayashree B, Ahuja YR, Sharma G. (1999). Detection of influenza virus induced DNA damage by comet assay. *Mutat Res*, 442(1):53-8.
- Visscher PM, Hill WG, Wray NR. (2008). Heritability in the genomics era. *Nat Rev Genet*, 9(4):255-66.
- Visscher PM, Goddard ME, Derks EM, Wray NR. (2011). Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry*, (Jun 14) in press.



- Vockley J, Rinaldo P, Bennett MJ, Matern D, Vladutiu GD. (2000). Synergistic heterozygosity: disease resulting from multiple partial defects in one or more metabolic pathways. *Mol Genet Metab*, 71(1-2):10-8.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. (2006). A map of recent positive selection in the human genome. *PLoS Biol*, 4(3):e72.
- Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351):380-4.
- Walker MG, Volkmut W, Klingler TM. (1999). Pharmaceutical target discovery using Guilt-by-Association: schizophrenia and Parkinson's disease genes. *Proc Int Conf Intell Syst Mol Biol*, 1999:282-6.
- Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, et al. (1991). A de novo Alu insertion results in neurofibromatosis type 1. *Nature*, 353(6347):864-6.
- Wang J, Cao Z, Zhao L, Li S. (2011). Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). *Int J Mol Sci*, 12(5):3205-19.
- Ward AJ, Cooper TA. (2010). The pathobiology of splicing. *J Pathol*, 220(2):152-63.
- Waring SC, Rosenberg RN. (2008). Genome-wide association studies in Alzheimer disease. *Arch Neurol*, 65(3):329-34.
- Watson JD, Laskowski RA, Thornton JM. (2005). Predicting protein function from sequence and structural data. *Curr Opin Struct Biol*, 15(3):275-84.
- Westerlund M, Hoffer B, Olson L. (2010). Parkinson's disease: Exit toxins, enter genetics. *Prog Neurobiol*, 90(2):146-56.
- Williams RS, Boeckeler K, Gräf R, Müller-Taubenberger A, Li Z, et al. (2006). Towards a molecular understanding of human diseases using *Dictyostelium discoideum*. *Trends Mol Med*, 12(9):415-24.
- Williams TN. (2006). Red blood cell defects and malaria. *Mol Biochem Parasitol*, 149(2):121-7.
- Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. (2011). Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiol Aging*, 32(10):1923.e1-8.
- World Health Organization (1993). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*, World Health Organization, Geneva, Switzerland.
- Wray NR, Goddard ME, Visscher PM. (2008). Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev*, 18(3):257-63.
- Yang F, Babak T, Shendure J, Disteché CM. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res*, 20(5):614-22.
- Yatsenko SA, Brundage EK, Roney EK, Cheung SW, Chinault AC, et al. (2009). Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3 microdeletion syndrome. *Hum Mol Genet*, 18(11):1924-36.
- Yook SH, Oltvai ZN, Barabási AL. (2004). Functional and topological characterization of protein interaction networks. *Proteomics*, 4(4):928-42.
- Young JW, Zhou X, Geyer MA. (2010). Animal models of schizophrenia. *Curr Top Behav Neurosci*, 4:391-433.
- Zhu J, Xiao H, Shen X, Wang J, Zou J, et al. (2010). Viewing cancer genes from co-evolving gene modules. *Bioinformatics*, 26(7):919-24.
- Zwijnenburg PJ, Meijers-Heijboer H, Boomsma DI. (2010). Identical but not the same: the value of discordant monozygotic twins in genetic research. *Am J Med Genet B Neuropsychiatr Genet*, 153B(6):1134-49.



## **Protein Structure**

Edited by Dr. Eshel Faraggi

ISBN 978-953-51-0555-8

Hard cover, 396 pages

**Publisher** InTech

**Published online** 20, April, 2012

**Published in print edition** April, 2012

Since the dawn of recorded history, and probably even before, men and women have been grasping at the mechanisms by which they themselves exist. Only relatively recently, did this grasp yield anything of substance, and only within the last several decades did the proteins play a pivotal role in this existence. In this expose on the topic of protein structure some of the current issues in this scientific field are discussed. The aim is that a non-expert can gain some appreciation for the intricacies involved, and in the current state of affairs. The expert meanwhile, we hope, can gain a deeper understanding of the topic.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Azhari Aziz, Jasmina Ilievska, Paul R. Fisher and Naomi E. Bishop (2012). An Evolutionary Biology Approach to Understanding Neurological Disorders, Protein Structure, Dr. Eshel Faraggi (Ed.), ISBN: 978-953-51-0555-8, InTech, Available from: <http://www.intechopen.com/books/protein-structure/an-evolutionary-biology-approach-to-understanding-complex-human-genetic-disease>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen