Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease

David N. Cooper · Michael Krawczak · Constantín Polychronakos · Chris Tyler-Smith · Hildegard Kehrer-Sawatzki

Abstract Some individuals with a particular disease-causing mutation or genotype fail to express most if not all features of the disease in question, a phenomenon that is known as ‘reduced (or incomplete) penetrance’. Reduced penetrance is not uncommon; indeed, there are many known examples of ‘disease-causing mutations’ that fail to cause disease in at least a proportion of the individuals who carry them. Reduced penetrance may therefore explain not only why genetic diseases are occasionally transmitted through unaffected parents, but also why healthy individuals can harbour quite large numbers of potentially disadvantageous variants in their genomes without suffering any obvious ill effects. Reduced penetrance can be a function of the specific mutation(s) involved or of allele dosage. It may also result from differential allelic expression, copy number variation or the modulating influence of additional genetic variants in cis or in trans. The penetrance of some pathogenic genotypes is known to be age- and/or sex-dependent. Variable penetrance may also reflect the action of unlinked modifier genes, epigenetic changes or environmental factors. At least in some cases, complete penetrance appears to require the presence of one or more genetic variants at other loci. In this review, we summarize the evidence for reduced penetrance being a widespread phenomenon in human genetics and explore some of the molecular mechanisms that may help to explain this enigmatic characteristic of human inherited disease.

Introduction

One old conundrum in human genetics is that not everyone with a given pathological mutation (or mutations) will eventually develop the disease in question. The proportion of those individuals harbouring a particular pathogenic mutation or genotype who exhibit clinical signs of the associated disorder within a specific and clearly defined time period is termed the penetrance of that disorder. If this proportion equals 100 %, the disease and/or disease genotype(s) are said to show complete penetrance. If not, they are said to exhibit reduced (or incomplete) penetrance. Reduced penetrance is likely to be a consequence of the combination of a variety of different genetic and environmental factors. A classic example is phenylketonuria, where inactivating mutations in the PAH gene encoding the enzyme phenylalanine hydroxylase lead to severe intellectual disability in the context of a normal diet, whereas a life-long phenylalanine-restricted diet makes possible a relatively healthy life (Blau et al. 2010). Regrettably, few
other examples of reduced penetrance are as simple, well understood or clinically manipulable as this one, and the precise mechanisms by which different factors give rise to reduced penetrance remain largely unknown (Zlotogora 2003).

In formal terms, penetrance measures the proportion of individuals in a given population with a specific disease-associated genotype who also express the corresponding disease phenotype. Large family/population studies have traditionally been considered necessary to measure penetrance, either for specific mutations/genotypes or as a disease average. ‘Cascade genetic screening’, whereby relatives of previously identified carriers are screened both for the mutation(s) in question and for the presence/absence of clinical symptoms, is one means to determine the degree of penetrance for a given genotype (Berge et al. 2008). As we discuss below, large-scale sequencing and genotyping studies of apparently healthy individuals from the general population provide a powerful new approach to understanding the penetrance of pathological mutations/genotypes. The outcomes of such studies should allow us to predict how likely it is that a given disease will manifest itself in an individual who carries a specific genotype.

Reduced penetrance is most obviously evident in disorders that follow an autosomal dominant mode of inheritance. In these instances, reduced penetrance is a characteristic of the underlying mutation, rather than a genotype. However, reduced penetrance can also occur in autosomal recessive disorders where one and the same mutation can have different phenotypic effects, depending at least in part upon the second disease allele present. Irrespective of the mode of inheritance, in most cases penetrance is likely to be a function of the specific mutation(s) involved. Thus, in some conditions normally characterized by an autosomal dominant mode of inheritance, two incompletely penetrant (or otherwise non-penetrant) alleles may act in recessive fashion while mimicking the normal dominant form of the disease (e.g. Grundy et al. 1991; Croxen et al. 2002; Kowalewski et al. 2007; Castaman et al. 2007; Rossetti et al. 2009; Vujic et al. 2010; Schaaf et al. 2011a). For a dominantly inherited condition, one consequence of reduced penetrance is that the clinical phenotype may not be evident in one generation, but can nevertheless still be transmitted (through an apparently unaffected parent) to subsequent generations where it again manifests itself; specimen examples of this from clinical practice include hereditary hyperekplexia (Kwok et al. 2001), cherubism (Preda et al. 2010), retinitis pigmentosa (Rí o Frio et al. 2009), rhabdoid tumour predisposition syndrome (Ammerlaan et al. 2008), autism spectrum disorder (Fujita-Jimbo et al. 2012) and hypercholesterolaemia (García–García et al. 2011). For all the above reasons, reduced penetrance presents a major challenge to genetic counsellors attempting to interpret the medical history of a patient’s family to quantify the disease risk to the patient’s offspring (Emery 1986; Otto and Maestrelli 2000).

Reduced penetrance is not uncommon; indeed, there are many known examples of bona fide disease-causing variants or genotypes that fail to cause disease in at least a proportion of individuals who carry them (Zlotogora 2003; Waalen and Beutler 2009). By definition, penetrance refers to the black and white issue of whether the clinical phenotype associated with a certain genotype is present or not. We routinely distinguish it from variable expressivity which refers to the degree of variation of the clinical phenotype in those individuals with a particular genotype. Although, in principle, penetrance and expressivity are distinct terms with specific meanings (depending upon the way a given clinical phenotype is defined), in practice they are closely inter-related and likely to manifest via similar mechanisms. We also distinguish reduced penetrance from small effect size. Thus, most carriers of the risk alleles discovered by genome-wide association studies (GWAS) may never develop the disease in question; this is because these variants generally only make a small contribution to the multifactorial aetiology of the condition. To be able instead to say that the variant is non-penetrant in some individuals, we require the same variant in other individual(s) to make the crucial difference between the phenotype being manifested or not. In what follows, we shall focus specifically on the molecular mechanisms that could account for the phenomenon of reduced penetrance. This notwithstanding, the discussion of genetic variants that modulate the expressivity of a particular disease has sometimes also been deemed appropriate.

In this review, we present the evidence for reduced penetrance being a widespread phenomenon in human genetics, evidence that comes not only from a plethora of case studies of monogenic disorders but also more recently from the next generation sequencing of entire exomes or genomes of apparently normal healthy individuals from the general population. Here, we have attempted to explore the individual genetic components that contribute to the complexity of primarily monogenic disorders with reduced penetrance. This notwithstanding, the action of modifier genes is one of the mechanisms responsible for reduced penetrance, and one that has increasingly become recognized as blurring the distinction between monogenic conditions and complex disease (Nadeau 2003; Badano and Katsanis 2003; Sidransky 2006). Finally, we explore some of the molecular mechanisms which could account for the reduced penetrance of many human inherited disorders and provide evidence to support the view that, at least in some instances, complete penetrance requires the presence of one or more genetic variants at other loci.
Deleterious and disease alleles in the general population

In the wake of the sequencing of multiple human genomes, it has become apparent that healthy individuals can harbour quite large numbers of potentially disadvantageous variants without suffering any obvious ill effects (The 1000 Genomes Project Consortium 2010; MacArthur et al. 2012; Xue et al. 2012; Shen et al. 2013a). The underlying reasons are likely to be many and varied: thus, the variants may damage the protein in question but the intact protein may not be necessary for the health of the carrier; individuals may be asymptomatic carriers of single-mutant alleles that could, in homozygosity or compound heterozygosity, cause recessive disease; the mutation may be dominant but the clinical phenotype might only be mild and classed as lying within the range of normal healthy variation; the disorder might be late in onset with expression being age- or sex-dependent; or the disorder may require additional genetic and/or environmental factors for it to manifest clinically.

Assessing the magnitude of the ‘genetic burden’ imposed on the general population by the presence of deleterious alleles has been a key aim of medical and population genetics for many decades. With the advent of large-scale sequencing technologies, it has become possible to estimate the number of amino acid substitutions in the human exome that would be predicted in silico to be damaging (Kryukov et al. 2007; Lohmueller et al. 2008; Boyko et al. 2008; Goode et al. 2010). However, personal genome sequences have not only provided estimates of the number of disease variants carried by each subject (Asan et al. 2011), but have also given us a glimpse of the likely complexity of the functional interpretation of such data (Ashley et al. 2010; Strom and Gorin 2013). More recently, Tennessen et al. (2012) suggested that 2.3% of the 13,595 single nucleotide variants carried by the average person impact upon protein function, involving ~313 genes per human genome. Adopting an alternative approach, Bell et al. (2011) surveyed 437 genes known to be related to a recessive Mendelian disease and identified 2.8 mutations per individual (range 0–7). Taken together, these studies suggest that individuals typically carry hundreds of mildly disadvantageous variants and perhaps several tens of potentially severe disease alleles.

In a pilot study for the 1000 Genomes Project Consortium (2010), we reported the prevalence of disease alleles, defined by reference to the disease-causing mutations (DMs) listed in the Human Gene Mutation Database (HGMD: http://www.hgmd.org; Stenson et al. 2009), in population samples of African, European and East Asian origin. These numbers were surprisingly high: 57–80 disease alleles per individual in a sample of 179 participants. Moreover, further examination of these data showed that 191 disease alleles were present in the homozygous state in at least one individual, and hence were not simply present because their phenotypic-clinical effects had been masked by a normal allele. Although little phenotypic information other than sex, ethnicity, place of origin and relationship to other participants is available for the 1000 Genomes Project donors, the Project’s ethical framework requires that sample donors are non-vulnerable adults who are competent to consent to participation in the project and hence are likely to lack any obvious severe disease phenotype. Instead, we need to seek some other explanation for the high number of DMs present. One possibility is that the penetrance of disease alleles and genotypes could be much lower and more variable than previously realized. Conventionally, most studies of human inherited disease that have contributed mutation data to HGMD have sought to identify a disease genotype, given a clinical phenotype (Cooper et al. 2010). This is a very different scenario from identifying a phenotype given a (potentially disease-associated) genotype. Indeed, the former strategy inevitably avoids the whole issue of penetrance because, by definition, it focuses exclusively upon those individuals in whom the mutation of interest has been penetrant. It follows that a given variant could be genuinely causative in a set of individuals manifesting a particular disease, yet may also be present in a set of healthy individuals who differ in any one of a number of ways to be discussed below.

Since the 1000 Genomes Project Consortium paper was published in 2010, two further reports have appeared that served to improve our knowledge of deleterious mutations in the genomes of apparently healthy individuals. The most readily recognized deleterious variants in the human genome are those that disrupt a protein-coding gene, either by leading to a loss of function (e.g. a nonsense or frameshift variant) or by altering an amino acid in the encoded protein (missense variants). The former category of mutation has been studied by MacArthur et al. (2012) who identified 1,285 putative loss-of-function variants (i.e. nonsense mutations, splice site-disrupting single nucleotide variants, micro-insertions/micro-deletions, etc.) in the genome sequences of 185 humans from the 1000 Genomes Project. From these data, they estimated that an average human genome typically contains ~100 genuine loss-of-function variants with ~20 genes having both copies inactivated. Following on from this study, Xue et al. (2012) focused on missense mutations, ascertaining the numbers of potentially deleterious missense variants in the genomes of apparently healthy individuals using low-coverage whole-genome sequence data from 179 individuals in the 1000 Genomes Pilot Project. Each individual was found to carry 281–515 missense substitutions predicted with a high degree of confidence to be damaging to the gene product, 40–85 of which were present in the homozygous state. Taken together, these studies suggest that a typical healthy
individual has about 80 of their genes severely damaged or inactivated in both copies, further emphasizing the stark contrast between damage to gene and protein on the one hand, and damage to health on the other. The 1000 Genomes Project participants also carried 40–110 variants (3–24 homozygous) classified by HGMD as DMs. Whereas many of these DMs could conceivably represent disease attribution errors of some kind, between 0 and 8 DMs per individual (0–1 homozygous) were predicted to be highly damaging.

Among the missense DMs, Xue et al. (2012) identified known pathological variants such as HBB (c.20A>T; p.Glu7Val), which leads to increased resistance to malaria in heterozygotes but to sickle cell disease in homozygotes [confined to Africans (Yoruba, YRI) in whom there were 12 heterozygotes and 1 homozygote]. In addition, Xue et al. (2012) identified an USH2A variant (c.2138G>C; p.Gly713Arg), previously reported as being causal for Usher syndrome type 2, a recessive disorder characterized by combined deafness and blindness; three homozygotes were noted in the YRI. Manual curation of the HGMD-1000GP overlap revealed the presence of three types of DM: (1) plausible severe disease-causing variants, (2) variants convincingly causative for pathological conditions, yet quite compatible with adult life and (3) variants probably incorrectly assigned as disease causing. After applying various filtering criteria designed to enrich for true disease alleles, the list was reduced to 45 candidates (Xue et al. 2012). Of these putative disease alleles, 34 were present in the heterozygous state and were deemed likely to

<table>
<thead>
<tr>
<th>Disease (MIM number)</th>
<th>Inheritance</th>
<th>Gene</th>
<th>HGVS cDNA mutation</th>
<th>Protein alteration</th>
<th>Total homozygotes</th>
<th>Total heterozygotes</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ataxia telangiectasia (MIM# 607585)</td>
<td>AR</td>
<td>ATM</td>
<td>NM_000051.3: c.4258C&gt;T</td>
<td>p.Leu1420Phe</td>
<td>1</td>
<td>3</td>
<td>Low-penetrance breast cancer susceptibility allele</td>
</tr>
<tr>
<td>Usher syndrome type II (MIM# 276901)</td>
<td>AR</td>
<td>USH2A</td>
<td>NM_206933.2: c.2137G&gt;C</td>
<td>p.Gly713Arg</td>
<td>3</td>
<td>22</td>
<td>Probable complex pathogenicity; neutral in YRI?</td>
</tr>
<tr>
<td>Nephronophthisis 4 (MIM# 606966)</td>
<td>AR</td>
<td>NPHP4</td>
<td>NM_015102.3: c.2542C&gt;T</td>
<td>p.Arg848Trp</td>
<td>1</td>
<td>2</td>
<td>Growth retardation; adult-onset renal disease</td>
</tr>
<tr>
<td>Cushing syndrome (MIM# 607397)</td>
<td>AR</td>
<td>MC2R</td>
<td>NM_000529.2: c.833T&gt;G</td>
<td>p.Phe278Cys</td>
<td>1</td>
<td>10</td>
<td>Hormonal disorder; variable gender-specific symptoms; variant functionally defective in vitro</td>
</tr>
<tr>
<td>Low-phospholipid-associated cholelithiasis (MIM# 171060)</td>
<td>AR</td>
<td>ABCB4</td>
<td>NM_000443.3: c.2363G&gt;A</td>
<td>p.Arg788Gln</td>
<td>2</td>
<td>9</td>
<td>Adult onset</td>
</tr>
<tr>
<td>Cardiomyopathy, hypertrophic (MIM# 115197)</td>
<td>AD</td>
<td>MYBPC3</td>
<td>NM_000256.3: c.1519G&gt;A</td>
<td>p.Gly507Arg</td>
<td>0</td>
<td>2</td>
<td>Late onset; incomplete penetrance</td>
</tr>
<tr>
<td>Glaucoma, primary open angle (MIM# 609887)</td>
<td>AD</td>
<td>WDR36</td>
<td>NM_139281.2: c.1586G&gt;A</td>
<td>p.Arg529Gln</td>
<td>0</td>
<td>1</td>
<td>Adult onset; variant functionally defective in vitro</td>
</tr>
<tr>
<td>Colorectal cancer, nonpolyposis (MIM# 609310)</td>
<td>AD</td>
<td>MLH1</td>
<td>NM_000249.3: c.1742C&gt;T</td>
<td>p.Pro581Leu</td>
<td>0</td>
<td>1</td>
<td>Adult onset; variant functionally defective in vitro</td>
</tr>
<tr>
<td>Renal cell carcinoma (MIM# 144700)</td>
<td>AD</td>
<td>FLCN</td>
<td>NM_144997.5: c.715C&gt;T</td>
<td>p.Arg239Cys</td>
<td>0</td>
<td>1</td>
<td>Late onset</td>
</tr>
<tr>
<td>Heparin cofactor 2 deficiency (MIM# 612356)</td>
<td>AD</td>
<td>SERPIND1</td>
<td>NM_000185.3: c.623G&gt;A</td>
<td>p.Arg208His</td>
<td>0</td>
<td>2</td>
<td>Deficiency state, but no overt disease; risk factor for thrombophilia</td>
</tr>
<tr>
<td>Lp(a) deficiency (MIM# 152200)</td>
<td>AD</td>
<td>LPA</td>
<td>NM_005577.2: c.4289+1G&gt;A</td>
<td>essential splice site</td>
<td>0</td>
<td>5</td>
<td>Late onset; risk factor for heart disease</td>
</tr>
</tbody>
</table>
be present in asymptomatic carriers. Among the remaining 11 (Table 1), the 6 linked to dominant disorders were explicable in terms of either late onset (e.g. Gly507Arg in MYBPC3) or covert disease (Arg208His in SERPIND1). In similar vein, the presence of homozygotes for four of the five recessive disorders could be explicable in terms of late onset and/or reduced penetrance of disease. The USH2A mutation (Gly713Arg) was, however, intriguing: this variant was predicted to be damaging to the protein, and pathogenic in some populations but not in others (e.g. YRI). One explanation put forward to explain this apparent contradiction was that, in the YRI population, the USH2A locus is subject to copy number variation (Matsuzaki et al. 2009) that could provide functional complementation of the mutant gene. In the majority of cases, however, the most likely explanation for the absence of disease at the time of recruitment was considered to be the probable late onset of disease, although clinical penetrance was often variable, and some phenotypes, such as loose anagen hair syndrome [caused by Glu337Lys in KRT75 (MIM 600628)], might not even be regarded as “diseases” sensu stricto. These factors notwithstanding, the findings of Xue et al. (2012) suggest that incidental findings which are potentially relevant to health and well-being might be made in as many as 11 % of individuals sequenced.

Reduced penetrance is one of several possible explanations for why some variants of putative pathological significance, listed in HGMD and/or Locus-specific Mutation Databases, nevertheless occur in apparently healthy individuals (Ashley et al. 2010; Bell et al. 2011; Xue et al. 2012; Golbus et al. 2012; Wang et al. 2013a; Kenna et al. 2013; Shen et al. 2013a). It is not hard to see why reduced penetrance might be much more common among described mutations than originally thought: whereas known pathological mutations have almost invariably been identified through retrospective analyses of families or well-defined groups of clinically symptomatic patients, relatively few prospective studies of asymptomatic carriers have so far been performed to derive estimates of penetrance (e.g. Jensen et al. 2013; Mavaddat et al. 2013). Indeed, establishing that a specific mutation identified in a particular patient with a given disease is the pathological lesion responsible for that individual’s clinical phenotype does not allow one automatically to judge whether this mutant genotype will invariably give rise to the same clinical phenotype in all other individuals harbouring it. This can only be established (or refuted) empirically by comprehensive, ideally prospective, studies of the genotype in question.

One fairly obvious reason why comparing the output of genome sequencing projects (e.g. the 1000 Genomes Project) with a comprehensive database of putatively pathological mutations (e.g. HGMD) is likely to generate a considerable number of potentially pathogenic mutations in the general population, is that many such mutations are quite frequent in the population at large. In particular, carrier frequencies for mutations underlying recessive conditions can often be quite high. Thus, the ABCA4 Gly863Ala mutation causing Stargardt disease has a carrier frequency of 1.8 % in Europe (Maugeri et al. 2002), the GJB2 35delG mutation causing congenital deafness has a carrier frequency of 2.9 % in southern Europe (Gasparini et al. 2000), the ERCC8 Tyr322Ter mutation causing Cockayne syndrome has a carrier frequency of 6.8 % in Israeli Christian Arabs (Khayat et al. 2010) and the SPG7 Ala510Val mutation associated with adult-onset neurogenic disease has a carrier frequency of 3–4 % in the UK population (Roxburgh et al. 2013). Disease allele frequencies can be as high as 10 % in certain ethnic groups, e.g. Jews (Zlotogora et al. 2007; Ostrer and Skorecki 2013). A recent screen of the Korean population for 20 common mutations contributing to six autosomal recessive disorders yielded a combined carrier frequency of 6.7 % (Song et al. 2012). Screening an ethnically diverse US population sample (N = 364,890) for 87 different CFTR mutations responsible for causing cystic fibrosis yielded a combined carrier frequency of 2.6 % (Rohlf et al. 2011). As part of the NHLBI-Go Exome Sequencing Project, a screen of 5,400 individuals from the general population for variants in eight long QT syndrome genes yielded a total of 33 different missense mutations (affecting 173 alleles), representing a carrier frequency of 3.2 % (Refsgaard et al. 2012). A similar screen for cardiomyopathy-associated gene variants yielded combined carrier frequencies for mutations reported to be disease associated of 25 % (1,474/5,810) for hypertrophic cardiomyopathy, 15 % (963/6,334) for dilated cardiomyopathy and (22 %) 1,393/6,359 for arrhythmogenic right ventricular cardiomyopathy (Andreasen et al. 2013); the high number of detected cardiomyopathy-associated gene variants suggests, however, that many are only modest disease modifiers or even non-pathogenic. Nishiguchi and Rivolta (2012) screened 46 complete genome sequences from the general population for mutations in 106 genes associated with recessively inherited retinal degeneration and identified null mutations in ten individuals (22 %). Finally, a routine screen of 23,453 individuals for 417 pathogenic mutations associated with a total of 108 recessive diseases concluded that 24 % of individuals were carriers of at least one disorder, whilst 5.2 % were carriers of two or more disorders (Lazarin et al. 2013). Clearly, there is a veritable abundance of actual and potential pathological variants segregating in the general population.

The examples that we cite in the text that follows and in the accompanying tables are by no means comprehensive and have been provided simply to illustrate the many and varied mechanisms that are already known to underlie the phenomenon of reduced penetrance in relation to clinical disorders. However, these examples also demonstrate that
it is often very hard for authors, diagnosticians and mutation database curators alike to classify the pathogenicity of identified variants with any degree of certainty. Thus, is a given missense mutation a bona fide pathological lesion that exhibits greatly reduced penetrance, or is it essentially a neutral or near-neutral variant that occasionally finds itself in cis to other variants that are responsible for conferring the disease phenotype upon the individual concerned, or is it a variant with small effect that is sometimes present in affected individuals whose major causative lesions remain unknown?

### Incomplete penetrance in dominant and recessive conditions

A considerable number of autosomal dominant disorders are characterized by incomplete penetrance. Well-studied examples include the hair disease monilethrix (KRT86; De Cruz et al. 2012), congenital cataract (GJA3; Burdon et al. 2004), different types of retinitis pigmentosa (PRPF8 and PRPF31; Maubaret et al. 2011; Saini et al. 2012), LMNA mutation-associated muscular phenotypes (Rankin et al. 2008) and long QT syndrome (Giudicessi and Ackerman 2013; Mathias et al. 2013).

One rather well-understood example of incomplete penetrance of a dominantly inherited mutation is factor V Leiden (F5, Arg506Gln; Arg506Gln in legacy nomenclature; rs6025) which occurs at polymorphic frequencies (2–5 %) in European populations, but is associated with a sixfold increased risk of venous thrombosis and a two- to threefold increased risk of pregnancy loss (Kujovich 2011). Despite these very evident disease associations, the vast majority of factor V Leiden carriers appear to be clinically unaffected. This could help to account for the high frequency of this variant in the general population, together perhaps with the survival advantage conferred by factor V Leiden carriership in various other clinical contexts including severe sepsis (Kerlin et al. 2003; van Mens et al. 2013). In similar vein, the asymptomatic (clinically covert) state is much more common than the clinically overt state in several other dominant disorders of haemostasis, including protein C deficiency (Tait et al. 1995; McColl et al. 1996), protein S deficiency (Dykes et al. 2001), antithrombin deficiency (Tait et al. 1994; McColl et al. 1996) and von Willebrand disease (Rodeghiero et al. 1987; Castaman et al. 2003). Thus, in disorders of haemostasis and thrombosis, even with well-characterized variants that are known to confer a significantly increased disease risk, the clinical penetrance is often so low that more healthy individuals carry the variant than those who actually manifest disease.

Reduced or incomplete penetrance has also been described for autosomal recessive disorders. Probably the best characterized example of incomplete penetrance in a recessive disorder is provided by the Cys282Tyr (rs1800562) mutation in the haemochromatosis (HFE) gene (Beutler 2003). The Tyr282 homozygous genotype is present in approximately 1 in 200 people of Northern European origin and is responsible for 80–90 % of hereditary haemochromatosis (Weiss 2010; Rochette et al. 2010). Although Tyr282 homozygosity displays a relatively high biochemical penetrance (i.e. iron accumulation), its clinical penetrance is low (McCune et al. 2006). Available data suggest that 38–50 % of Tyr282 homozygotes develop iron overload and 10–25 % develop some type of haemochromatosis-associated morbidity (Whitlock et al. 2006). However, these statistics conceal what appears to be a gender effect: large-scale studies of newly diagnosed Tyr282 homozygotes, in whom liver disease had been specifically assessed, revealed that disease manifested in 24–43 % of males, but only 1–14 % of females (Rossi et al. 2008). Various genetic modifiers have been identified as influencing the clinical expression of haemochromatosis. These include mutations in the HAMP, HFE2 and TFR2 genes and polymorphisms in the BMP2, BMP4, CYBRD1, HP, LTA, MPO, TMRSS6 and TNF genes (Milet et al. 2007; Rochette et al. 2010; Valenti et al. 2012; Pelucchi et al. 2012). In addition, several environmental modifiers (e.g. diet, alcohol intake) are also known to affect the penetrance of the HFE genotype.

In another recessive disorder, Gaucher disease, the most common GBA mutation, Asn370Ser (Asn409Ser in HGVS nomenclature; rs76763715), is also characterized by low penetrance and exhibits extensive phenotype heterogeneity even in the homozygous state (Sibille et al. 1993; Horowitz et al. 1998; Fairley et al. 2008). However, close examination of asymptomatic Ser370 homozygotes, serendipitously diagnosed by prenatal carrier screening, revealed a variety of previously unidentified disease manifestations indicating that the clinical penetrance of this disease genotype may be greater than previously appreciated (Balwani et al. 2010).

### Influence of mutation type on penetrance

Clinical penetrance is in part a function of the mutation(s) in question. For a given disease, some causal mutations may exhibit complete clinical penetrance, whereas other mutations in the same gene show incomplete or even very low penetrance. Thus, whereas the penetrance of the most common CFTR gene lesion, ΔF508 (rs113993960), in cystic fibrosis is very high, the penetrance of the CFTR Arg117His (rs78655421) mutation (in any allele combination) appears to be so low as to call into question its putative role as a pathological mutation (Thauvin-Robinet et al. 2009).

As yet, relatively few studies have been performed on low-penetrance mutations with a view to identifying the features responsible at the molecular level for their low
penetrance. One exception is in retinoblastoma where it has been found that low-penetrance RBL mutations tend either to lead to a reduction in the amount of Rb protein produced (through promoter or splice site mutations) or yield a partially functional Rb molecule through missense mutation or in-frame deletion (Onadim et al. 1992; Kratzke et al. 1994; Bremner et al. 1997; Otterson et al. 1997; Scheffer et al. 2000; Genuardi et al. 2001; Harbour 2001; Klutz et al. 2002; Lefèvre et al. 2002; Valverde et al. 2005; Sánchez-Sánchez et al. 2005; Sampieri et al. 2006; Gámez-Pozo et al. 2007; Park et al. 2008; Hung et al. 2011). Of particular interest is the category of temperature-sensitive mutations in the Rb pocket domain (ΔAsn480, Arg661Trp, Cys712Arg) whose ‘reversible fluctuations’ in a threshold level of Rb pocket-binding activity could be responsible for their characteristic low penetrance (Otterman et al. 1999).

On the basis of studies performed to date, it would appear that a high proportion of RBL mutations with reduced penetrance are splice site mutations, although not all splice site mutations display low penetrance.

Some mutations are associated with specifically reduced penetrance as compared to other mutations of the same type in the same gene. For example, BRCA1 Arg1699Gln is characterized by a cumulative risk of breast or ovarian cancer by the age of 70 years of only 24 % (Spurdle et al. 2012), much lower than for the average pathogenic BRCA1 mutation (~71 %; van der Kolk et al. 2010). Individuals with GJB2-associated deafness who harbour two nonsense/truncating mutations exhibit a much more severe clinical phenotype, and hence are more likely to come to clinical attention, than those harbouring two missense mutations (Azaiez et al. 2004). This serves to illustrate that, in recessive disorders, the clinical penetrance of one mutation may be strongly influenced by the nature of the other mutation in trans. In Ehlers–Danlos syndrome type IV, however, null COL3A1 mutations tend to exhibit lower penetrance than missense and splicing mutations (Leistritz et al. 2011); this is presumably because a faulty gene product can disrupt the entire triple helical collagen molecule, whereas a null mutation merely reduces the amount of normal collagen produced (Arnold and Fertala 2013). In the same vein, patients with heritable pulmonary arterial hypertension, due to missense mutations in the BMPR2 gene, present earlier and with more severe disease than patients harbouring truncating mutations (Austin et al. 2009a). It is thought that the missense mutations are associated with stable BMPR2 transcripts encoding BMPR2 protein which exerts a dominant negative effect on BMP signalling, thereby rendering missense mutations more detrimental than truncating mutations. The majority of BMPR2 missense mutations were penetrant prior to the age of 36 years, whereas the majority of truncating mutations became penetrant only after the age of 36 years (Austin et al. 2009a). With mutations of the TNFRSF1A gene causing TNF receptor-associated periodic syndrome, missense mutations in cysteine residues have been reported to be more penetrant than missense mutations in non-cysteine residues (Aksentijevich et al. 2001; Aganna et al. 2001). Intriguingly, a SOD1 Leu117Val missense mutation which yields a mutant protein indistinguishable from wild-type SOD1 (in terms of its activity, stability and folding) causes amyotrophic lateral sclerosis, but with unusually low penetrance and slow progression (Synofzik et al. 2012). Finally, clinical penetrance may vary not only with the mutation type, but also with the location of the mutation in the gene/protein (Jackson et al. 1999; Risch et al. 2001; Yatsenko et al. 2001; van der Werf et al. 2012; Ho et al. 2012).

In autosomal dominant hereditary pancreatitis, the penetrance of the PRSS1 Arg122His mutation has been calculated to be 86 %, whereas that of PRSS1 Ala16Val is of the order of 55–65 % (Grocock et al. 2010; Joergensen et al. 2010). It is thought that the comparatively low penetrance of this latter mutation may be related to its particularly mild biochemical phenotype. The Ala16Val substitution alters the N-terminal residue of the trypsinogen activation peptide, thereby increasing the rate of N-terminal processing by chymotrypsin C by ~5.8-fold (Szabó and Sahin-Tóth 2012). Since the activation peptide is released during the activation process, the Ala16Val mutation is absent from active trypsin and hence cannot influence trypsin function.

Missense mutations in the MEFV gene responsible for familial Mediterranean fever can differ quite dramatically in terms of their clinical penetrance. For example, Met694Val is generally characterized by high penetrance, whereas both Glu148Gln and Val726Ala exhibit reduced penetrance (Shohat and Halpern 2011). In this disorder, the carrier frequency is higher than would be expected from the prevalence of the disease, suggesting that the penetrance of pathogenic MEFV mutations may often be incomplete in the compound heterozygous state (Gershoni-Baruch et al. 2002; Zaks et al. 2003; Caglayan et al. 2010; Camus et al. 2012; Soriano and Manna 2012).

Double missense mutations in cis are not infrequently encountered in patients with an inherited disease. One of the two mutations may represent a hypomorphic (i.e. less functional) allele, as for example with the GLA Asp313Tyr occurring in cis to the pathogenic Gly411Asp in patients with Fabry disease (Yasuda et al. 2003). Double missense mutations in cis may however be associated with a highly variable clinical phenotype (e.g. MEFV, Pro369Ser/Arg408Gln as a cause of familial Mediterranean fever; Ryan et al. 2010). A low-penetrance missense mutation may be associated with a particularly severe clinical phenotype when it occurs in cis with a second known pathogenic mutation, e.g. MYH7 Val606Met and Ala728Val in hypertrophic cardiomyopathy (Blair et al. 2001). Similarly,
two missense mutations in cis, each individually exerting a comparatively mild or no effect on the clinical phenotype, can act in concert leading to a more severe effect on the phenotype than either acting alone (e.g. CFTR Arg347His and Asp979Ala in cystic fibrosis; Clain et al. 2001 or RET Cys634Tyr and Tyr791Phe resulting in pheochromocytoma with high penetrance; Toledo et al. 2010). By contrast, Brugnoni et al. (2013) have intriguingly claimed that two different CLCN1 mutations do not give rise to myotonia congenita when they occur in cis on the same allele, although both lesions cause the disease when inherited on their own.

In diseases that exhibit locus heterogeneity, clinical penetrance may vary between mutations in different genes. For example, in pancreatitis, penetrance may vary from virtually 100 % in the case of the most common mutations with high penetrance manifesting as a later onset of clinical symp-
toms (Kokkos et al. 2004; Quarrell et al. 2004; Toledo et al. 2009). Another example of the modulatory effect of a missense polymorphism on disease allele penetrance is provided by

Modulating influence of additional allelic variants in cis or in trans

Some allelic variants may influence the expression of their host gene so as to alter the penetrance of a potentially pathological mutation in the same gene. Such modulatory variants may reside within exons, introns or regulatory regions. For example, the common Arg413Gln F7 polymorphism (Arg353Gln in legacy nomenclature; rs6046), which serves to reduce the level of secreted coagulation factor VII by ~25 % (Arbini et al. 1994; Hunault et al. 1997), is over-represented among individuals with clinically symptomatic factor VII deficiency (Millar et al. 2000). This is consistent with the view that Arg413Gln is a functional polymorphism and that the presence of Gln413 increases the likelihood that an individual, whose haemo-
static potential is already compromised by a heterozygous F7 mutation, will come to clinical attention due to a bleeding diathesis. The same principle applies to the functional Arg202Gln MEFV polymorphism (rs224222) where the Gln202 allele occurs in the homozygous state at a disproportionately higher frequency (15 %) in familial Mediterranean fever patients than in normal controls (2.7 %) (Yigit et al. 2012). In similar vein, the T allele of a functional C/T polymorphism (rs11024595) in the pro-
motor region of the SAA1 gene is significantly over-rep-
resented in familial Mediterranean fever patients as compared with normal controls (Migita et al. 2013). The MLH1 Lys618Ala mutation (AAG> GCC; rs35502531), initially supposed to be a benign polymorphism, has been found to be significantly over-represented in sporadic cancers associated with Lynch syndrome; MLH1 Ala618 appears to have a reduced ability to bind PMS2, one of the MLH1 protein’s mismatch repair partners (Medeiros et al. 2012). Finally, the functional KCNE1 Asp85Asn poly-
morphism (rs1805128), which occurs in the general pop-
ulation with a frequency of 0.8 %, occurs at a frequency of 3.9 % in long QT syndrome patients (Nishio et al. 2009).

Various reported examples of the modulation of the impact of pathogenic missense mutations by allelic single nucleotide polymorphisms (SNPs) are given in Table 2. Thus, the missense polymorphism Asp216His (rs1801968) in the TOR1A (DYT1) gene serves to moderate the clinical impact, in both cis and in trans, of the TOR1A c.904-906 del GAG mutation, the major mutation underlying early-onset dystonia (Kock et al. 2006; Risch et al. 2007; Martino et al. 2013). Similarly, in long QT syndrome, the genotype of a missense polymorphism (Lys897Thr; rs1805123) in the KCNH2 gene appears to distinguish symptomatic from asymptomatic individuals carrying a low-penetrance Ala1116Val pathogenic mutation (Crotti et al. 2005). Another example of the modulatory effect of a missense polymorphism on disease allele penetrance is provided by
Table 2  Examples of pathogenic microlesions whose penetrance has been found to be modulated by allelic SNPs

<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene</th>
<th>Pathological mutation</th>
<th>Modifying SNP</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor VII deficiency</td>
<td>F7</td>
<td>Various</td>
<td>NM_000131.3: c.1238G&gt;A Arg143Gln</td>
<td>Millar et al. (2000)</td>
</tr>
<tr>
<td>Brugada syndrome</td>
<td>SCN5A</td>
<td>NM_198056.2: c.5243G&gt;A Gly1748Asp</td>
<td>NM_198056.2: c.1673A&gt;G His558Arg (rs1805124)</td>
<td>Núñez et al. (2013)</td>
</tr>
<tr>
<td>Familial sick sinus syndrome</td>
<td>SCN5A</td>
<td>Various</td>
<td>NM_198056.2: c.1673A&gt;G His558Arg (rs1805124)</td>
<td>Gui et al. (2010)</td>
</tr>
<tr>
<td>Early-onset dystonia</td>
<td>TOR1A</td>
<td>NM_000113.2: c.907-909delGAG</td>
<td>NM_000113.2: c.646G&gt;C Asp216His (rs1801968)</td>
<td>Kock et al. (2006) and Kisch et al. (2007)</td>
</tr>
<tr>
<td>Familial Mediterranean fever</td>
<td>MEFV</td>
<td>Various</td>
<td>NM_000243.2: c.605G&gt;A Arg202Gln (rs224222)</td>
<td>Yigit et al. (2012)</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>MLH1</td>
<td>Various</td>
<td>NM_000249.3: c.1852_1853delAAinsGC Lys618 Ala</td>
<td>Medeiros et al. (2012)</td>
</tr>
<tr>
<td>Long QT syndrome</td>
<td>KCNQ2</td>
<td>NM_000238.3: c.3347C&gt;T Ala1116Val</td>
<td>NM_000238.3: c.2690A&gt;C Lys897Thr (rs1805123)</td>
<td>Croatti et al. (2005)</td>
</tr>
<tr>
<td>Medullary thyroid carcinoma</td>
<td>RET</td>
<td>NM_020975.4: c.196A&gt;G Gly666Glu</td>
<td>NM_020975.4: c.2071G&gt;A Gly691Asp (rs1799990)</td>
<td>Goldfarb et al. (1992) and Aperti et al. (2005)</td>
</tr>
<tr>
<td>Sudden unexplained death</td>
<td>SCN5A</td>
<td>NM_198056.2: c.2039G&gt;A Arg680His</td>
<td>NM_198056.2: c.3308G&gt;A Ser1103Tyr (rs7626962)</td>
<td>Cheng et al. (2011)</td>
</tr>
<tr>
<td>Medullary thyroid carcinoma</td>
<td>RET</td>
<td>NM_020975.4: c.1597G&gt;T Gly533Cys</td>
<td>NM_020975.4: c.74-126G&gt;T (rs2565206)</td>
<td>Tamanaha et al. (2009)</td>
</tr>
<tr>
<td>Homocystinuria</td>
<td>MTRR</td>
<td>NM_002454.2: c.166G&gt;A Val56Met</td>
<td>NM_002454.2: c.66G&gt;A Ile22Met (rs1801394)</td>
<td>Gherasim et al. (2007)</td>
</tr>
<tr>
<td>Childhood absence epilepsy</td>
<td>CACNA1H</td>
<td>NM_021098.2: c.2318G&gt;A Gly773Asp</td>
<td>NM_021098.2: c.2362C&gt;T Arg788Cys (rs3751664)</td>
<td>Vitko et al. (2005)</td>
</tr>
<tr>
<td>Primary hyperoxaluria type 1</td>
<td>AGXT</td>
<td>NM_000030.2: c.731T&gt;C Ile244Thr</td>
<td>NM_000030.2: c.32C&gt;T Pro111Leu (rs34116584)</td>
<td>Santana et al. (2003)</td>
</tr>
<tr>
<td>Erythropoietic protoporphyria</td>
<td>FECH</td>
<td>Various</td>
<td>NM_000140.3: c.315-48T&gt;C (rs2272783)</td>
<td>Gouya et al. (2006)</td>
</tr>
<tr>
<td>Autosomal dominant osteopetrosis type II</td>
<td>CLCN7</td>
<td>Various</td>
<td>NM_001257.4: c.1252G&gt;A Val418Met (rs12926089)</td>
<td>Chu et al. (2005)</td>
</tr>
<tr>
<td>Haemolytic uraemic syndrome</td>
<td>CFH</td>
<td>Various</td>
<td>NM_000186.3: c.2808G&gt;T Glu936Asp (rs1065489)</td>
<td>Caprioli et al. (2003)</td>
</tr>
<tr>
<td>GM1 gangliosidosis</td>
<td>GLB1</td>
<td>NM_000404.2: c.601C&gt;T Arg201Cys</td>
<td>NM_000404.2: c.1306C&gt;T Leu436Phe (rs34421970)</td>
<td>Caciotti et al. (2003)</td>
</tr>
<tr>
<td>Primary cortisol resistance</td>
<td>NR3C1</td>
<td>NM_001018077.1: c.2035G&gt;A Gly679Ser</td>
<td>NM_001018077.1: c.68G&gt;A Arg231Lys (rs6190)</td>
<td>Raef et al. (2008)</td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>SPINK5</td>
<td>NM_006846.3: c.2468dupA</td>
<td>NM_006846.3: c.1258G&gt;A Gly420Lys (rs2303067)</td>
<td>Di et al. (2009)</td>
</tr>
<tr>
<td>Hyperinsulinism</td>
<td>HADH</td>
<td>NM_005327.4: c.636+471G&gt;T</td>
<td>NM_005327.4: c.636+385A&gt;G (rs732941)</td>
<td>Flanagan et al. (2013)</td>
</tr>
</tbody>
</table>

the His558Arg substitution (rs1805124) in the SCN5A gene in a case of Brugada syndrome type 1 caused by compound heterozygous mutations (Asp1690Asn and Gly1748Asp) in SCN5A. Both mutations reduced the peak Na⁺ current density due to limited trafficking of the SCN5A protein towards the membrane, but Gly1748Asp also profoundly...
affected channel gating. The His558Arg polymorphism was found to be capable of rescuing the defective trafficking of SCN5A Asn1690 towards the membrane when present in cis to the pathological lesion (Núñez et al. 2013). Intriguingly, cotransfection with Asn1690, either alone or together with the modulatory His558Arg polymorphism, completely restored the gating defect associated with the pathogenic Gly1748Asp mutation in trans, although it only slightly rescued its trafficking.

In passing, it is perhaps pertinent to note that interplay between functional SNPs may, depending upon the precise combination of alleles involved, also be sufficient to bring about disease in the absence of a pathogenic mutation sensu stricto. Thus, a particular allele of one SNP may contribute to pathogenesis, but only in the presence of specific allele of another SNP. These SNPs may be neighbouring, as in the case of Glu918Asp (rs16022) and Glu993Val (rs16023) in the CACNA1A gene, which appear to contribute to migraine susceptibility (D’Onofrio et al. 2009). In the same way, when they occur together in cis, two otherwise neutral missense polymorphisms in the FMO3 gene [Glu158Lys (rs2266782) and Glu308Gly (rs2266780)] result in a decrease in FMO3 enzymatic activity that is sufficient to give rise to a mild form of trimethylaminuria (Akerman et al. 1999; Zschocke et al. 1999; D’Angelo et al. 2013). However, the interacting SNPs can also be located within different genes as in the case of Pro589Ser (rs1049296) in the transferrin (HFE) gene, which appear to contribute to malarial susceptibility (D’Onofrio et al. 2009). In the same way, when they occur together in cis, two otherwise neutral missense polymorphisms in the FMO3 gene [Glu158Lys (rs2266782) and Glu308Gly (rs2266780)] result in a decrease in FMO3 enzymatic activity that is sufficient to give rise to a mild form of trimethylaminuria (Akerman et al. 1999; Zschocke et al. 1999; D’Angelo et al. 2013). However, the interacting SNPs can also be located within different genes as in the case of Pro589Ser (rs1049296) in the transferrin (TF) gene and Cys282Tyr (rs1800562) in the haemochromatosis (HFE) gene which interact so as to increase the risk of Alzheimer disease (Robson et al. 2004; Kauwe et al. 2010). A further example of the combined effect of two unlinked SNPs is provided by an intronic SNP in the thrombospondin 2 (THBS2) gene (c.1478-8C>T; rs9406328) and a missense SNP in the metalloproteinase 9 (MMP9) gene (Gln279Arg; rs17576), which together increase the risk of lumbar disc herniation (Hirose et al. 2008). This type of situation may occupy the middle ground between monogenic and complex disorders.

The modifying SNP may also be regulatory in nature and can serve to render the pathogenic coding mutation more or less deleterious (and hence more or less penetrant) depending upon whether the allele harbouring it is more or less expressed than the wild-type allele. An example of this is provided by the −30C>T variant (rs17249141) in the LDLR gene promoter that has been shown to act in concert with a low-penetrant missense mutation in cis so as to give rise to an unusually severe form of familial hypercholesterolaemia (Snozek et al. 2009). The modulating influence of regulatory SNPs on the penetrance of coding mutations located in cis appears to be a widespread phenomenon in medical genetics, with the high expressing SNP allele usually increasing the clinical penetrance of the linked coding mutation (Lappalainen et al. 2011). One regulatory SNP is thought to act as a low-penetrance cancer susceptibility factor in its own right: homozygosity for the intronic 309T>G MDM2 variant (rs2279744), which leads to enhanced binding of the Sp1 transcription factor and MDM2 up-regulation, appears to increase the risk for many types of tumour, presumably in concert with other lesions (Hu et al. 2007).

Introns may also harbour SNPs that are capable of modulating the clinical penetrance of a given pathogenic mutation. Thus, a short tract of five thyrimidines (5T) in intron 8 of the CFTR gene, found in ~10% of individuals from the general population, can give rise to either congenital absence of the vas deferens (CAVD), non-classical cystic fibrosis or a normal phenotype when found in trans to a severe CFTR mutation (Kiesewetter et al. 1993; Cuppens et al. 1998). The number of TG repeats immediately adjacent to 5T is not only significantly associated with the level of alternative splicing of exon 9 of the CFTR gene (Cuppens et al. 1998; Niksic et al. 1999), but also influences clinical penetrance both in the context of cystic fibrosis and CAVD (Groman et al. 2004; Buratti et al. 2004; Lebo and Grody 2007). Another intronic modifying polymorphism is found in the FECH gene responsible for erythropaetic protoporphyria, an autosomal dominant disorder characterized by incomplete penetrance. This polymorphism (c.315-48T>G; rs2272783) modulates the use of a cryptic acceptor splice site, yielding an aberrantly spliced FECH mRNA which is degraded via nonsense-mediated mRNA decay (Gouya et al. 1999, 2002). The hypomorph C allele increases the penetrance of erythropoetetic protoporphyria when it occurs in trans to a pathogenic FECH mutation (Gouya et al. 2002, 2006). Finally, in a family with hyperinsulinism, a c.636+385A>G SNP (rs732941) in intron 5 of the HADH gene, which creates a cryptic acceptor splice site, acts in concert with a pathogenic HADH mutation (c.636+471G>T) in the same intron, which creates a cryptic donor splice site, to generate a 141-bp pseudoexon that leads to premature termination of translation (Flanagan et al. 2013).

A common C>T variant within an enhancer in intron 1 of the RET gene (rs2435357) serves to increase the clinical penetrance of RET coding sequence mutations (Emison et al. 2005). The T allele disrupts a SOX10-binding site, thereby reducing RET transactivation (Emison et al. 2010). A comparable example is provided by an intronic enhancer SNP (rs2596623) in the thyroid hormone receptor β (THRB) gene, which was found to be responsible for the pituitary cell-specific over-expression of a mutant thyroid hormone receptor β2 (Arg338Trp) in a case of pituitary cell-specific resistance to thyroid hormone (Alberobello et al. 2011).

Variants in the 3′ untranslated region (3′UTR) of the KCNQ1 gene reportedly modify disease severity in individuals with type 1 long QT syndrome resulting from KCNQ1 gene mutations (Amin et al. 2012). These variants
serve to reduce KCNQ1 gene expression, such that patients with one or more variants on their mutated KCNQ1 alleles have a shorter QT interval and a milder clinical phenotype, whereas patients with the variants on their normal KCNQ1 alleles exhibit significantly longer QT intervals and a more severe clinical phenotype. Another example of a modifying 3’UTR variant is provided by the G>A polymorphism (rs1799963) at position 20210 in the prothrombin (F2) gene, which increases the risk of venous thrombosis by enhancing F2 mRNA 3’ end formation efficiency, thereby boosting thrombin formation (Gehring et al. 2001). This F2 20210G>A polymorphism has been claimed to interact with a common F13A1 Val34Leu (rs5985) variant to confer a greatly increased risk of myocardial infarction (Butt et al. 2003). Some 3’UTR variants are located within microRNA-binding sites and may constitute low-penetrance risk factors for disease in their own right (Ahluwalia et al. 2009; Kontorovich et al. 2010; Qiu et al. 2011; Arnold et al. 2012). Finally, various SNPs in other non-coding RNAs (e.g. lincRNAs; Jendrzejewski et al. 2012; Kumar et al. 2013) appear to be disease associated and may therefore also influence disease penetrance.

**Influence of the gene expression level on mutation penetrance**

Humans are characterized by marked inter-individual differences in the expression levels of their genes (Stranger et al. 2007; Skelly et al. 2009; Cowley et al. 2009; Cheng et al. 2012). Since gene expression is controlled by a combination of cis- and trans-acting regulatory factors, one means by which heritable differences in gene expression may be mediated is through polymorphism either of trans-acting regulatory (transcription) factors or of the cis-active target sequences to which they bind. In the case of disease genes, such inter-individual variation in gene expression levels and patterns can influence the penetrance of pathological mutations. However, it should be appreciated that there are also substantial environmental and stochastic (non-genetic) components to gene expression that are likely to contribute to variable penetrance, even between monozygotic twins (Grundberg et al. 2012).

Differential allelic expression is a widespread phenomenon and is thought to be relevant to as many as 50% of all human genes (Williams et al. 2007; Cheung and Spielman 2009; Palacios et al. 2009). In autosomal dominant conditions where the two alleles of the disease gene are expressed at different levels, this discrepancy can favour either the mutant or the wild-type allele and hence may influence clinical penetrance in either direction (de la Chapelle 2009). Thus, in pulmonary arterial hypertension, a disease caused by mutations in the bone morphogenetic protein receptor type 2 (BMPR2) gene, the penetrance of the BMPR2 disease allele is dependent upon the level of expression of the wild-type BMPR2 allele (Hamid et al. 2009a). Similarly, in erythropoietic protoporphyria, an autosomal dominant condition caused by mutations in the ferrochelatase (FECH) gene, the penetrance of the pathogenic FECH allele is influenced by the level of expression of the wild-type FECH allele (Gouya et al. 1999; 2002; Di Pierro et al. 2007). Other examples of autosomal dominant conditions where the degree of clinical penetrance is modulated by differential expression of the wild-type and mutant alleles include hereditary elliptocytosis (SPTA1, Wilmotte et al. 1993), Marfan syndrome (FBN1, Hutchinson et al. 2003), retinoblastoma (RB1, Taylor et al. 2007), colorectal cancer (APC, Yan et al. 2002; TGFBR1, Valle et al. 2008) and breast and ovarian cancer (BRCA1, Ginolhac et al. 2003).

Perhaps, the best understood example of penetrance depending upon the level of expression of the wild-type allele is retinitis pigmentosa type 11 (Utz et al. 2013). This autosomal dominant condition is caused by mutations in the pre-mRNA processing factor 31 (PRPF31) gene located on chromosome 19q13.42. The clinical penetrance of the underlying mutations has been shown to depend upon the level of wild-type PRPF31 mRNA expression displayed by the patient (Vithana et al. 2003; Rivolta et al. 2006; Liu et al. 2008). Cells from asymptomatic carriers of PRPF31 mutations express a higher level of the wild-type allele than cells from affected patients: high enough for the wild-type PRPF31 mRNA level to lie within the range of the unaffected general population (Rivolta et al. 2006; Liu et al. 2008). The penetrance of PRPF31 mutations is reduced by transcriptional repression mediated by the product of the CCR4-NOT transcription complex, subunit 3 (CNOT3) gene which is linked to PRPF31 (McGee et al. 1997; Venturini et al. 2012). PRPF31 expression has also been found to be strongly influenced by an unlinked eQTL on chromosome 14q21-q23 (Rio Frio et al. 2008). The penetrance of PRPF31 mutations is therefore determined at least in part by a trans-acting modifier located on a different chromosome. The trans-acting alleles are inherited from the parent lacking the PRPF31 mutation; these alleles are presumably present in the general population, but appear only to be relevant to disease when they modulate the penetrance of PRPF31 mutations.

A slightly different scenario is exemplified by Schimke immune-osseus dysplasia (SIOD), a recessive condition, which appears to result from biallelic mutations in the SMARCAL1 gene. Several examples of SIOD families with incomplete penetrance have been reported (Bökenkamp et al. 2005; Dekel et al. 2008; Elizondo et al. 2009). It has recently been shown that SMARCAL1, a protein involved in chromatin remodelling, influences the transcription level of many genes (Baradaran-Heravi et al. 2012). Although...
SMARCAL1 deficiency is insufficient in itself to cause SIOD in Drosophila and mouse models, the addition of environmental (viz. heat shock) and genetic insults affecting transcription can successfully recapitulate the pathophysiology of SIOD (Baradaran-Heravi et al. 2012). The penetrance of SIOD therefore appears to be dependent upon the magnitude of the alteration of gene expression consequent to SMARCAL1 deficiency.

In the case of a splicing mutation (c.291+1 G>A; rs71640277) in intron 3 of the GH1 gene causing growth hormone deficiency type II, the expression levels of the mutant and wild-type alleles were found to correlate with the penetrance and expressivity of the deficiency state in different members of the same family (Hamid et al. 2009b). Some splicing mutations associated with low penetrance affect splicing in such a way that both normal-length and truncated transcripts are expressed from the same mutant allele, but presumably to different extents in different individuals (e.g. RB1, c.2211G>A; Schubert et al. 1997). Other such examples of reduced penetrance due to ‘leaky splicing’ involve splicing mutations in the SPAST gene causing hereditary spastic paraplegia (Svensson et al. 2001) and in the BTK gene causing X-linked agammaglobulinemia (Kaneko et al. 2005). The reduced penetrance characteristic of some splicing mutations may also result from alternative splicing (Rave-Harel et al. 1997; Chiba-Falek et al. 1998; Nissim-Rafinia and Kerem 2005; Zimmer et al. 2009; Szymanski et al. 2011; Cogan et al. 2012; Lee et al. 2012b) or internal translational start site initiation (Sánchez-Sánchez et al. 2007).

Special cases of differential allelic expression are of course provided by X-inactivation (Dobyns et al. 2004) and imprinting (Lo et al. 2003), both of which are discussed below.

Allele dosage and its influence on penetrance

Formally, use of the term ‘autosomal dominant’ implies that the homozygotes exhibit the same or a similar clinical phenotype to the heterozygotes, as is the case in Hirschsprung disease where the length of the expanded HTT CAG triplet repeat appears to be predictive of the age of onset irrespective of the presence or absence of a second expanded HTT allele (Lee et al. 2012a). However, in practice, for most ‘dominant’ human disorders in which homozygotes have been reported, their clinical symptoms tend to be significantly more severe than in the heterozygotes (Vogel and Motulsky 1997). This would seem to be especially true in the context of low-penetrance mutations such as those identified in the SCN4A and CLCN1 genes, causing muscle channelopathies, conditions which are usually held to be transmitted in an autosomal dominant fashion. Patients homozygous for sodium channel mutations causing paramyotonia congenita (SCN4A, Ile1393Thr), hypokalemic periodic paralysis (SCN4A, Arg1132Gln) and myotonia congenita (CLCN1, Gly190Ser, Ile556Asn, Ala313Thr, Ile556Asn) display much more severe clinical features than patients heterozygous for these mutations (Plassart-Schiess et al. 1998; Arzel-Hézode et al. 2010; Shalata et al. 2010). The aforementioned mutations were also found to exhibit reduced penetrance in heterozygotes.

Mutations in the RET gene, associated with isolated Hirschsprung disease, are dominant loss-of-function mutations with incomplete penetrance and variable expressivity. Baseline-Vanagaite et al. (2007) reported a c.1263+5G>A splicing mutation in the homozygous state in three females with severe Hirschsprung disease and in the heterozygous state in a male patient with short-segment Hirschsprung disease. In addition, a hypomorphic RET-predisposing allele, rs2435357, located in the first intron of the RET gene, was found in the heterozygous state in the male patient but not in the three affected females. Whilst the heterozygous c.1263+5G>A mutation is known to be low penetrance for short-segment Hirschsprung disease, the homozygous state is fully penetrant for total aganglionosis or long-segment Hirschsprung disease. Thus, the penetrance of RET gene mutations in Hirschsprung disease depends not only on the nature of the mutation but also on the allele dosage.

Influence of copy number variation on mutation penetrance

Estimates of the clinical penetrance of recurrent pathogenic copy number variants (CNVs) vary widely, depending upon CNV size, genomic location and the disorder in question (Ben-Shachar et al. 2009; Vassos et al. 2010; Breckpot et al. 2011; Čiuladaitė et al. 2011; Hosak et al. 2012; Klopocki et al. 2012; Rosenfeld et al. 2013; Vaags et al. 2012; Weischenfeldt et al. 2013; Dabell et al. 2013; Carvill and Mefford 2013; Tropeano et al. 2013). In their study of children known to carry a CNV associated with intellectual disability and congenital abnormalities, Girirajan et al. (2012) reported synergy between multiple large CNVs leading to a particularly severe clinical presentation. Such a two-hit model, or ‘oligogenic heterozygosity’ as it has been termed, also appears to be characteristic of autism (Pinto et al. 2010; Schaaf et al. 2011b; Klei et al. 2012; Gau et al. 2012).

The penetrance of a given CNV may also be influenced by genetic variants in the vicinity. Thus, a submicroscopic deletion of 1q21.1 (encompassing the RBM8A gene) has been reported to interact with a low-frequency functional SNP in the regulatory region of the wild-type RBM8A allele to cause thrombocytopenia with absent radii (Albers et al. 2012).
Several papers have now suggested that CNVs can also act as genetic modifiers of phenotype severity in a variety of different disease contexts (Beckmann et al. 2007; Chaudru et al. 2009; El-Hattab et al. 2010; Mulley et al. 2011; Jiang et al. 2011; Carvalho et al. 2012; Shen et al. 2013b). CNVs may influence the penetrance of a clinical phenotype indirectly as well as directly. For example, a mutant gene might be ‘covered’ by a CNV in a given individual, so that the expected clinical phenotype would be masked by the presence of an additional wild-type copy in cis to the gene in question. Consistent with this postulate, Ng et al. (2008) reported that ~30% of nonsense SNPs occur in genes residing within segmental duplications, a proportion some threefold larger than that noted for synonymous SNPs. Genes harbouring nonsense SNPs were also found to belong to larger gene families (Ng et al. 2008) suggesting that some functional redundancy could also exist between paralogous human genes. In support of this idea, Hsiao and Vitkup (2008) reported that those human genes which have a homologue with >90% sequence similarity are ~3 times less likely to harbour disease-causing mutations than genes with less closely related homologues. Hsiao and Vitkup (2008) interpreted their findings in terms of ‘genetic robustness’ against null mutations, with the duplicated sequences providing ‘backup’ by potentiating the functional compensation/complementation of homologous genes in the event that the latter acquired deleterious mutations. The capacity to be functionally compensated appears to vary, in the order non-disease genes > monogenic disease genes > polygenic disease genes (Podder and Ghosh 2011). One example of how a CNV can ameliorate the clinical phenotype is spinal muscular atrophy where an increased copy number of the SMN2 gene can greatly reduce the severity of the disease caused by the homozygous deletion of the SMN1 gene, because the SMN2 gene, which lacks a splicing enhancer, can nevertheless generate some functional product thereby compensating functionally in a copy number-dependent fashion for the loss of the SMN1 gene (Vitali et al. 1999; Harada et al. 2002; Wirth et al. 2006). Another example of how the clinical and/or phenotypic impact of a mutant gene can be nullified by a CNV is provided by a foetus that possessed paternal (Gln318Term) and maternal (Arg356Term) nonsense mutations of the CYP21A2 gene but lacked the normal clinical sequelae of congenital adrenal hyperplasia; this was found to be due to a duplication of the CYP21A2 gene on the paternal allele (Kleinle et al. 2009; Lekarev et al. 2013).

Influence of modifier genes on disease penetrance

“For a so-called single gene disorder, there is one gene that may be primarily responsible for the pathogenesis with one or more independently inherited modifier genes that influence the phenotype. On the other hand, for a complex trait, the primacy of any individual gene is not perceptible, and the interaction of two or more independently inherited pairs of alleles, most likely influenced by additional modifier genes, results in the disease. The consequence of this conceptual framework is that there is no such thing as a ‘single’ gene disorder. In other words, there is no obvious clear distinction between simple Mendelian and complex traits: genetic diseases represent a continuum with diminishing influence from a single primary gene influenced by modifier genes, to increasingly shared influence by multiple genes”.

Dipple and McCabe (2000)

It is sometimes claimed that sickle cell anaemia is the simplest of all Mendelian disorders in that it is caused by one specific mutation (Glu7Val) in the β-globin (HBB) gene. However, this single-mutation monogenic disorder is not as simple as it might at first appear; indeed, it is characterized by marked clinical heterogeneity and incomplete penetrance of subphenotypes which is due in part to allelic variation and in part to variants in unlinked modifier genes (Steinberg and Sebastiani 2012). If we extrapolate from the archetypal example of sickle cell disease to other Mendelian disorders, it is not unreasonable to expect the action of modifier genes to be the rule rather than the exception. Indeed, variants in unlinked modifier genes have been reported to influence penetrance in a variety of different inherited diseases including pancreatitis (Khalid et al. 2006), breast cancer (Wolf et al. 2010; Wang et al. 2010a; Walker et al. 2010; Antoniou and Chenevix-Trench 2010; Esteban Carderios et al. 2012; Harlid et al. 2012), Gaucher disease (Taddei et al. 2009; Zhang et al. 2012), retinitis pigmentosa (Rio Frio et al. 2008; Venturini et al. 2012), haemochromatosis (Krayenbuehl et al. 2010), hypertrophic cardiomyopathy (Daw et al. 2007), frontotemporal lobar degeneration (Finch et al. 2011) and amyloid polyneuropathy (Soares et al. 2005) among others. In familial late-onset Alzheimer disease, modifying loci may either influence the risk (Sleegers et al. 2009; Cruchaga et al. 2012) or the age of onset of disease (Wijsman et al. 2005; Marchani et al. 2010). Importantly, a significant excess of rare coding APP, PSEN1 and PSEN2 variants was noted in probands from late-onset Alzheimer disease families even though these variants did not actually co-segregate with the disease; this suggests that the variants in question may nevertheless serve to modulate the risk of disease (Cruchaga et al. 2012). An excess of rare variants as compared to controls has also been noted in patients with various other disorders including hypertriglyceridaemia (Johansen et al. 2012; Talmud 2007), hypertrophic
cardiomyopathy (Lopes et al. 2013) and autism spectrum disorder (Mondal et al. 2012).

A typical example of a modifier gene in action is provided by long QT syndrome. The clinical penetrance of KCNQ1 (potassium voltage-gated channel, KQT-like subfamily, member 1) mutations in this disorder is influenced by two coding sequence polymorphisms [Ser49Gly (rs1801252) and Arg389Gly (rs1801253)] in the ADRB1 gene. Individuals homozygous for the Arg389 allele tend to have shorter QT intervals, whereas individuals homozygous for Ser49 tend to have longer QT intervals than those with other genotypes (Paavonen et al. 2007). Interestingly, those individuals doubly homozygous for Arg389 and Ser49 were found to be indistinguishable from the remainder of the patient cohort, both in terms of their QT intervals and in terms of clinical penetrance.

The minor allele of a variant in the complement receptor 1 (CRI) gene, Ser1610Thr (rs4844609), which has a population frequency of 0.02, is associated with episodic memory decline and susceptibility to Alzheimer disease (Keenan et al. 2012). However, this effect appears largely dependent upon an interaction with APOE-ε4, itself an important risk factor for Alzheimer disease (Mayeux et al. 1993).

Hirschsprung disease is one of the most complex genetic disorders in terms of the number of modifier genes (Garcia-Barcelo et al. 2009; Tang et al. 2010) known to influence the penetrance of its causative mutations, which has been estimated to be of the order of 50–70 % (Bolk et al. 2000). The best characterized of these modifier genes is the neurogenin 1 gene (NRG1; Tang et al. 2011, 2012a); however, most probably still remain to be identified. Some Bardet–Biedl syndrome patients also present with Hirschsprung disease. It appears that RET, the major gene involved in the aetiology of Hirschsprung disease, acts as a modifier of the Hirschsprung disease phenotype in Bardet–Biedl syndrome (de Pontual et al. 2007). Some families with Hirschsprung disease and Bardet–Biedl syndrome harbour mutations in their BBS4, BBS5, BBS7 and RET genes (de Pontual et al. 2009). Sánchez-Mejías et al. (2009) reported a Hirschsprung disease family in which mutations in three different genes (RET, NTRK3 and EDN3) contributed to the disease phenotype; the RET and NTRK3 mutations were both necessary and sufficient to give rise to the clinical phenotype, whereas the EDN3 mutation appeared to act as a modifier. More recently, copy number variations in various neurodevelopmental genes (MAPK10, ZFHX1B, SOX2 and NRG2) have been shown to modify the penetrance of Hirschsprung disease (Jiang et al. 2011; Tang et al. 2012b). Taken together, these findings are consistent with an impact of both common and rare variants on the inheritance (and hence penetrance) of this highly complex disorder (Sánchez-Mejías et al. 2009; Núñez-Torres et al. 2011; Alves et al. 2013).

Table 3 lists a number of well-characterized examples of specific variants in modifier genes that serve to modulate the clinical penetrance of diseases caused by mutation(s) at unlinked loci.

The plastin 3 (PLS3) gene acts as a modifier of the clinical penetrance of autosomal recessive spinal muscular atrophy, caused by the homozygous deletion of the SMN1 gene. Oprea et al. (2008) studied spinal muscular atrophy-discordant families with affected and unaffected SMN1-deleted siblings and found that all unaffected SMN1-deleted siblings were characterized by a high PLS3 expression level in blood cells, considerably higher than in their affected counterparts (and despite high PLS expression being evident in only 5 % of healthy controls). Although it is still unclear whether PLS3 expression is regulated by cis- or trans-acting factors, it would appear that high PLS3 expression serves to rescue the spinal muscular atrophy patient from the detrimental effects of SMN1 deletion by promoting axonogenesis through elevation of the level of F-actin (Oprea et al. 2008) and ultimately by improving neuromuscular transmission (Ackermann et al. 2013).

A unique kind of modifying effect is exemplified by α-thalassaemia/mental retardation syndrome, caused by mutations of the ATRX gene. The ATRX protein binds to variable number tandem repeat sequences (VNTRs) in the human genome, and genes associated with these VNTRs are dysregulated when ATRX is mutated (Law et al. 2010). Law et al. (2010) identified 917 ATRX targets in primary human erythroid cells, including one in the α-globin (HBA) locus. This ψδ VNTR was found to be highly polymorphic in terms of its length and acted as a length-dependent negative regulator of gene expression, its length serving to influence the degree of α-thalassaemia observed in an α-thalassaemia/mental retardation syndrome patient. Thus, the length of the ψδ VNTR could explain the incomplete penetrance of α-thalassaemia noted in individuals with identical ATRX mutations. A similar mechanism could underlie other genetic traits characterized by reduced penetrance.

**Digenic mutations and disease penetrance**

In a typical autosomal Mendelian condition, a single mutation (dominant) or two mutations (recessive) at a specific locus give rise to a clinical phenotype. By contrast, digenic inheritance occurs in cases where the interaction of mutations in two different genes is required for the expression of the clinical phenotype. In this situation, a mutation in one copy of each gene is required for the full clinical phenotype to manifest. In the absence of one of the component mutations, the other mutation may be non-
Table 3  A selection of well-characterized examples of polymorphic variants in modifier genes that serve to modulate the clinical penetrance and/or severity of an inherited disease caused by mutation(s) at an unlinked locus

<table>
<thead>
<tr>
<th>Disease</th>
<th>Primary disease gene</th>
<th>Modifier gene/variant</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-linked retinitis pigmentosa</td>
<td>RPGR</td>
<td>IQCB1/NM_001023570.2:c.1178T&gt;A Ile393Asn (rs1114152)</td>
<td>Fahim et al. (2011)</td>
</tr>
<tr>
<td>X-linked retinitis pigmentosa</td>
<td>RPGR</td>
<td>RPGR1/NM_015272.2:c.2231G&gt;A Arg744Gln (rs2302677)</td>
<td>Fahim et al. (2011)</td>
</tr>
<tr>
<td>Retinal generation in ciliopathies</td>
<td>RPGR or NPHP5</td>
<td>RPRGR1/NM_015272.2:c.685G&gt;A Ala229Thr (rs6174707)</td>
<td>Khanna et al. (2009)</td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td>RB1</td>
<td>MDM2/NM_002392.4:c.14+309T&gt;G (rs2279744)</td>
<td>Castéra et al. (2010)</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia</td>
<td>LDLR</td>
<td>PCSK9/NM_174936.3:c.63_65dupGCT (rs35574083)</td>
<td>Abifadel et al. (2009)</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia</td>
<td>LDLR</td>
<td>APOB/NM_000384.2:c.10580G&gt;C (rs5742904)</td>
<td>Benlian et al. (1996) and Taylor et al. (2010)</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia</td>
<td>LDLR</td>
<td>CFH/NM_000186.3:c.1204T&gt;C Tyr402His (rs1061170)</td>
<td>Koeijers et al. (2009)</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia</td>
<td>LDLR</td>
<td>APOH/NM_000042.4:c.1204T&gt;C Leu266Val (rs4581)</td>
<td>Takada et al. (2003a)</td>
</tr>
<tr>
<td>Familial breast cancer</td>
<td>BRCA2</td>
<td>RAD51/NM_0002875.4:c.98G&gt;C (rs1801320)</td>
<td>Antoniou et al. (2007)</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>BRCA1 or BRCA2</td>
<td>IRS1/NM_000163.4:c.1630A&gt;G (rs1800470)</td>
<td>Drumm et al. (2005)</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>MSH2 or MLH1</td>
<td>RNASEL/NM_021133.3:c.1385G&gt;A Arg462Gln (rs486907)</td>
<td>Kru¨ger et al. (2007)</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>MSH2 or MLH1</td>
<td>RNASEL/NM_021133.3:c.1385G&gt;A Arg462Gln (rs486907)</td>
<td>Krüger et al. (2005)</td>
</tr>
<tr>
<td>Lynch syndrome</td>
<td>MSH2 or MLH1</td>
<td>TP53/NM_000546.3:c.215G&gt;C Arg72Pro (rs1042522)</td>
<td>Krüger et al. (2007)</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>CFTR</td>
<td>TGFBI/NM_000660.4:c.29C&gt;T Pro10Leu (rs1800470)</td>
<td>Phillips et al. (2008)</td>
</tr>
<tr>
<td>Familial pulmonary arterial hypertension</td>
<td>BMPR2</td>
<td>TGFBI/NM_000660.4:c.-1347C&gt;T (rs1800469) &amp; NM_000660.4:c.29C&gt;T Leu10Pro (rs1800470)</td>
<td>Phillips et al. (2008)</td>
</tr>
<tr>
<td>Paget’s disease</td>
<td>SQSTM1</td>
<td>TNFRSF11A/NM_003839.2:c.575T&gt;C Val192Ala (rs1805034)</td>
<td>Gianfrancesco et al. (2012)</td>
</tr>
<tr>
<td>X-linked variable immunodeficiency</td>
<td>XIAP</td>
<td>CD40LG/NM_000074.2:c.655G&gt;A Gly219Arg (rs148594123)</td>
<td>Rigaud et al. (2011)</td>
</tr>
<tr>
<td>Haemochromatosis</td>
<td>HFE</td>
<td>CYP2D6/NM_024843.3:c.-399T&gt;G (rs884409)</td>
<td>Constantine et al. (2009)</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>GBA</td>
<td>MTH1/NM_002455.3:c.187T&gt;A Ser63Thr (rs760077)</td>
<td>Gan-Or et al. (2011)</td>
</tr>
<tr>
<td>Recessive dystrophic epidermolysis bullosa</td>
<td>COL7A1</td>
<td>MMP1/NM_002421.4:c.-1673delG (rs1797950)</td>
<td>Titeux et al. (2008)</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>SOD1</td>
<td>CHGB/NM_001819.3:c.1238C&gt;T Pro413Leu (rs742710)</td>
<td>Gros-Louis et al. (2009)</td>
</tr>
<tr>
<td>Huntington disease</td>
<td>HTT</td>
<td>HAP1/NM_177977.2:c.1322C&gt;T Thr441Met (rs4523977)</td>
<td>Metzger et al. (2008)</td>
</tr>
<tr>
<td>Fatal keratitis</td>
<td>G6PD</td>
<td>UGT1A1(TA)1(TA)7 (rs8175347), HGVS nomenclature not available</td>
<td>Zang et al. (2009)</td>
</tr>
<tr>
<td>Atypical haemolytic uraemic syndrome</td>
<td>MCP or CFH</td>
<td>C4BPA/NM_000715.3:c.719G&gt;A Arg240His (rs45574833)</td>
<td>Blom et al. (2008)</td>
</tr>
<tr>
<td>Spinal muscular atrophy</td>
<td>SMN1</td>
<td>SMN2/NM_017411.3:c.859G&gt;C Gly287Arg (rs121909192)</td>
<td>Prior et al. (2009)</td>
</tr>
<tr>
<td>Long QT syndrome</td>
<td>KCNQ1</td>
<td>KCNH2/NM_000238.3:c.2690A&gt;C Lys897Thr (rs1805123)</td>
<td>Cordeiro et al. (2010)</td>
</tr>
<tr>
<td>Long QT syndrome</td>
<td>KCNQ1</td>
<td>ADRB1/NM_000684.2:c.145A&gt;G Ser49Gly (rs1801252) &amp; NM_000684.2:c.1165G&gt;C Arg389Gly (rs1801253)</td>
<td>Paavonen et al. (2007)</td>
</tr>
<tr>
<td>Long QT syndrome</td>
<td>KCNQ1</td>
<td>NOS1AP/NC_000001.10:g.162029907A&gt;T (rs4657139)</td>
<td>Crotti et al. (2009)</td>
</tr>
</tbody>
</table>
penetrant (digenic inheritance sensu stricto) or could be responsible for a less severe clinical phenotype (digenic inheritance sensu lato). At least 100 cases of probable and possible examples of digenic inheritance causing human inherited disease have been reported to date (Table 4).

Digenic inheritance may occur as a result of mutations in genes encoding different subunits of the same multimeric protein (e.g. PRPH2 and ROM1), an oligomeric protein complex (e.g. KCNJ10 and SLC26A4; KRT14 and KRT5) or simply two proteins that interact functionally with each other (CDH23 and PCDH15; DSG2 and DSC2; PARK7 and PINK1). However, mutations in receptor/ligand pairs can also give rise to digenic inheritance (e.g. PROKR2 and PROKR2). Alternatively, digenic inheritance can involve mutations located in different genes, but compromising the same regulatory (e.g. KCNN10 and SLC26A4; KRT14 and KRT5) or biosynthetic (e.g. ZMPSTE24 and LMNA; CPOX and PPOX; MYOC and CYP1B1) or degradative (e.g. PCSK9 and LDLR) pathway. Finally, the combination of a mutation in a transcription factor with a mutation in a target gene of that transcription factor can also serve to reduce the amount of the protein in question to a level sufficient to cause a disease phenotype (e.g. MITF and TYR; FOXC1 and PITX2).

In practice, it is not always clear if a given situation constitutes true digenic inheritance (e.g. Kajiwara et al. 1994) or whether it is simply the coinheritance of two mutations in different genes (e.g. Gruber et al. 2009; Serrano-Fernández et al. 2009; Ekvall et al. 2011). In the former case, the expression of the disease phenotype actually requires the presence of both gene lesions. In the latter case, coinheritance of the two gene lesions may serve to aggravate the clinical phenotype, but each lesion is independently associated with its own characteristic clinical sequelae. In true digenic inheritance, mutations in both genes must be present for the genetic disorder to be manifest. Since many of the disorders reported (in Table 4) to be characterized by digenic inheritance also have monogenic forms in which just one of the two genes has been mutated (whether in the heterozygous or homozygous/compound heterozygous state), it is unclear how many of the examples listed really represent digenic inheritance sensu stricto. However, in many of the listed examples, the doubly heterozygous probands exhibit earlier onset or a more severe clinical phenotype than their singly heterozygous relatives (although this is not invariably so; Marras et al. 2010). Further, whereas the truly ‘digenic’ patients tend to be characterized by complete penetrance, the monogenic disease genotypes often exhibit reduced penetrance as in e.g. normosmic idiopathic hypogonadotropic hypogonadism (see Table 4).

In some of the cases listed in Table 4, the digenic inheritance may be confined to a single proband and hence it is not always straightforward to distinguish true digenic inheritance from the chance coinheritance of two mutations in unlinked genes. This notwithstanding, the requirement for the involvement of a second mutated gene may depend upon the specific mutations that are segregating in the pedigrees. If digenic inheritance eventually turns out to be more frequent than previously appreciated, it could provide yet another reason why some potentially pathogenic alleles are present in the general population in the absence of overt disease. At the very least, in many cases the clinical penetrance of the condition in question is likely to be greater when two relevant genes have been functionally compromised by mutation than if only one had been mutated.

### Table 3 continued

<table>
<thead>
<tr>
<th>Disease</th>
<th>Primary disease gene</th>
<th>Modifier gene/variant</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial venous thrombosis</td>
<td>PROC</td>
<td>F5’NM_00130.4:c.1601G&gt;A Arg534Gln (rs6025; Factor V Leiden)</td>
<td>Koeleman et al. (1994); Gandrille et al. (1995) and Cafolla et al. (2012)</td>
</tr>
<tr>
<td>Familial venous thrombosis</td>
<td>PROS1</td>
<td>F5’NM_00130.4:c.1601G&gt;A Arg534Gln (rs6025; Factor V Leiden)</td>
<td>Koeleman et al. (1995)</td>
</tr>
<tr>
<td>Familial venous thrombosis</td>
<td>SERPINCl</td>
<td>F5’NM_00130.4:c.1601G&gt;A Arg534Gln (rs6025; Factor V Leiden)</td>
<td>Van Boven et al. (1996)</td>
</tr>
<tr>
<td>Hypertrophic cardiomyopathy</td>
<td>MYBPC3 or MYH7</td>
<td>CALM3/NN_005184.2:c.–157T&gt;A (rs150954567)</td>
<td>Friedrich et al. (2009)</td>
</tr>
<tr>
<td>Familial Mediterranean fever</td>
<td>MEFV</td>
<td>SAA1/NN_000331.4:c.–197C&gt;T (rs11024595)</td>
<td>Migita et al. (2013)</td>
</tr>
</tbody>
</table>

Oligogenic inheritance and its implications for disease penetrance

Triallelic inheritance has been described as being “a bridge between Mendelian and multifactorial traits” (Eichers et al. 2004). There are a burgeoning number of reported examples of digenic triallelic inheritance including nephropathiasis (Hoefele et al. 2007), venous thrombosis (Formstone et al. 1996; Brenner et al. 1996) and cortisone
<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waardenburg syndrome type 2</td>
<td>MITF</td>
<td>PAX3 or OCA3 or TYR or</td>
<td>Morell et al. (1997); Chiang et al. (2009); Yan et al. (2011) and Yang et al. (2013)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GJB2</td>
<td></td>
</tr>
<tr>
<td>Retinitis pigmentosa*</td>
<td>PRPH2</td>
<td>ROM1 or RHO or PDE6B</td>
<td>Kajiwara et al. (1994); Loewen et al. (2001); Sullivan et al. (2006) and Jin et al. (2008)</td>
</tr>
<tr>
<td>Retinitis pigmentosa</td>
<td>RHO</td>
<td>PRPF31</td>
<td>Lim et al. (2009)</td>
</tr>
<tr>
<td>Retinitis pigmentosa*</td>
<td>PDE6B</td>
<td>GPR98</td>
<td>Hmani-Aifa et al. (2009)</td>
</tr>
<tr>
<td>Progressive cone dystrophy</td>
<td>CNGA3</td>
<td>CNGB3</td>
<td>Thiadens et al. (2010)</td>
</tr>
<tr>
<td>Frontotemporal dementia</td>
<td>PSEN1</td>
<td>PRNP</td>
<td>Bernardi et al. (2011)</td>
</tr>
<tr>
<td>Leber congenital amaurosis</td>
<td>RPE65</td>
<td>GUCY2D</td>
<td>Silva et al. (2004)</td>
</tr>
<tr>
<td>Idiopathic hypogonadotropic hypogonadism</td>
<td>FGFR1</td>
<td>GNRHR or NELF</td>
<td>Pitteloud et al. (2007)</td>
</tr>
<tr>
<td>Bilateral cystic renal dysplasia</td>
<td>DACH1</td>
<td>BMP4</td>
<td>Schild et al. (2013)</td>
</tr>
<tr>
<td>Glaucoma, early onset</td>
<td>MYOC</td>
<td>CYP1B1 or LTBP2</td>
<td>Vincent et al. (2002), Geyer et al. (2011) and Azmanov et al. (2011)</td>
</tr>
<tr>
<td>Severe insulin resistance</td>
<td>PPARG</td>
<td>PPP1R3A</td>
<td>Savage et al. (2002)</td>
</tr>
<tr>
<td>Usher syndrome type 2</td>
<td>PDZD7</td>
<td>GPR98</td>
<td>Ebermann et al. (2010)</td>
</tr>
<tr>
<td>Usher syndrome type 1-associated deafness</td>
<td>CDH23</td>
<td>PCDH15</td>
<td>Zheng et al. (2005)</td>
</tr>
<tr>
<td>Hidrotic ectodermal dysplasia</td>
<td>GJB2</td>
<td>GJA1</td>
<td>Kellermayer et al. (2005)</td>
</tr>
<tr>
<td>Non-syndromic deafness</td>
<td>GJB2</td>
<td>GJB3</td>
<td>Liu et al. (2009)</td>
</tr>
<tr>
<td>Hearing loss</td>
<td>GJB2</td>
<td>SLC26A4</td>
<td>Sagong et al. (2013)</td>
</tr>
<tr>
<td>Non-syndromic hearing loss associated with an enlarged vestibular</td>
<td>KCNJ10</td>
<td>SLC26A4</td>
<td>Yang et al. (2009)</td>
</tr>
<tr>
<td>vestibular aqueduct/Pendred syndrome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porphyria*</td>
<td>CPOX</td>
<td>PPOX</td>
<td>van Tuyl van Serooskerken et al. (2011)</td>
</tr>
<tr>
<td>Atypical haemolytic uraemic syndrome</td>
<td>CF1</td>
<td>CD46 or C3 or CFB or CFHR1</td>
<td>Esparza-Gordillo et al. (2006), Westra et al. (2010) and Bresin et al. (2013)</td>
</tr>
<tr>
<td>Atypical haemolytic uraemic syndrome</td>
<td>CFH</td>
<td>CD46 or CF1 or C3 or CFB</td>
<td>Sullivan et al. (2011), Bresin et al. (2013) and Fan et al. (2013)</td>
</tr>
<tr>
<td>Epidermolysis bullosa simplex</td>
<td>KRT14</td>
<td>KRT5</td>
<td>Padaloon-Brauch et al. (2012)</td>
</tr>
<tr>
<td>Junctional epidermolysis bullosa</td>
<td>COL17A1</td>
<td>LAMB3</td>
<td>Floeth and Bruckner-Tuderman (1999)</td>
</tr>
<tr>
<td>Long QT syndrome*</td>
<td>KCNQ1</td>
<td>KCNH2 or KCNE1 or SCN5A</td>
<td>Schwartz et al. (2003), Westenskow et al. (2004), Tester et al. (2005) and Itoh et al. (2010)</td>
</tr>
<tr>
<td>Long QT syndrome*</td>
<td>KCNH2</td>
<td>SCN5A or KCNE1</td>
<td>Schwartz et al. (2003), Westenskow et al. (2004) and Tester et al. (2005)</td>
</tr>
<tr>
<td>Long QT syndrome*</td>
<td>SCN5A</td>
<td>SNTA1 or KCNE1</td>
<td>Westenskow et al. (2004) and Hu et al. (2013)</td>
</tr>
<tr>
<td>Haemochromatosis*</td>
<td>HFE</td>
<td>HAMP or TFR2</td>
<td>Merryweather-Clarke et al. (2003), Jacolot et al. (2004), Island et al. (2009), Aliès et al. (2009) and Del-Castillo-Rueda et al. (2012)</td>
</tr>
<tr>
<td>Kallmann syndrome</td>
<td>PROK2</td>
<td>PROKR2</td>
<td>Cole et al. (2008), Sarfati et al. (2010) and Shaw et al. (2011)</td>
</tr>
<tr>
<td>Kallmann syndrome</td>
<td>NELF</td>
<td>KALI or TACR3</td>
<td>Xu et al. (2011) and Quaynor et al. (2011)</td>
</tr>
<tr>
<td>Kallmann syndrome</td>
<td>PROKR2</td>
<td>KALI</td>
<td>Dodé et al. (2006), Canto et al. (2009) and Shaw et al. (2011)</td>
</tr>
<tr>
<td>Kallmann syndrome</td>
<td>KALI</td>
<td>TACR3 or WDR11 or CHD7</td>
<td>Quaynor et al. (2011) and Shaw et al. (2011)</td>
</tr>
<tr>
<td>Normosmic idiopathic hypogonadotropic hypogonadism</td>
<td>GNRHR</td>
<td>KALI</td>
<td>Quaynor et al. (2011)</td>
</tr>
<tr>
<td>Normosmic idiopathic hypogonadotropic hypogonadism</td>
<td>WDR11</td>
<td>GNRHR</td>
<td>Quaynor et al. (2011)</td>
</tr>
<tr>
<td>Normosmic idiopathic hypogonadotropic hypogonadism</td>
<td>FGFR1</td>
<td>GNRHR or PROKR2 or FGFR3 or KALI or GPR54</td>
<td>Raivo et al. (2009), Sykiotis et al. (2010) and Shaw et al. (2011)</td>
</tr>
<tr>
<td>Systemic amyloid A amyloidosis</td>
<td>TNFRSF1A</td>
<td>MEFV</td>
<td>Cigni et al. (2006) and Mereuta et al. (2013)</td>
</tr>
<tr>
<td>Disease</td>
<td>Gene 1</td>
<td>Gene 2</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia*</td>
<td>LDLR</td>
<td>PCSK9</td>
<td>Pisciotta et al. (2006), Noguchi et al. (2010) and Bertolini et al. (2013)</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia</td>
<td>LDLR</td>
<td>APOB</td>
<td>Bertolini et al. (2013)</td>
</tr>
<tr>
<td>Familial hypercholesterolaemia</td>
<td>LDLR</td>
<td>LDLRAP1</td>
<td>Tada et al. (2011)</td>
</tr>
<tr>
<td>Severe congenital neutropenia</td>
<td>ELANE</td>
<td>6FPC3  or HAX1</td>
<td>Germeshausen et al. (2010)</td>
</tr>
<tr>
<td>McArdle’s disease*</td>
<td>PYGM</td>
<td>CPT2</td>
<td>Vockley et al. (2000)</td>
</tr>
<tr>
<td>Parkinson’s disease, early onset</td>
<td>PINK1</td>
<td>PARK2  or PARK7</td>
<td>Tang et al. (2006) and Funayama et al. (2008)</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>LRRK2</td>
<td>PRKN</td>
<td>Dächsel et al. (2006)</td>
</tr>
<tr>
<td>Emery–Dreifuss muscular dystrophy*</td>
<td>LMNA</td>
<td>DES</td>
<td>Muntoni et al. (2006)</td>
</tr>
<tr>
<td>Joubert syndrome and nephronophthisis*</td>
<td>NPhip1</td>
<td>NPhip6</td>
<td>Tory et al. (2007)</td>
</tr>
<tr>
<td>Axenfeld–Rieger syndrome</td>
<td>FOXC1</td>
<td>PITX2</td>
<td>Kelberman et al. (2011)</td>
</tr>
<tr>
<td>Cortisone reductase deficiency</td>
<td>HSD11B1</td>
<td>H6PD</td>
<td>Draper et al. (2003) and San Millán et al. (2005)</td>
</tr>
<tr>
<td>Hypertrophic cardiomyopathy*</td>
<td>MYBPC3</td>
<td>TNNT2  or TNNT3 or MYH7 or TPM1</td>
<td>Richard et al. (2003), Van Driest et al. (2004) Ingles et al. (2005), Millat et al. (2010), Kubo et al. (2011) and Zou et al. (2013)</td>
</tr>
<tr>
<td>Hypertrophic cardiomyopathy*</td>
<td>MYH7</td>
<td>TNNT2  or MYL2 or TNNT3 or ACTC1</td>
<td>Millat et al. (2010) and Zou et al. (2013)</td>
</tr>
<tr>
<td>Restrictive cardiomyopathy*</td>
<td>MYL2</td>
<td>MYL3</td>
<td>Caleshu et al. (2011)</td>
</tr>
<tr>
<td>Rasopathy phenotype with severe hypertrophic cardiomyopathy</td>
<td>PTPN11</td>
<td>SOS1</td>
<td>Fahrner et al. (2012)</td>
</tr>
<tr>
<td>Arrhythogenic right ventricular cardiomyopathy</td>
<td>DES</td>
<td>PKP2</td>
<td>Lorenzon et al. (2013)</td>
</tr>
<tr>
<td>Arrhythogenic right ventricular cardiomyopathy*</td>
<td>DES</td>
<td>DSG2</td>
<td>Rasmussen et al. (2013)</td>
</tr>
<tr>
<td>Arrhythogenic right ventricular cardiomyopathy</td>
<td>PKP2</td>
<td>DSP or DSG2 or PKP4 or DSC2</td>
<td>Xu et al. (2010)</td>
</tr>
<tr>
<td>Arrhythogenic right ventricular dysplasia/cardiomopathy</td>
<td>DSG2</td>
<td>DSC or PKP2</td>
<td>Bhuiyan et al. (2009) and Nakajima et al. (2012)</td>
</tr>
<tr>
<td>Familial dilated cardiomyopathy*</td>
<td>LMNA</td>
<td>TTN</td>
<td>Roncarati et al. (2013)</td>
</tr>
<tr>
<td>Dent’s disease</td>
<td>CLCN5</td>
<td>OCRL</td>
<td>Addis et al. (2013)</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>SOD1</td>
<td>CNTF</td>
<td>Giess et al. (2002)</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>VAPB</td>
<td>C9orf72</td>
<td>van Blitterswijk et al. (2012b)</td>
</tr>
<tr>
<td>Dravet syndrome</td>
<td>PCDH19</td>
<td>TSPYL4</td>
<td>Kwong et al. (2012)</td>
</tr>
<tr>
<td>Dravet syndrome*</td>
<td>SCN9A</td>
<td>SCN1A</td>
<td>Singh et al. (2009)</td>
</tr>
<tr>
<td>Dravet syndrome*</td>
<td>CACNA1A</td>
<td>SCN1A</td>
<td>Ohmori et al. (2013)</td>
</tr>
<tr>
<td>Severe myoclonic epilepsy</td>
<td>CACNB4</td>
<td>SCN1A</td>
<td>Ohmori et al. (2008)</td>
</tr>
<tr>
<td>Severe myoclonic epilepsy</td>
<td>POLG</td>
<td>SCN1A</td>
<td>Bolszak et al. (2009)</td>
</tr>
<tr>
<td>Progressive external ophthalmoplegia</td>
<td>POLG</td>
<td>SLC25A4</td>
<td>Galassi et al. (2008)</td>
</tr>
<tr>
<td>Bartter syndrome</td>
<td>CLCNKA</td>
<td>CLCNKB</td>
<td>Nozu et al. (2008)</td>
</tr>
<tr>
<td>Chronic pancreatitis</td>
<td>SPINK1</td>
<td>CASR or CFTR or CTRC or PRSS1</td>
<td>Felderbauer et al. (2003), Masson et al. (2007), Tzetis et al. (2007), Schneider et al. (2011), LaRusch et al. (2012) and Rosendahl et al. (2013)</td>
</tr>
<tr>
<td>Oculocutaneous albinism</td>
<td>OCA2</td>
<td>TYRP1 or SLC45A2 or TYR</td>
<td>Chiang et al. (2008) and Wei et al. (2013)</td>
</tr>
<tr>
<td>Oculocutaneous albinism</td>
<td>TYR</td>
<td>SLC45A2</td>
<td>Wei et al. (2013)</td>
</tr>
<tr>
<td>Cystinuria</td>
<td>SLCS3A1</td>
<td>SLCL7A9</td>
<td>Font-Llitjós et al. (2005)</td>
</tr>
<tr>
<td>Transposition of the great arteries</td>
<td>ZIC3</td>
<td>FOXH1 or NKX2-5</td>
<td>De Luca et al. (2010)</td>
</tr>
<tr>
<td>Congenital heart disease</td>
<td>MYH6</td>
<td>NKX2-5 or GATA4</td>
<td>Granados-Riveron et al. (2012)</td>
</tr>
<tr>
<td>Charcot–Marie–Tooth disease*</td>
<td>PMP22</td>
<td>ABCD1 or LITAF</td>
<td>Meggouh et al. (2005), Hodapp et al. (2006)</td>
</tr>
<tr>
<td>Charcot–Marie–Tooth disease</td>
<td>GJB1</td>
<td>EGR2</td>
<td>Chung et al. (2005)</td>
</tr>
<tr>
<td>Charcot–Marie–Tooth disease*</td>
<td>GDAP1</td>
<td>MFN2</td>
<td>Vital et al. (2012)</td>
</tr>
</tbody>
</table>
reductase deficiency (Draper et al. 2003). It would not be altogether surprising, in conditions where digenic or even trigenic inheritance has been reported, if the individual component mutations were found to exhibit a reduced clinical penetrance as compared to mutations underlying the monogenic forms of the disease.

In some disorders, incomplete penetrance of a particular mutation can be due to the oligogenic nature of the disease and hence to the requirement for multiple genes to be mutated for the condition in question to manifest. An inherited predisposition to cancer can be monogenic, but is also very likely to have an oligogenic aetiology in many instances (Fearnhead et al. 2004; Koren-Michowitz et al. 2005; Okkels et al. 2006; Küry et al. 2008; Wasielewski et al. 2010; Martinez and Kolodner 2010; Plon et al. 2011; Morak et al. 2011; Gracia-Aznarez et al. 2013). In

<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refractory auto-inflammatory syndrome</td>
<td>TNFRSF1A</td>
<td>CIAS1</td>
<td>Tuitou et al. (2006)</td>
</tr>
<tr>
<td>Short-rib polydactyly syndrome type 2</td>
<td>NEK1</td>
<td>DYNC2H1</td>
<td>Thiel et al. (2011)</td>
</tr>
<tr>
<td>Maturity-onset diabetes of the young</td>
<td>HNF1A</td>
<td>HNF1B</td>
<td>Karges et al. (2007)</td>
</tr>
<tr>
<td>Maturity-onset diabetes of the young</td>
<td>HNF1A</td>
<td>HNF4A</td>
<td>Forlani et al. (2010) and Shankar et al. (2013)</td>
</tr>
<tr>
<td>Polycystic kidney disease*</td>
<td>PKD1</td>
<td>PKD2</td>
<td>Pei et al. (2001) and Dedoussis et al. (2008)</td>
</tr>
<tr>
<td>Hyperimmunoglobulinemia D and periodic fever syndrome</td>
<td>MKV</td>
<td>TNFRSF1A</td>
<td>Hoffmann et al. (2005)</td>
</tr>
<tr>
<td>Obesity, hyperinsulinaemia and insulin resistance</td>
<td>TCF1</td>
<td>NR0B2</td>
<td>Tonooka et al. (2002)</td>
</tr>
<tr>
<td>Neuronal ceroid lipofuscinosis</td>
<td>POLG</td>
<td>CLN5</td>
<td>Staropoli et al. (2012)</td>
</tr>
<tr>
<td>Chronic lung disease</td>
<td>SFTPC</td>
<td>ABCA3</td>
<td>Bullard and Nogee (2007)</td>
</tr>
<tr>
<td>Lafora disease</td>
<td>EPM2B</td>
<td>PPP1R3C</td>
<td>Guerrero et al. (2011)</td>
</tr>
<tr>
<td>Congenital erythropoietic porphyria</td>
<td>UROS</td>
<td>ALAS2</td>
<td>To-Figuera et al. (2011)</td>
</tr>
<tr>
<td>Familial venous thrombosis*</td>
<td>PROC</td>
<td>PROS1</td>
<td>Formstone et al. (1996), Brenner et al. (1996), Boinot et al. (2003), Knoll et al. (2001) and Hayashida et al. (2003)</td>
</tr>
<tr>
<td>Familial venous thrombosis*</td>
<td>PROC</td>
<td>SERPIND1</td>
<td>Bernardi et al. (1996)</td>
</tr>
<tr>
<td>Breast cancer*</td>
<td>BRCA1</td>
<td>BRCA2</td>
<td>Leegte et al. (2005), Lavie et al. (2011) and Heidemann et al. (2012)</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>BRCA1</td>
<td>PALB2</td>
<td>Pern et al. (2012)</td>
</tr>
<tr>
<td>Multiple tumours of different types</td>
<td>BRCA1</td>
<td>MLH1</td>
<td>Pedroni et al. (2013)</td>
</tr>
<tr>
<td>Familial pulmonary arterial hypertension</td>
<td>BMPR2</td>
<td>THBS1</td>
<td>Maloney et al. (2012)</td>
</tr>
<tr>
<td>Hereditary nonpolyposis colorectal cancer</td>
<td>MUTYH</td>
<td>MSH6</td>
<td>Van Puijenbroek et al. (2007) and Giráldez et al. (2009)</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>EPCAM</td>
<td>MSH2</td>
<td>Li-Chang et al. (2013)</td>
</tr>
<tr>
<td>Colorectal cancer, juvenile onset*</td>
<td>APC</td>
<td>MSH2</td>
<td>Uhrhammer and Bignon (2008)</td>
</tr>
<tr>
<td>Autoimmune lymphoproliferative syndrome</td>
<td>FAS</td>
<td>CASP10</td>
<td>Cerutti et al. (2007)</td>
</tr>
<tr>
<td>Autoimmune lymphoproliferative syndrome*</td>
<td>FAS</td>
<td>PRF1</td>
<td>Clementi et al. (2004)</td>
</tr>
<tr>
<td>Steroid-resistant focal segmental glomerulosclerosis</td>
<td>NPHS2</td>
<td>NPHS1 or CD2AP</td>
<td>Lówik et al. (2008)</td>
</tr>
<tr>
<td>Severe infantile liver disease</td>
<td>AKR1D1</td>
<td>SKIV2L</td>
<td>Morgan et al. (2013)</td>
</tr>
<tr>
<td>Ataxia, dementia and hypogonadotropism</td>
<td>RNF216</td>
<td>OTUD4</td>
<td>Margolin et al. (2013)</td>
</tr>
<tr>
<td>Paediatric inflammatory bowel disease</td>
<td>NOD2</td>
<td>GSDMB or ZNF365 or ERAP2 or SEC16A or GMPBB</td>
<td>Christodoulou et al. (2013)</td>
</tr>
<tr>
<td>Paediatric inflammatory bowel disease</td>
<td>BACH2</td>
<td>IL10</td>
<td>Christodoulou et al. (2013)</td>
</tr>
<tr>
<td>Hutchinson-Gilford progeria syndrome</td>
<td>ZMPSTE24</td>
<td>LMNA</td>
<td>Denecke et al. (2006)</td>
</tr>
</tbody>
</table>

In this table, we considered only those examples of digenic mutations that are unlikely to be merely coincidental and which are predicted to affect genes that are both functionally associated with the disease in question.

* Patients, with heterozygous mutations affecting two different genes, exhibiting earlier disease onset or a more severe clinical phenotype than either of their singly heterozygous parents or their siblings.
amyotrophic lateral sclerosis, van Blitterswijk et al. (2012a) detected FUS and TARDBP mutations in combination with ANG mutations, and C9orf72 repeat expansions with TARDBP, SOD1 and FUS mutations. At least five relatively common polymorphisms in four different genes, CFB (Arg32Gln, rs641153), C2 (Glu318Asp, rs9332739), CFH [Tyr402His (rs1061170) and non-coding variant rs1410996] and ARMS2 (Ala69Ser, rs10490924), interact so as to confer increased risk of age-related macular degeneration (Maller et al. 2006). Common polymorphisms in the LEPR (Gln223Arg, rs1137101) and ADRB2 [Arg16Gly (rs1042713) and Gln27Glu (rs1042714)] genes jointly confer increased risk of obesity even though none of these polymorphisms exhibits a significant influence on their own (Pereira et al. 2011). Other examples of oligo-genic inheritance, involving the mutation or polymorphism of multiple unlinked genes in the same individual, include isolated gonadotropin-releasing hormone deficiency (KAL1, PROK2 and NELF; Sykiotis et al. 2010), hypertrophic cardiomyopathy (MYH7, MYBPC3, TNNI3 and TNNT2; Girolami et al. 2010; Lopes et al. 2013), iminoglycinuria (SLC36A2, SLC6A20, SLC6A18, SLC6A19; Bröer et al. 2008), long QT syndrome (KCNH2, SCN5A and KCNE1; Yoshikane et al. 2013), chronic pancreatitis (SPINK1, CFTR and CTRC; Rosendahl et al. 2013), atypical haemolytic uraemic syndrome (CFH, CD46 and CFI; Roumenina et al. 2012; Bresin et al. 2013), Parkinson disease (LRRK2, SNCA, MAPT, GBA, BST1, PARK16; Wang et al. 2012), acrocallosal syndrome (KIF7, AH11, BBS2 and BBS4; Walsh et al. 2013), autism spectrum disorders (Neale et al. 2012) and a low plasma level of HDL cholesterol, a major risk factor for atherosclerosis (Cohen et al. 2004; Wang et al. 2008; Johansen and Hegele 2012).

One of the best characterized oligogenic disorders is Bardet–Biedl syndrome (BBS) where at least 17 genes are known to contribute to the clinical phenotype, and the severity of the disease phenotype may vary as a result of the interaction of mutations in different BBS genes. Although mutations at more than one locus have often been found to segregate with the disease, thereby modulating both its penetrance and expressivity (Badano et al. 2006; Zaghloul and Katsanis 2009; Cardenas-Rodriguez et al. 2013), the jury is still out in relation to claims of digenic triallelic inheritance in BBS (Katsanis et al. 2001; Beales et al. 2003; Badano et al. 2003; Fauser et al. 2003; Smaoui et al. 2006; Chen et al. 2011; Abu-Safieh et al. 2012). This notwithstanding, some common BBS gene variants appear to be detrimental to protein function and may well interact with the much rarer pathogenic BBS mutations so as to influence the severity of the BBS phenotype (Zaghloul et al. 2010).

Some idea of the likely complexity of genotype–phenotype relationships in complex disease has come from the comparative exome sequencing of 237 ion channel genes performed in sporadic idiopathic epilepsy patients and unaffected controls (Klassen et al. 2011). Both rare and common variants were identified in the two groups. Although these variants were more numerous in the patient group, they were not found to be predictive of disease: as the authors opined, “absolute numerical counts of SNP burden hold little predictive value as a global pathogenic measure”. However, 51 % of cases and 14 % of controls had ≥2 non-synonymous variants in their sodium channel genes. In similar vein, 24 % of cases and 6 % of controls had ≥2 non-synonymous variants in their GABA receptor alpha genes (Klassen et al. 2011). Such findings are quite consistent with an oligogenic model of sporadic epilepsy (Dibbens et al. 2007). What is required here, however, is not a simple variant number count, but rather an assessment, on an individual basis, of the net effect of an oligogenic variant profile (requiring ascertainment of the various gains or losses of function associated with specific variants and computed with an eye to the nature of potential joint effects) on a clinically or phenotypically relevant output measure such as the electrical signature of a cell type or brain region.

Perhaps, the best characterized oligogenic disorder to date is familial venous thrombosis. The risk of venous thromboembolism is known to be increased in patients who carry more than one genetic variant disrupting the 100+ genes of the ‘hemostaseome’ (Fechtel et al. 2011). Thus, 19 % of symptomatic individuals harbouring a protein C (PROC) gene mutation were also found to be heterozygous for factor V Leiden (F5 Arg534Gln; Koeleman et al. 1994), a functional polymorphism which occurs at a frequency of 2–5 % in European populations. In a replication study, 9.5 % of venous thrombosis patients were found to carry both mutations (Gandrille et al. 1995) suggesting that their co-occurrence increases the likelihood of their coming to clinical attention. Similar findings have been noted in families with protein S (PROS1) deficiency; among symptomatic individuals, 38 % also carried the factor V Leiden mutation (Koeleman et al. 1995). Likewise, co-inheritance of antithrombin (SERPINC1) deficiency and factor V Leiden not only increases clinical penetrance, but also reduces the age of clinical presentation (van Boven et al. 1996). In a larger-scale study involving 132 thrombophilic families, the risk of thrombosis was increased and the age of onset lowered in cases of double heterozygosity for two gene variants (combinations of variants in PROC, PROS1, F5 and F2) as compared to individuals carrying single variants of these genes (Tirado et al. 2001). ABO blood group is also known to modify the risk of venous thrombosis in individuals with hereditary thrombophilia through an influence on the plasma levels of factor VIII and the factor VIII carrier protein, von Willebrand factor (Tirado et al. 2005; Nossent et al. 2006; Cohen et al. 2012). Such studies provide strong circumstantial
support for the joint impact of multiple mutations in thrombotic disease (Martinelli et al. 2008). The elevation of risk for each individual variant is, however, low, and incomplete penetrance is evident for all prothrombotic variants. This means that the vast majority of individuals bearing these variants do not suffer from thrombotic disease. It is nevertheless reasonable to suppose that patients with recurrent venous thrombosis will tend to have a greater number of prothrombotic variants than those who have experienced a single thrombotic event, with those individuals who never experienced thrombosis harbouring even fewer prothrombotic variants (Fechtel et al. 2011). Since the number of genes known to influence haemostasis is large and the number of variants with potential impact larger still, we may expect that a substantial number of different variant combinations will be capable of conferring an increased risk. This genetic risk will accompany each prothrombotic challenge (such as pregnancy, long haul air travel, contraceptive pill usage and immobilization after surgery), with the greatest risk of venous thrombosis accruing to those possessing the largest number of prothrombotic variants in their genomes (Fechtel et al. 2011). Our task is to come to understand how specific DNA sequence changes in the large number of genes known to play a role in haemostasis and thrombosis act either synergistically or antagonistically so as to confer disease predisposition upon the individual, thereby influencing the clinical penetrance, by shifting their haemostatic balance towards either a prothrombotic or anticoagulant phenotype (Franchini and Mannucci 2009; Westrick and Ginsburg 2009; Fechtel et al. 2011).

**Influence of sex on penetrance**

The sex dependence of the penetrance of inherited mutations has been reported in a variety of different heritable disorders including haemochromatosis (*HFE*; Rossi et al. 2008), hypertrophic cardiomyopathy (*MYBPC3, MYH7*; Michels et al. 2009; Page et al. 2012), arrhythmogenic right ventricular dysplasia/cardioiomyopathy (*PKP2; Dalal et al. 2006*), long QT syndrome (* KCNQ1, KCNH2, SCN5A*; Zareba et al. 2003), hypokalaemic periodic paralysis (*CACNA1S; Kawamura et al. 2004; Li et al. 2012; SCN4A; Ke et al. 2013*), familial pulmonary arterial hypertension (*BMPR2; Austin et al. 2009b*), hereditary spastic paraplegia (*SPAST; Mitne-Neto et al. 2007*), hereditary dystonia/dopa-responsive dystonia (*GCH1; Furukawa et al. 1998*), cardiac disease (*LMNA*; van Rijsingen et al. 2013), Hirschsprung disease (*RET; Emison et al. 2005*), autism spectrum disorder (*SHANK1; Sato et al. 2012*), amyotrophic lateral sclerosis (*C9ORF72; Le Ber et al. 2013; Williams et al. 2013*), and familial obesity (*SHP; Yang et al. 2010*). A male-biased effect on the penetrance of duplications and deletions at 16p13.11 is evident in a range of neurodevelopmental conditions including autism, attention deficit hyperactivity disorder, intellectual disability and schizophrenia (Tropeano et al. 2013). In the case of familial pulmonary arterial hypertension, both genetic and metabolic marker data were consistent with a modifying role for variation in oestrogens and/or oestrogen metabolism upon disease risk (Austin et al. 2009b). The low penetrance of hypokalaemic periodic paralysis due to *SCN4A* mutations in females is also likely to be due to the effect of oestrogens (Ke et al. 2013).

Allelic variation may also influence the clinical phenotype in a sex-specific fashion. Thus, Lahtinen et al. (2011) reported that the common * KCNE1* Asp85Asn (rs1805128) polymorphism was associated with a QT-interval prolongation in male but not female type 1 long QT syndrome patients harbouring the * KCNQ1* Gly589Asp mutation. * KCNE1* Asp85Asn may thus be a sex-specific QT-interval modifier in type 1 LQTS. Similarly, the Ile148Met (rs738409) *PNPLA3* polymorphism is a disease modifier in primary sclerosing cholangitis with bile duct stenosis, but only in male patients (Friedrich et al. 2013). Finally, a Val89Leu polymorphism (rs523349) in the steroid 5α-reductase type 2 ( *SRD5A2*) gene, which serves to reduce the conversion of testosterone to dihydrotestosterone, has been claimed to influence the severity of post-traumatic stress symptoms but in a male-specific fashion (Gillespie et al. 2013).

An intriguing parent-of-origin effect has been noted in two apparently unrelated retinoblastoma families with a heterozygous, low-penetrance splice site mutation (c.607+1G>T) in the *RB1* gene which causes skipping of exon 6 (Klutz et al. 2002). The abundance of the resulting nonsense (frameshifted) *RB1* mRNA relative to the wild-type was found to vary between members of one and the same family. Those individuals in family #1 who inherited the mutant *RB1* allele from their mother displayed a similar level of nonsense and wild-type *RB1* transcripts, and only one of eight carriers developed retinoblastoma. By contrast, those individuals in family #2 who inherited the mutant *RB1* allele from their fathers displayed a reduced abundance of the nonsense transcript with six of eight carriers developing retinoblastoma, indicating that the mutant transcript has residual function. Assuming that this is not a chance result (Fisher’s exact test: *p* = 0.04), it may be that the gender of the transmitting parent can influence the penetrance of the pathogenic mutation.

There is good evidence to suggest that sex-specific genomic architecture can influence the expression of human phenotypes, including disease traits (Ober et al. 2008). It is likely that the underlying mechanism is differential gene regulation in males and females, particularly in relation to sex steroid-responsive genes (Zhang et al. 2007; Dimas et al. 2012).

Another mechanism by which sex influences penetrance is via genomic imprinting. Genomic imprinting results
from the epigenetic modification of a gene or gene region that leads to the mutually exclusive expression of either the maternal or the paternal allele. Imprinted alleles are silenced (by DNA methylation or histone modification), so that the corresponding genes are expressed only from the non-imprinted allele inherited from the other parent. In the case of disease genes, imprinting can influence the penetrance of pathological mutations depending upon whether the wild-type or the mutant allele is imprinted. Genomic imprinting can give rise to markedly different levels of clinical penetrance depending upon the parental origin of the disease allele. Examples include *SGCE* mutations in myoclonus dystonia (Zimprich et al. 2001; Müller et al. 2002; Grabowski et al. 2003) and *SDHD* mutations in paraganglioma (Badenhop et al. 2001; Simi et al. 2005; Baysal et al. 2011). In both cases, maternal imprinting ensures that the pathologically effective mutations are almost invariably inherited from the father. Intriguingly, in one family with pseudohypoparathyroidism type 1b, Jan de Beur et al. (2003) reported a case of the incomplete penetrance of an imprinting mutation. These authors found that both the clinically affected and unaffected siblings had inherited the same *GNAS1* allele from their affected mother, indicating that some dissociation must have occurred between the genetic *GNAS1* defect responsible for the disease and its epigenetic mark. The inconsistent acquisition of a paternal epigenotype on a maternal *GNAS1* allele would appear to provide evidence for the incomplete expression of a reprogramming defect that affects imprinting.

**Age-dependent penetrance**

Age-dependent penetrance is present if the clinical symptoms of a given disease are increasingly likely to manifest themselves with increasing age of the at-risk individual. Age-dependent penetrance has been reported for mutations in a wide variety of different human disease genes, e.g. *MYBPC3* in hypertrophic cardiomyopathy (Michels et al. 2009; Page et al. 2012), *LMNA* in Emery–Dreifuss muscular dystrophy (Vytopil et al. 2002), *MC4R* in familial obesity due to melanocortin-4 receptor deficiency (Stuttman et al. 2008), *GBA* in Parkinson disease (Anheim et al. 2012; Rana et al. 2013), *BRCA1* and *BRCA2* in breast cancer susceptibility (Chen and Parmigiani 2007; Al-Mulla et al. 2009; Mavaddat et al. 2013), *MEN1* in multiple endocrine neoplasia type 1 (Machens et al. 2007), *RET* in multiple endocrine neoplasia type 2A (Frank-Raue et al. 2011) and *SDHD* and *SDHB* in predisposition to paragangliomas (Hensen et al. 2010; Hes et al. 2010). The *APOE* ε4 allele (comprising the T allele of rs429358 and the C allele of rs7412 in *cis*) serves to reduce the age of onset of Alzheimer disease from 78.4 years in patients lacking the allele, to 75.3 in heterozygous carriers to 72.9 in carriers of two *APOE* ε4 alleles (Sando et al. 2008).

Age-dependent penetrance is particularly evident where large numbers of heterozygous carriers harbouring specific gene mutations have been identified by cascade screening, e.g. *LRRK2* Gly2019Ser (rs34637584) in Parkinson disease (Latourelle et al. 2008; Healy et al. 2008; Sierra et al. 2011), *GLUT1* Arg232Cys in familial idiopathic generalized epilepsy (Striano et al. 2012), *RET* Cys634Trp (rs77709286) in multiple endocrine neoplasia type 2A (Milos et al. 2008), *ACADM* Lys329Glu (rs77931234) in medium-chain acyl-CoA dehydrogenase deficiency (Andresen et al. 2012), *PKP2* Gln59Leu in arrhythmogenic right ventricular cardiomyopathy (Lahtinen et al. 2008) and *MYBPC3* c.2308+1G>A (rs112738974) in hypertrophic cardiomyopathy (Oliva-Sandoval et al. 2010). However, there are always anomalous cases; thus, in a family segregating a pathogenic missense mutation (Arg1205His) in the vacuolar protein sorting 35 (VPS35) gene, six family members between the ages of 54 and 73 years exhibited signs of Parkinson disease, but one individual was still asymptomatic at age 86 (Nuytemans et al. 2013).

Specific mutations may sometimes differ from each other in terms of the average age of onset of clinical symptoms. Thus, for example, patients with maturity-onset diabetes of the young (MODY) who harbour mutations in exons 9 or 10 of the *HNF4A* gene have been found to develop disease much later (average 40 vs. 24 years) than MODY patients with mutations in exons 2–8 (Harries et al. 2008). This difference in age-related penetrance is thought to be a consequence of the exon 9 and 10 mutations being absent from three of the nine HNF4A isoforms encoded by the *HNF4A* gene, whereas the mutations located in exons 2–8 affect all nine isoforms.

In some cases, the clinical penetrance of a particular mutation can change quite dramatically with age. For example, the cumulative incidence among carriers of the Arg1441Gly mutation in the *LRRK2* gene causing Parkinson disease was found to be 12.5 % until the age of 65 years, but 83 % until age 80 (Ruiz-Martínez et al. 2010). However, the penetrance of the common *TTR* Val30Met mutation causing autosomal dominant familial amyloid polyneuropathy has been estimated to be 1.7 % until the age of 30 years, 22 % until the age of 60, but still only 69 % until age 90 (Hellman et al. 2008). Majounie et al. (2011) showed that the pathogenic GGGGCC hexanucleotide expansion in the C9orf72 gene associated with a high proportion of cases of amyotrophic lateral sclerosis and frontotemporal dementia was non-penetrant in individuals younger than 35 years, 50 % penetrant by age 58 but almost fully penetrant by age 80. Age-dependent penetrance could thus provide another explanation for why some putatively pathological mutations listed in HGMD
are present in apparently healthy individuals from the 1000 Genomes Project.

A glimpse of the way ahead is provided by a recent study of symptomatic and asymptomatic carriers of a specific granulin (GRN) mutation (Thr272Ser) responsible for autosomal dominant frontotemporal lobar degeneration (FTLD), a disease whose onset typically occurs in the sixth decade of life (Milanesi et al. 2013). Unsurprisingly, both the symptomatic and asymptomatic GRN mutation carriers had lower serum levels of progranulin than non-carriers. However, using whole-genome microarray analysis, the leukocyte expression of the TMEM40 and LY6G6F genes was found to be significantly higher in FTLD patients harbouring GRN mutations as compared to asymptomatic carriers. Further, elevated expression of the genes was correlated with increased brain damage and could therefore be directly related to the pathology of the disease (Milanesi et al. 2013).

Epigenetic influences on disease penetrance

As briefly discussed above in the context of the influence of gender upon penetrance, epigenetic modifications may also account for incomplete penetrance. Thus, when monozygotic twins are discordant for disease phenotypes, epigenetic differences should be considered (Wong et al. 2005; Kaminsky et al. 2009; Gordon et al. 2011; 2012). Indeed, monozygotic twins have been reported who differ both in relation to a specific clinical phenotype and in terms of an epigenotype. For example, monozygotic twins discordant for childhood leukaemia have been found to have discordant BRCA1 methylation status (Galetzka et al. 2012). Similarly, hypermethylation of SLC6A4, encoding the serotonin transporter, has been reported in one member of a monozygotic twin pair discordant for bipolar disorder (Sugawara et al. 2011).

Epigenetic differences may also contribute to incomplete penetrance in other conditions such as asthma, where DNA methylation has been reported to modulate the risk of disease conferred by genetic variants at the zona pellucida binding protein 2 (ZPBP2; Berlivet et al. 2012), forkhead box P3 (FOXP3; Runyon et al. 2012), interferon-γ (IFNG; Runyon et al. 2012) and interleukin-4 receptor (IL4R; Soto-Ramirez et al. 2013) gene loci.

A special case of imprinting is provided by X-inactivation (Dobyns et al. 2004). When a disease gene is X-linked, skewed X-inactivation can cause variable penetrance of pathogenic mutations in female carriers (Van den Veyver 2001). Examples of this phenomenon involve mutations in the TIMM8A gene (Xq22.1) in dystonia-deafness syndrome (Plenge et al. 1999), the EBP gene (Xp11.23) in X-linked dominant chondrodysplasia punctata (Shirahama et al. 2003), the FLNA gene (Xq28) in otopalatodigital type 1 syndrome (Hidalgo-Bravo et al. 2005), the ABCD1 gene (Xq28) in a family with X-linked adrenoleukodystrophy (Wang et al. 2013b) and the ZIC3 gene (Xq26.3) in a family with complex heart defects (Chhin et al. 2007). It should, however, be pointed out that some ZIC3 mutations are characterized by reduced penetrance in males, a finding that cannot be explained by skewed X-inactivation (Mégarbané et al. 2000).

Gene–environment interactions and penetrance

The environment, in its broadest sense, will often influence clinical penetrance, either ameliorating or exacerbating the impact of heritable genetic variants (Hunter 2005). Indeed, environmental modifiers of disease penetrance (e.g. diet, alcohol intake, drugs, metabolic syndrome) have long been known to influence the penetrance of HFE C282Y homozygosity in haemochromatosis (Beutler 2003; Rossi et al. 2008; Deugnier and Mosser 2008).

One way to explore the relative contribution of genes and environment is by studying monozygotic twins harbouring the same pathogenic mutation(s) and sharing the same genetic background. Although the vast majority of such monozygotic twin pairs have been found to be concordant in terms of their clinical phenotypes (e.g. Miesfeldt et al. 1998; Munhoz et al. 2008; McDade et al. 2012), others are quite discordant (Matsuo et al. 2000; Amann et al. 2001; Martin et al. 2003; Holmgren et al. 2004; Lachmann et al. 2004; Czlonkowska et al. 2009; Biegstraaten et al. 2011; Fencel et al. 2012; Iatropoulos et al. 2012), suggesting that the environment can often play an important role in determining both the penetrance and expressivity of pathological mutations. (It should be borne in mind that there are various alternative genetic explanations for discordant phenotypes in monozygotic twins, including de novo post-zygotic mutation (Kentsis et al. 2009; Vogt et al. 2011), compensatory mutations (Mankad et al. 2006) and somatic copy number variation (Bruder et al. 2008) that obviate the need for a major contribution from the environment, as well as acquired epigenetic differences (Galetzka et al. 2012; Bennett et al. 2008)).

An environmental influence on penetrance is perhaps at its most evident in cancer susceptibility (Houlston and Peto 2004; Shen 2009). Indeed, an environmental component is very important in colorectal cancer where inherited genetic variants at a number of different loci interact primarily with dietary variables and overweight to confer risk (Hutter et al. 2012; Siegert et al. 2013). In similar vein, inherited differences in skin pigmentation influence the risk of melanoma, but this risk is further modified both by latitude of habitation and lifestyle choices (van der Velden et al. 2001; Bishop et al. 2002; Begg et al. 2005; Meyle and
Parity and breast feeding are both known to be modifiers of risk of breast/ovarian cancer in BRCA1 mutation carriers (McLaughlin et al. 2007; Jernström et al. 2004; Cullinane et al. 2005; Antoniou et al. 2006). Another example of a gene–environment interaction in the context of cancer is provided by cytochrome P450 gene variants that may influence cancer risk by virtue of their roles in xenobiotic metabolism, detoxification of carcinogens, and to a lesser extent the bioactivation of procarcinogens (Rodriguez-Hung et al. 20110). Lung cancer provides an excellent example of the interaction of genes and environment. Three different GWAS, published virtually simultaneously, provided the first convincing evidence for an association between heritable genetic variation at the nicotinic acetylcholine receptor CHRNA5/CHRNA3/CHRNA4 locus on chromosome 15q25.1 and lung cancer (Amos et al. 2008; Hung et al. 2008; Thorgerisson et al. 2008). Although allele T of SNP rs1051730, a synonymous variant located within exon 5 of the CHRNA3 gene, was found to be strongly associated with smoking quantity, the issue of whether the association with lung cancer was direct or indirect (i.e. mediated through cigarette smoking and nicotine dependence) remained unclear. Galvan and Dragan (2009) performed a meta-analysis of reported studies of the 15q25 region and found that this locus was not associated with lung cancer risk in never-smokers. This lack of effect argued for an indirect effect of genetic variation at the 15q25 locus on lung cancer risk via an association between these variants and smoking/nicotine dependence. However, Wang et al. (2010b) subsequently examined the relationship between rs1051730 and lung cancer and concluded that, in addition to its indirect influence on disease risk (through smoking behaviour), this variant also exerted a rather larger (and direct) effect. Kaur-Knudsen et al. (2011) concurred, demonstrating that homozygosity for rs1051730 was associated with a smoking behaviour-adjusted relative risk of lung cancer of 1.6, indicating that rs1051730 is associated with an additional risk of lung cancer over and above that derived from its effect on smoking behaviour. Finally, in a lung cancer case–control study, VanderWeele et al. (2012) employed two 15q25.1 SNPs, rs8034191 and rs1051730, to show that the proportion of increased risk due to smoking was only 3.2 % and that the association of the 15q25 variants with lung cancer operates primarily through pathways other than smoking behaviour. All of the above notwithstanding, the risk of lung cancer conferred directly or indirectly by genetic variants on 15q25 would be small if the individual concerned simply opted not to smoke (Brennan et al. 2011).

The penetrance of genetic variants conferring susceptibility to infectious disease is clearly contingent upon exposure to the specific pathogens concerned (Vannberg et al. 2011; Chapman and Hill 2012). One example is the CCR5 32-bp (c.554del32) deletion which is associated with a lower rate of HIV infection and a delay in the onset of AIDS (Smith et al. 1997). Sex may also play a role in some conditions; thus, in multiple sclerosis, women appear to be more responsive to the environmental risk factors that cause the disease (Goodin 2012; O’Gorman et al. 2012).

Diet is also an important modifier of clinical penetrance. Thus, an inherited predisposition to obesity (exemplified by the association between dietary fat intake and obesity in carriers of the PPARG2 Pro12Ala allele; Memisoglu et al. 2003) is in principle modifiable by diet (Walters et al. 2010; Ramachandrappa and Farooqi 2011). Similarly, the impact of genetic variants at the FTO locus on risk of obesity can be attenuated by physical activity (Kilpeläinen et al. 2011). Diet is also an important modifier of clinical penetrance in phenylketonuria, as mentioned in the “Introduction” to this review, where the penetrance of the condition can be very substantially reduced by restricting dietary phenylalanine (van Spronsen 2010).

Heavy coffee drinkers have been known for some time to have a reduced risk of developing Parkinson disease. However, the risk of developing Parkinson disease has been found to be reduced even further for heavy coffee drinkers by a specific variant in the GRIN2A gene; compared to light coffee drinkers with an rs4998386_CC genotype, heavy coffee drinkers with the same genotype have an 18 % lower risk, whereas heavy coffee drinkers with an rs4998386_TC genotype have a 59 % lower risk (Hamza et al. 2011).

More unusually, altitude has been reported to act as a modifier of the phenotypic severity of hereditary paraganglioma type 1 caused by mutations in the succinate dehydrogenase D (SDHD) gene (Astrom et al. 2003). Since chronic hypoxic stimulation at high altitude causes sporadic carotid body paragangiomas, Aström et al. (2003) proposed that SDHD might be involved in oxygen sensing. Thus, whilst SDHD mutations could impair oxygen sensing, low altitude may serve to reduce the penetrance of these mutations.

The clinical penetrance of psychological disorders and traits has long been known to be strongly influenced by gene–environment interactions (Dick 2011). For example, a 44-bp deletion/insertion polymorphism in the promoter region of the serotonin transporter gene SLC6A4 was reported to be associated with depression after stressful life events (Caspi et al. 2003). Recently, Klengel et al. (2013) gave us a glimpse of the likely complexity of the mechanisms underlying gene–environment interactions in the context of psychological disorders. These authors demonstrated that a risk variant for post-traumatic stress disorder and major depression in the FK506-binding protein 5 gene (FKBP5) is demethylated in several cell types in children exposed to trauma. This demethylation persists into
adulthood and confers an increased risk of developing disease. In \textit{FKBP5} risk allele carriers, excessive cortisol release during early life stress leads to demethylation within the glucocorticoid-responsive elements of \textit{FKBP5} in intron 7 with the consequence of long-lasting disruption of the ultra-short feedback loop that balances \textit{FKBP5} and glucocorticoid receptor activity, causing dysregulation of the stress hormone system. The \textit{FKBP5} risk allele corresponds to a functional polymorphism located in intron 2 of the \textit{FKBP5} gene that alters the chromatin interaction between the transcriptional start site and long-range enhancers, thereby increasing the transcriptional activity of \textit{FKBP5} over and above that of the wild-type allele. Only the risk allele is able to form a three-dimensional complex which includes RNA polymerase II and a glucocorticoid-responsive element located within intron 7 of \textit{FKBP5}. Enhanced transcription of the risk allele facilitates the PolII-dependent demethylation in intron 7 in response to elevated glucocorticoids under early life stressful conditions. The reduced methylation of intron 7 CpGs leads to increased induction of \textit{FKBP5} by glucocorticoid receptor activation, especially in risk allele carriers, representing an enhancement of the ultra-short feedback loop leading to increased glucocorticoid receptor resistance. If this occurs during developmentally critical periods, then the methylation patterns will remain stable over time (Klengel et al. 2013). Hence, the demethylation in \textit{FKBP5} intron 7 depends upon both childhood trauma and the sequence variant in intron 2, in a tripartite gene–environment interaction (Szyf 2012).

The above examples represent the tip of the iceberg because the clinical penetrance of most monogenic conditions and all complex disease is likely to be influenced by the environment in some way or another. A further glimpse of this complexity is perhaps provided by Kallberg et al. (2007) who reported interactions between \textit{HLA-DRB1} SE alleles, the A allele of the \textit{PTPN22} Arg620Trp (rs2476601) polymorphism and smoking in conferring risk of rheumatoid arthritis. Gene–environment interactions are also evident in asthma (Custovic et al. 2012; Chang et al. 2012) and eczema (Bisgaard et al. 2008) and will become increasingly apparent in other disorders as new techniques are developed to identify them (Aschard et al. 2012). To this end, mouse models are beginning to come into their own as a means to study the role of gene–environment interactions in the aetiology of human disorders; by these means, short-term gestational hypoxia has been found to increase the penetrance of vertebral defects in congenital scoliosis (Sparrow et al. 2012).

Conclusions

A holy grail of human medical genetics is to be able to deduce the likely clinical phenotype of an individual from their genotype or genomic sequence. It was once perhaps naively assumed that, at least for “monogenic” disorders, genotype–phenotype relationships would be that simple, and also fairly straightforward to discern. However, it has been clear for some time that it is inappropriate to regard such disorders as either simple or monogenic in any strict sense. Further, in many cases, the reality is that we cannot readily draw straight lines of causation from known genotypes to specific clinical phenotypes. This is because instances abound of individuals who harbour a disease-associated mutation/genotype, but who do not express certain features of the disease or who may even be asymptomatic. This phenomenon of reduced penetrance may or may not be the norm, but it is far from being a rare exception. Our appreciation of its full extent is still emerging, although some of the different mechanisms underlying reduced penetrance are now becoming apparent (Fig. 1).

It has become clear from large-scale sequencing studies that many individuals in the general population harbour...
large numbers of potentially disadvantageous variants without suffering any obvious ill effects (The 1000 Genomes Project Consortium 2010; MacArthur et al. 2012; Xue et al. 2012; Shen et al. 2013a). Thus, it would appear that many mutations are, on their own, insufficient to cause disease and need to occur in the presence of other genetic variants, either allelic or non-allelic, as well as facultative environmental factor(s), for a disease state to ensue. Indeed, many pathological mutations may only be conditionally pathogenic, exerting a detrimental effect only if and when the genetic and external environments interact to push the phenotype over some notional threshold into pathology.

Penetrance is best thought of as being a genotype-specific rather than a gene-specific or disease-specific phenomenon. Thus, in any given disease gene, some mutations may exhibit complete penetrance, whereas others may show incomplete or even quite low penetrance. Generally speaking, mutations that display low penetrance also tend to exert milder effects on the clinical phenotype and/or protein function, while the more highly penetrant a mutation is, the less frequent it is likely to be in the general population (Coventry et al. 2010; Marth et al. 2011; Gorlov et al. 2011; Tennessen et al. 2012; Nelson et al. 2012; Subramanian 2012; Fu et al. 2013). Whereas highly penetrant mutations may exert their pathogenic effects with relatively little interaction with other genetic or environmental factors, low-penetrance mutations are generally characterized by significant gene–gene and gene–environment interactions (Cordell 2009). Different combinations of such variants may contribute to the variable penetrance characteristic of both monogenic and complex disease.

Whatever the molecular basis may be in the case of a given mutation, reduced penetrance is in general likely to present a serious impediment to the implementation of any scheme designed to classify the pathological significance of human genetic variants (e.g. Plon et al. 2008; Tavtigian et al. 2008). Reduced penetrance is also likely to present problems in identifying pathological mutations in whole-genome/exome sequencing programmes unless it is explicitly built into the disease models being considered (Varga et al. 2013). Despite the complexities it introduces, it also offers hope that if we are able to identify environmental factors, drugs or other types of intervention that serve to reduce the penetrance of a given pathological variant, or alternatively delay the onset of its pathological sequelae beyond the natural lifespan of its carrier, we shall have a whole new range of therapeutic approaches at our disposal.

Human genetic variation occurs as a continuum ranging from neutral polymorphisms, through functional polymorphisms and disease susceptibility variants to true pathological mutations with high penetrance. However, in addition and as discussed above, it has become increasingly clear that our genomes contain many ‘putatively disadvantageous variants’ that are probably insufficient on their own to cause disease, but nevertheless still have the potential to contribute to pathogenesis. Since it has also become clear that many genetic disorders are not monogenic as originally supposed, but may instead involve mutations in two or more genes, we speculate that different combinations of pathological mutations with low penetrance, functional polymorphisms, disease susceptibility variants and ‘putatively disadvantageous variants’ may vary quite considerably in terms of their net functional and hence clinical effect. Such combinations are likely to exert an influence on the age of onset and/or clinical severity of the disease in question.

The rationale of genetic studies of complex phenotypes has generally relied upon either the ‘common disease, common variant (CDCV)’ hypothesis or the ‘common disease, rare variant (CDRV)’ hypothesis. The former postulates that complex phenotypes result from the cumulative effects of a number of common variants, each with a modest effect size and relatively low penetrance. The latter proposes that complex phenotypes result from multiple rare variants, each with relatively high penetrance and large effect sizes (Schork et al. 2009). It is highly likely that both rare and common alleles will contribute to complex phenotypes and so effect sizes may be expected to differ quite widely, with rare variants with large effects complemented by a large number of frequent variants with small effects. The clinical phenotypes of complex phenotypes are therefore likely to be due to individual effects of, and interaction between, multiple causative or contributory alleles, as well as non-genetic determinants.

The full relevance of digenic and oligogenic inheritance to the phenomenon of incomplete or variable penetrance remains to be elucidated. If, however, it turns out that digenic and oligogenic conditions are more common than originally anticipated, then many disease contributory variants will have evaded purifying selection, and hence those variants that in combination (but not individually) have significant pathogenic potential will not be as frequent as might be expected under the CDRV hypothesis. Since both multiple common and rare variants may be involved in conferring disease susceptibility, we are not obliged to favour either the CDCV hypothesis or the CDRV hypothesis. Moreover, in view of the likely complexity of the gene–gene interactions involved, we concur with Lupski et al. (2011) that “for a given individual, what is important to know is not only the number and location of pathogenic variants taken one at a time, but also the unique composition of his or her genome-wide mutational burden”.

Lupski et al. (2011) charted progress on the road to a unified genetic model for human disease and opined that such a model should unite categories of diseases, previously held to be distinct entities, as part of a continuum.
which would include chromosomal syndromes, genomic disorders, Mendelian traits and common diseases or complex traits. Concurring with this view, we envisage an integrated concept of genetic aetiology in which different types of mutation (from single base substitutions to copy number variants), different combinations of mutations in multiple genes (whether in homozygosity or heterozygosity), cis-acting or trans-acting modifiers, common variants, rare variants, de novo variants and even somatic variants, jointly serve to exacerbate or ameliorate a given clinical phenotype. Further, to explain the scale of reduced penetrance, we need to conceptualize clinical phenotypes as being derived, potentially at least, from the expression of different genetic variants in two or more genes. On the basis of the data collated for this review, it seems reasonable to conclude that digenic, oligogenic and polygenic influences are much more frequent than has perhaps hitherto been realized. Unravelling such influences will undoubtedly be key to understanding the molecular basis of reduced penetrance. The impact of disease genotypes may also be modified by epigenetic and environmental factors, allowing both for synergistic and antagonistic interactions resulting in highly individualized contributions to the phenotype (whether deleterious or protective) that will variably perturb the balance of specific biological pathways so as to give rise to disease.

With the advent of next-generation sequencing, very large numbers of genetic variants are being detected in individual genomes and it has been necessary to develop new algorithms to identify those variants which are of key functional/clinical importance. However, if in using these tools, we focus exclusively on single infrequent variants under the assumption that they will invariably exert their effects in splendid isolation, then there is a very real danger that we shall inadvertently exclude from consideration those more frequent variants with modest effects, blithely ignoring their potential for interaction with the rare variants. The irony would then be that, despite having the requisite mutation and polymorphism data available, the molecular basis of genotype–phenotype relationships in many inherited diseases (including, of course, the phenomenon of reduced penetrance) could still remain unintelligible. The alternative, anticipating multigenic influences on the clinical phenotypes associated with disorders traditionally regarded as being monogenic, should not only to lead to new insights into the nature of reduced penetrance, but is also likely to improve our understanding of the nature of complex disease.

Acknowledgments The authors are grateful to Peter Stenson for provision of HGMD mutation data. DNC acknowledges receipt of financial support from BIOBASE GmbH through a licence agreement with Cardiff University. CTS is supported by the Wellcome Trust (098051).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References


regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet 44:435–439


© Springer
pulmonary arterial hypertension penetrance. Circulation 126:1907–1916


de Pontual L, Pelet A, Clement-Ziza M, Trochet D, Antonarakis SE, Attie-Bitach T, Beales PL, Blouin JL, Dastot-Le Moal F, Dollfus... Springer
of retinal disease and a contributor to digenic Usher syndrome. J Clin Invest 120:1812–1823


Fechtel K, Osterbur ML, Kehler-Sawatzki H, Stenson PD, Cooper DN (2011) Delineating the Hemostaseome as an aid to individualize the analysis of the hereditary basis of thrombotic and bleeding disorders. Hum Genet 130:149–166


 Springer


functional CYP21A2 gene are primarily restricted to Q318X alleles: evidence for a founder effect. J Clin Endocrinol Metab 94:3954–3958


among relatives of HFE C282Y homozygotes identified either by population genetic testing or presenting as patients. Gut 55:554–562


Nishiguchi, KM, Rivolta, C (2012) Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. PLoS ONE 7:e41902


predominant-negative effects in arrhythmogenic right ventricular cardiomyopathy. Hum Mutat 34:697–705
by exome sequencing in a severe form of dilated cardiomyopathy. Eur J Hum Genet (in press)


Tomooka N, Tomura H, Takahashi Y, Onigata K, Kikuchi N, Horikawa Y, Mori M, Takeda J (2002) High frequency of...


Yatsenko AN, Shroyer NF, Lewis RA, Lupski JR (2001) Late-onset genotypes in hereditary hemochromatosis (HbH) associated with enlarged ventricular cardiomyopathy. J Am Coll Cardiol 38:25–26