

Saturation of the Human Phenome

Mark E. Samuels*

Centre de Recherche de Ste-Justine, 3175, Côte Ste-Catherine, Montréal QC H3T 1C5, Canada

Abstract: The phenome is the complete set of phenotypes resulting from genetic variation in populations of an organism. Saturation of a phenome implies the identification and phenotypic description of mutations in all genes in an organism, potentially constrained to those encoding proteins. The human genome is believed to contain 20-25,000 protein coding genes, but only a small fraction of these have documented mutant phenotypes, thus the human phenome is far from complete. In model organisms, genetic saturation entails the identification of multiple mutant alleles of a gene or locus, allowing a consistent description of mutational phenotypes for that gene. Saturation of several model organisms has been attempted, usually by targeting annotated coding genes with insertional transposons (*Drosophila melanogaster*, *Mus musculus*) or by sequence directed deletion (*Saccharomyces cerevisiae*) or using libraries of antisense oligonucleotide probes injected directly into animals (*Caenorhabditis elegans*, *Danio rerio*). This paper reviews the general state of the human phenome, and discusses theoretical and practical considerations toward a saturation analysis in humans. Throughout, emphasis is placed on high penetrance genetic variation, of the kind typically associated with monogenic versus complex traits.

Received on: May 08, 2010 - Revised on: June 22, 2010 - Accepted on: June 22, 2010

Keywords: Human genome, phenome, genetics, saturation mutagenesis.

INTRODUCTION

The phenome is the complete set of phenotypes resulting from genetic variation in populations of an organism [1-4]. The concept of phenomic saturation cannot be approached without a careful consideration of the meaning of phenotype, and the genotype-phenotype map. In the most general sense, the phenotype of an individual organism is the sum total of its physiology. Such a broad definition is not particularly useful. In the context of genetics, phenotypes may be defined more specifically as those physiological traits which vary measurably as a function of genomic sequence differences among individuals in a population. Thus, many aspects of physiology can be ignored depending on the type of sequence differences being studied. For example, mutations in the gene encoding the low density lipoprotein receptor (LDLR) have huge impact on plasma cholesterol, atherosclerosis, heart disease, but have no obvious or known impact on eye color or height. In contrast, mutations in the genes encoding pigment-producing enzymes in the iris may be irrelevant to heart disease. One of the advantages of forward genetics is that one begins with an unusual physiological trait observed in actual organisms (human or model animal), and proceeds to search for genetic variation that underlies that trait. In contrast, if one starts with a gene of interest, and mutates it in a directed way (or even through screening undirected libraries of mutants), one has no idea what the physiological effects might be so one is obliged to explore all possible phenotype consequences – a time consuming and difficult task with animals that cannot tell us how they are

feeling! Both approaches have been employed in model studies, but human genetics in intact organisms is generally restricted to forward genetics except in some special cases of clinical trials involving somatic cell genetic manipulation (*i.e.* gene therapy).

MONOGENIC VERSUS COMPLEX PHENOTYPES

To some extent, the distinction between monogenic and complex phenotypes is operational. All genetic variants in nuclear chromosomes are *de facto* transmitted in a Mendelian fashion. Thus even weakly penetrant variants, such as most SNPs found by association through large scale genome-wide studies in case control cohorts, are nonetheless “Mendelian”. Monogenic or Mendelian conditions are defined as being caused by variants in single genes, with the variants having a high penetrance (*i.e.* added risk). As a simple example, the clinical phenotype of cystic fibrosis has been made for many thousands of patients across all human populations. In all cases that have been studied, mutations are found in a single protein-coding gene, CFTR. Thus there is a clear one gene-one phenotype correlation for this condition. In contrast, oligogenic or polygenic or complex phenotypes are usually thought of as those for which combined input of many different genetic variants is required, potentially with significant environmental factors required as well (Fig. 1).

Even for nominally monogenic disorders the case of CF is far from typical. There are many medical examples in which a particular clinical diagnosis has been linked to high penetrance genetic variation in many different genes (e.g. ataxias [5], sensorineural deafness [6-8], retinitis pigmentosa [9-11]). In Charcot-Marie-Tooth disease, causal genes have been found with a broad spectrum of cellular processes including specialized neuronal structures, protein turnover,

*Address correspondence to this author at the Centre de Recherche du CHU Ste-Justine, 3175, Côte Ste-Catherine, Montréal QC H3T 1C5, Canada; Tel: 514-345-4931 p4265; Fax: 514-345-4801; E-mail: mark.e.samuels@umontreal.ca

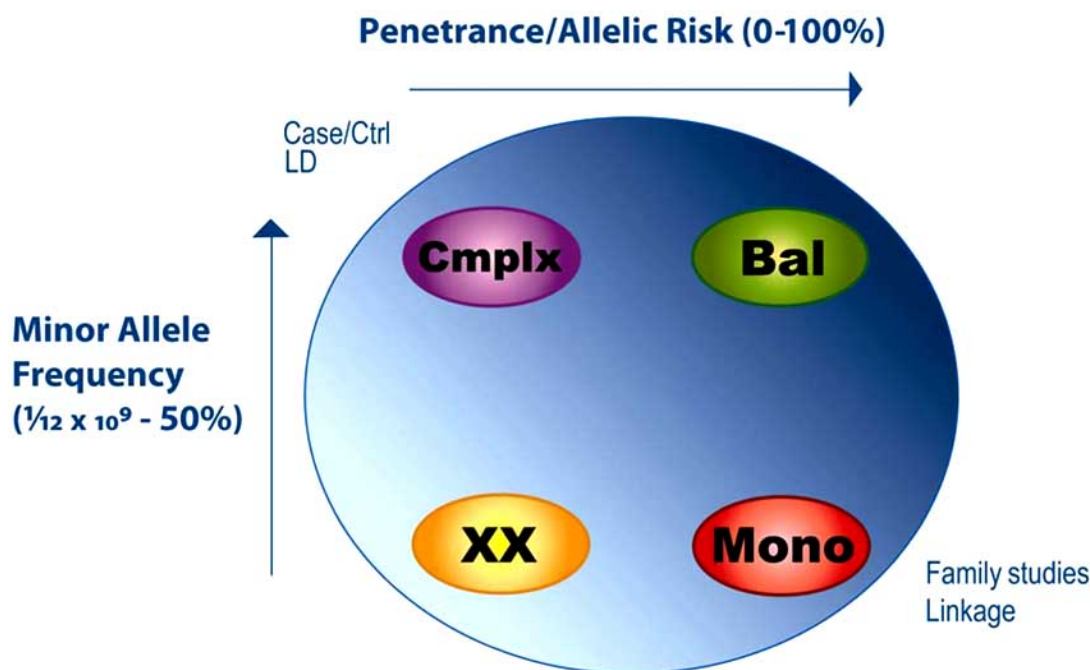


Fig. (1). Monogenic versus complex phenotypes.

Genetic variants may be anywhere between vanishingly rare (with a minor allele frequency at a minimum of one in the entire human population of approximately 6 billion, thus an allele frequency of $1/12 \times 10^9$ chromosomes) up to 50% (after which the minor allele becomes the major allele). The physiological effect of a genetic variant may be individually very strong (high penetrance) or weak (low penetrance). Rare high penetrance alleles have historically been identified in families, and studied by genome-wide scanning with dense polymorphic anonymous markers, currently SNPs, followed by statistical linkage analysis and gene resequencing to identify causal variants in the family. Common low penetrance alleles are typically identified in large case/control cohorts, and studied by genome-wide SNP genotyping followed by simple statistical tests of differential allele frequency in the two classes, or by more sophisticated linkage disequilibrium (haplotype) mapping. Common high penetrance variants are unusual, since these would normally be either quickly fixed or eliminated from breeding populations. One circumstance under which such variants can be maintained is balanced selection, where there are opposing selections on heterozygotes versus homozygotes (malaria and sickle cell anemia being the best documented example). Rare variants of small functional effect are easily discovered through random resequencing but difficult to study mechanistically.

vesicle fusion, microtubule transport, transcription factors, and even several tRNA-synthetases [12-19]. Clearly, some disease diagnoses represent a common end-point for different types of defects in gene function.

Somewhat less common in human genetics are examples of different allele-specific physiological outcomes. For obvious quantitative traits, such as plasma cholesterol, there are alleles in genes such as PCSK9 and APOB which may alternatively cause hyper- or hypocholesterolemia. These appear to correlate with gain or loss of function, such that loss of function alleles cause either high end or low end mean values in the measured trait, but not both, whereas gain of function alleles cause the opposite mean value. In a developmental example, Huntington disease is caused by triplet repeat expansions of a particular region of the encoding gene leading to some form of aberrant protein which causes neuronal toxicity through an incompletely understood mechanism. Formally such an allele is a clear gain of function (more specifically, a neomorph or new function versus a hypermorph or higher level of normal function). The alternative type of allele, a loss of function, is not known for the HD gene in humans, however the orthologous HD gene directed knockout is lethal in homozygous mice [20], and it is likely that a human loss of function allele would have a similar extreme phenotype clinically unrelated to Huntington disease.

There are also a variety of relatively well-documented examples of digenic inheritance in human disease, such that a clinical phenotype depends on simultaneous genetic variation in two different protein-coding genes. Retinitis pigmentosa resulting from simultaneous heterozygous mutations in the genes peripherin/RDS and ROM1 is perhaps the first documented example [21-28]. There are also likely digenic models for hemochromatosis [29, 30], Bardet-Biedl syndrome [31-36], sensorineural deafness [37-46] and other phenotypes.

For many monogenic genetic disorders there is significant inter-patient variability. This may be due to allele specific effects of the primary causal gene, or to secondary genetic modifiers [47, 48]. In some cases, it can be attributed to well-understood environmental factors. For example, mutations of the gene encoding phenylalanine hydroxylase cause childhood onset phenylketonuria, but there is great variation in clinical severity among patients [49], and disease progression depends on dietary intake levels of phenylalanine, such that dietary restrictions have beneficial effects in clinical practice. Similarly, hereditary hemochromatosis is generally more severe clinically in males than females carrying the same risk alleles; this difference is generally interpreted as females being partially rescued from iron overload through regular blood loss during menstruation [50]. It is widely as-

sumed that combinations of genetic and environmental factors play a major role in clinical severity for many monogenic disorders [51-53]. This has been difficult to address experimentally for most such disorders, as they are individually rare and affected patients are dispersed across the globe. Significant attempts have been made to identify genetic cofactors or modifiers for cystic fibrosis [54-57] and hemochromatosis [58, 59]. The distinction between digenic (or oligogenic) inheritance and genetic modifiers is somewhat semantic, depending on the ease of clinical ascertainment and severity of the physiological effect of the primary gene. In any case, the ability to combine spontaneous, mutagen-induced, and engineered genetic variants in model organisms such as mouse and fruit fly, make it clear that physiological states can be readily decomposed into the effects of multiple genes acting in concert in an individual. The commonly used term “synthetic phenotype” in *Drosophila* genetics refers to the same phenomenon as “digenic inheritance”, or “major genetic modifier”, as used more often in human genetics.

Monogenic disorders are operationally identified primarily by their observed transmission in families. In the case of weakly penetrant genetic variants, or those with major genetic modifiers, it is difficult or impossible to trace the inheritance of the physiological state (whether a disease, a quantitative trait, or a benign trait such as height) in families, in a fashion consistent with a single chromosomal locus. Furthermore, for every anecdotal case of traits “running in families” such as hair or eye color or height, there are examples where this does not occur. Clearly, some physiological states require multiple genetic inputs of similar strength penetrance, and some allelic variants on their own have little physiological impact. Assuming that all physiological traits can be defined by some quantitative variable or variables, although these may not be known, the effect of an individual variant allele can be visualized by its effect on the mean trait measurement. The further from the population mean a particular allelic variant is, the easier it is to detect and follow in individual pedigrees. The distinction between dominant and recessive acting alleles, while a practical complication in the detection of familial transmission, is theoretically not relevant to this point.

Even so, examples are known which blur these distinctions. Age-related macular degeneration (ARMD) has been strongly associated with genetic variation in the gene encoding complement factor H (CFH) [60-63]. The gene was identified independently by family-based linkage and in genome wide association methods, although the GWAS approach was clearly stronger. However, there is a very strong age-of-onset effect, which combined with the natural human lifespan makes it difficult to observe familial transmission of the phenotype. Factoring in the age-of-onset, the penetrance of CFH variants is extremely high, and this can be arguably considered a monogenic trait in the affected patients. If humans lived several hundred years, large dominant pedigrees for ARMD caused by CFH variants might well be ascertained.

Not all traits are diseases, obviously, although diseases receive the most attention from geneticists and are often the most readily ascertained through routine clinical surveillance. However, non-medical traits such as hair or eye color,

and ability to smell or taste particular chemicals, have been studied genetically [64-70].

HOW MANY GENES ARE THERE?

Determining the number of actual genes encoded by human (or other) genomes has proved surprisingly vexing. From early estimates in excess of 80,000 protein coding genes in the human genome based on quantitative measurements of CpG islands [71], estimates have dropped repeatedly, to 28-34,000 based on exon comparison between the pufferfish *Tetraodon nigroviridis* and human genomes [72], to 21,037 based on large scale cDNA library sequencing [73, 74], to 20,488 based on comparison of proposed human genes to sequences of other primate genomes [75]. Currently RefSeq, a widely used NCBI library of gene annotations, includes 21,515 unique entries. These include substantial numbers of non-protein coding genes such as 363 small nucleolar RNA (SNO) genes, 28 SNAR genes and 637 micro-RNA (MIR) genes, so RefSeq probably includes approximately 20,000 protein coding genes. There are several ongoing manual gene curation efforts, specifically Havana, VEGA [76, 77] and CCDS [78], as well as the Mammalian Gene Collection (MGC). The CCDS has currently curated 18,173 different protein coding genes, although some well documented genes are visibly missing from the database which is thus still in progress. The MGC currently includes 17,592 genes for which full length cDNA clones are available.

Although a consensus appears near that the total count of protein coding genes in the human genome is between 20-22,000, the advent of next-generation technologies has allowed further exploration of alternative splicing through deep resequencing of cDNA libraries. Several recent studies have documented novel alternatively spliced forms of known genes, as well as completely novel exons within previously studied genes [79, 80]. Thus the number of potential alternative protein isoforms is many times greater than the actual gene count.

The human mitochondrial genome has been comparatively straightforward to annotate, as it resembles bacterial genomes in gene structure. It encodes 13 proteins, all involved in respiratory electron transport and oxidative phosphorylation, as well as 22 mitochondrion specific tRNAs, and the 12S and 16S ribosomal RNAs (rRNAs) required for the mitochondrial ribosome [81]. However, the human mitochondrion itself contains 900-1000 different proteins, based on direct proteomic analyses [82, 83]. Thus the large majority of mitochondrial proteins are encoded by nuclear chromosomal genes.

STUDIES IN MODEL ORGANISMS

Saturation mutagenesis of model organisms has been widely used to identify mutable genes either for viability or for various interesting biological phenotypes [84]. In diploid organisms, dominant mutations are readily detected in heterozygotes after mutagenesis and breeding to rule out direct effects of the mutagen, although dominant lethality is obviously a difficult phenotype to recover. Recessive mutations can be detected in fungi by sporulation and reduction to haploidy, or in metazoans by establishing mutagenized het-

erozygous lines and inbreeding to create homozygotes, or particularly in *Drosophila melanogaster* by complementation testing of mutagenized heterozygotes with chromosomal deletions of cytologically or molecularly defined extent. A longstanding question has been whether all biochemically defined genes (transcribed, spliced, protein coding/noncoding, etc) have detectable mutational phenotypes. Even now, only a few historically saturated regions can be well-defined in terms of chromosomal extent, number of genetically mutable loci, and annotated protein-coding transcripts. With the advent of the complete genomic sequence plus extensive gene annotation, such studies can be revisited. This brief overview will focus on animals, leaving plant genetics aside.

The budding yeast *S. cerevisiae* is relatively straightforward to saturate since most genes lack introns. As a result at least for protein-coding genes, predictions can be made based on the occurrence of long contiguous open reading frames. Although the lower limit cutoff is somewhat arbitrary, based on annotation of the complete genome sequence the yeast genome project in 2001 identified 6138 potential protein coding genes. Essentially all of these have been individually deleted, and the resulting strains analyzed under a few different sets of growth conditions [85-87]. According to the yeast deletion project, 18.2% of genes are essential for growth in rich medium. A large-scale effort is also under way to map genetic interactions by combining individually deleted genes in pairwise combination [88]. Ultimately over 36 million different pairwise tests must be performed; to date 5.4 million tests have been reported but the results are intriguing. Approximately 170,000 interactions were detected, meaning that a strain carrying both deletions differed significantly in fitness from either single deletion strain. These interactions could be further studied by overlaying orthogonal annotations such as functional cellular modules, expression profiles, and protein-protein interaction profiles.

In *D. melanogaster*, historical saturation mutageneses have been performed for multiple genomic regions, while the essentially complete genome sequence of the organism together with a preliminary gene annotation was published in 2000 [89]. Overall the fly genome is believed to contain 13-14,000 protein coding genes. Saturation of the *zeste-white* region of the X chromosome in *D. melanogaster* has been performed repeatedly, using various mutagens [90, 91]. Defining the region as that uncovered by a particular deletion (Df(1)w¹¹, encompassing cytological bands 3A2-3C2), mutations were initially identified in 17 loci, 15 generating lethality and 2 generating visible phenotypes. Currently Fly-Base annotates 31 genetic loci uncovered by this deletion, including genes with additional lethal, female sterile or visible phenotypes beyond the original 17. One caveat is that in rare cases, different mutations within the same gene may show genetic complementation, overestimating the number of actual genes. Through transcriptional and comparative genomic analyses, Flybase annotates a total of 43 genes in this chromosomal region, including 19 identified only as candidate genes. In some cases genetically defined loci have not yet been assigned to structurally defined genes. Thus up to three fourths (31/43) of annotated genes may be mutable to an observable phenotype in this region.

In another chromosomal region also intensively studied in *D. melanogaster*, a strikingly different result was obtained. In a 2.9 million base region around the alcohol dehydrogenase (*Adh*) gene, Ashburner *et al.* documented 218 potential protein coding genes, but only 73 genetically definable loci in the region, of which only 49 have been assigned to specific structural genes so far [92]. Thus in the *Adh* region only a third of structurally defined genes may have individual mutational phenotypes.

Thus a substantial fraction, and in some cases a majority of protein coding genes in an interval do in fact generate a mutable phenotype, typically recessive lethality. Attempts at saturation mutagenesis using P-elements have been widely performed in *D. melanogaster* [93], however it is well documented that this transposon shows significant site specificity thus is not a good mutagen for true saturation. Efforts are currently under way to generate genome-wide saturation with transposable elements demonstrating less site specificity than the original P-element [94-96]. Although the larger number of genes in the fly makes whole genome pairwise genetic mapping a daunting prospect, nonetheless *de novo* mutagenesis in genetically compromised fly strains is a routinely used tool to identify genetic interactions and members of shared physiological or regulatory pathways [97, 98]. Such modifier genes are often referred to as suppressors or enhancers of a phenotype, which may be either visible or lethal in nature.

Although the small nematode worm *Caenorhabditis elegans* has a shorter experimental history than flies, ease of manipulation quickly made it a favorite among developmental geneticists. Moreover its small number of cells has permitted the unique opportunity to map all of embryonic and postembryonic development at the cellular level [99, 100]. The *C. elegans* genome has also been sequenced, and annotated to contain slightly fewer than 20,000 protein coding genes [101]. It remains puzzling why the worm, a much 'simpler' organism than the fly, has more genes, and in fact has about the same number as mammals. *C. elegans* has been the subject of a large number of genetic analyses, and pilot transposon-mediated mutagenesis has been performed [102]. However, the use of antisense RNA (RNAi), originally discovered in *C. elegans* [103], has permitted large scale gene knockdown experiments using libraries of gene-specific antisense targeting reagents, formally similar to germ-line loss-of-function mutagenesis [104, 105]. A study of RNAi targeting 16,757 genes from a predicted set of 19,427, detected mutant phenotypes for 1,722 probes, or 10% of targeted genes. Of these, slightly more than half caused embryonic lethality (RNAi was fed to hermaphrodite worms, allowing analysis of the effects in their progeny), with an estimated success rate of 70% compared to a set of genes with known germ-line embryonic lethal mutations. This suggests that 8% of genes can mutate to an embryonic lethal state in this organism. The overall low rate of observed phenotypes is difficult to interpret; it may simply be that most genes are functionally redundant in the worm. Alternatively, RNAi reagents may not be effective in completely reducing gene function for some genes.

Large scale random mutageneses have been performed in *Mus musculus* using both chemical and genetic mutagens

[106-125]. Chemical mutageneses typically employ ethylnitrosourea (ENU), but the prohibitive logistics of generating and maintaining the huge numbers of heterozygous F1 lines, required to create homozygous F2 animals to screen systematically for recessive mutations, have precluded a true genome-wide saturation. Genetic approaches primarily involve transposon-mediated insertional mutagenesis in mouse embryonic stem (ES) cells. There are now multiple libraries of embryonic stem cells carrying random transposon insertions, for which insertion sites have been mapped by direct sequencing. This potentially permits researchers to study the insertional phenotype for a gene of interest, if it happens to exist in a library. The vast majority of these transposon insertions have not been developed from ES cells into whole animals. Moreover, many of these insertions are not explicit gene knockouts as insertions are not targeted to coding exons.

As an alternative to random mutagenesis, many targeted gene knockout mouse strains have been constructed. These are inevitably biased in favor of "interesting" genes, with the occurrence of known human genetic phenotypes being one component of interest. However there are other reasons for genes to be defined as interesting, and it is unclear whether existing knockouts are biased toward genes certain to yield a detectable mutant phenotype in mice. In a review of mouse knockouts as of 1995, it was reported that of 263 different genes targeted, 25% yielded an embryonic lethal phenotype, with another 10% leading to death within a few weeks postnatally [126-128]. Overall, 95% of knockouts were reported to show an observable phenotype, a very high proportion in comparison to the RNAi results with *C. elegans*. Subsequently it was reported that about 1/3 of direct knockouts yield embryonic lethality [129]. In an analysis of mouse knockouts for 34 genes which are known targets of commercial therapeutics for important human diseases, all but one knockout strain showed some detectable phenotype in the mouse, most of which were directly relevant to the equivalent human disease [130]. In one case, that of insulin, the gene is known to be duplicated in the mouse genome, such that the double knockout is lethal whereas either gene alone does not yield a phenotype [131, 132]. In contrast, a knockout disrupting the insulin receptor gene, which exists in a single haploid copy in the mouse genome, is severely ill and dies shortly after birth, validating that the insulin pathway is intrinsically mutable in the mouse [133, 134].

The zebrafish *Danio rerio* has been the subject of many large scale mutageneses, since it was originally developed as a research organism by Streisinger [135-140]. In the absence of a gene-specific targeting system, most mutageneses have involved ENU followed by screens for interesting biological phenotypes. Identifying the actual mutated genes requires large scale recombinational mapping using anonymous polymorphic DNA markers, either microsatellites or SNPs, followed by positional cloning. Since the advent of high throughput sequencing, an intermediate approach called TILLING has been developed [141-143]. Instead of mapping and positional cloning, mutations in a given gene are found directly by sequencing pools of very large numbers of mutagenized animals (or plants, where the method was originally developed [142]) until a variant in the desired locus is detected. Using these approaches mutations have been

found affecting many different physiological systems. Another effective approach in zebrafish is the use of antisense reagents. Modified oligonucleotides containing morpholino groups to reduce enzymatic breakdown can be injected in early fish embryos [144]. Oligos targeting genes are capable of producing very specific knockdowns, similar to the situation in worms. One limitation is that the morpholinos are eventually degraded, so only events of embryogenesis can be interrogated with this technology. Morpholinos have been used in other important experimental organisms such as the clawed frog *Xenopus laevis*, but genomic saturation screens have not been reported.

In sum, mutational studies in model organisms have demonstrated that a sizeable proportion of genes can be mutated to generate observable phenotypes. The exact number is difficult to determine from the available data sets, as there seem to be broad discrepancies between yeast and worm (small proportion of directed knockouts/knockdowns generating phenotypes) and flies (substantial portion of genes mutable) and mice (large proportion of directed knockouts generating phenotypes), nor is there a consensus as to the proportion of all genes that are essential for viability. It remains to be seen how these results will translate to human physiology, although it seems safe to assume that there must be substantial numbers of genes mutating to embryonic lethality, to viability with clinically observable phenotypes, or to complete viability.

HOW MANY MONOGENIC DISORDERS ARE KNOWN?

Keeping these points in mind, how many characterized monogenic phenotypes are there in humans? The entire human genetics literature is vast. There are special locus-specific databases and/or web sites for many genetic disorders, such as cystic fibrosis (<http://www.genet.sickkids.on.ca/cftr/app>) and familial hypercholesterolemia (http://www.umd.necker.fr/LDLR/Home_Page.html). There are at least three organizations attempting to curate the entire literature [145]. The best-known, Online Inheritance in Man (OMIM), functions as a mixed medical and genetic database, with entries for clinical disorders intermingled with entries for specific genes [146]. It has limited direct search functions. Querying with the universal term 0001 together with a limit for allelic variants, OMIM returns 2420 independent gene entries at the time of writing, while the statistics page reports 2763 phenotypes with a known molecular basis; the excess presumably resulting from some genes mutating to different phenotypes with independent OMIM entries [147]. Almost all of these represent monogenic disorders (as practically defined, high penetrance variants with reasonable evidence of genic pathogenicity based on the mutation itself). However, a fraction of these, possibly as many as 10%, may represent low penetrance risk alleles not clearly definable as part of monogenic disorders. OMIM does not curate all reported genetic variants in all genes, although entries are regularly updated when scientifically interesting results are published.

The Human Gene Mutation Database (HGMD) curated at Cardiff University, attempts to consolidate all reported genetic variants in the literature [148]. It has an incomplete

publicly available version, and an up-to-date commercial version. The commercial HGMD currently includes 3611 independent gene entries and 96,631 mutation entries. It is difficult to assess how many of these correspond to monogenic conditions, versus low penetrance risk factors for complex phenotypes. HGMD suggests that 2491 genes contain at least one presumptive high penetrance allele (frameshift or premature stop codon, splice site, missense or indel) (*P. Stenson, pers comm*) [149]. The difference of fewer than 100 genes between OMIM and HGMD may relate to subtleties of literature curation of suggested versus confirmed genotype/phenotype causation.

The Human Genome Variation Society (HGVS) is developing a program of comprehensive curation which may eventually provide similar information as OMIM and HGMD combined [150, 151]. Currently the database HGVSbaseG2P appears focused primarily on the results of GWAS studies for complex disorders. It includes links to HGMD, but presumably only the publicly available content not the full proprietary database.

The total number of molecularly characterized, true monogenic disorders remains slightly uncertain, but is slightly less than 2500, representing 10-12% of protein-coding genes. For the remaining ~90% of genes, no high penetrance genetic variation has been causally linked to a physiological phenotype in humans.

MUTATION RATES

Rates of spontaneous mutation in the human genome have been estimated in various ways. Combining early results of sequencing a small number of real or pseudogenes, Drake *et al.* suggested a rate of approximately 10^{-8} events per base pair per generation, equivalent to 64 new mutations in each zygote [152]. By comparing partial human and chimpanzee sequences, Nachman and Crowell estimated a neutral rate of $1.3-3.4 \times 10^{-8}$ events per bp per generation, equivalent to 91-238 new mutations per zygote [153]. More recently a direct test was performed by sequencing large segments of flow-sorted Y chromosomes from two males of a precisely known ancestral genealogy. Four new mutations were detected between the two samples, which correcting for chromosome size yielded a rate of 3×10^{-8} events per bp per generation, although the small number of individual events implies a reasonable sampling bias [154]. Finally, whole genomes were resequenced for a pedigree of two parents and two offspring, allowing the most direct possible measure. A total of 28 confirmed new mutations were detected, which after various corrections yielded a rate of 1.1×10^{-8} events per bp per generation, equivalent to 70 new events per zygote, with a substantially lower chance of sample bias [155].

These all represent essentially point mutation rates. It is now clear that larger structural rearrangements, generally termed copy number variants (CNVs), also arise at reasonably high frequencies in human populations, although it is difficult to assess true *de novo* mutation rates given the somewhat arbitrary definition of a CNV [156-158]. Many CNVs include genes and have potential functional consequences, although the proportion is not yet well established [159].

Not all new mutations need have functional consequences. It is generally considered that most of them are in fact functionally neutral. There is no easy way to assess this since most new mutations arise in non-coding regions of the genome, in either introns or intergenic regions, where our understanding of functional elements is still very incomplete. The fraction of haploid sequence coding for protein is typically considered to be 1-2%, thus if each zygote receives 70 new mutations, at most one of these is expected to be in a coding exon. Coding variants may be silent (synonymous) or change an amino acid (nonsynonymous). Some studies have argued that a large proportion of nonsynonymous changes have functional significance, although these may only be evident over evolution time frames and not be directly relevant to interpretation in individuals [160-163]. It should be noted that synonymous changes may also have measurable effects on gene function, by influencing splicing *via* exon splice enhancer or repressor elements, or translation efficiency.

The rate of new deleterious mutations (usually termed U when summed over the entire genome), while certain to be substantially less than the rate of overall nucleotide change, cannot readily be inferred from the global rate. A comparison of human and chimpanzee sequences, using some assumptions about evolutionary selection, led to an estimate for U of 1.6 per zygote [164]. A separate analysis, again using human/chimp comparison, obtained a value of $U = 3$ [153]. In a key study, Kondrashov combined and re-analyzed data for 20 different genes for which typical loss-of-function mutations lead to either autosomal dominant or X-linked high penetrance clinically ascertainable phenotypes [165]. Medical ascertainment rates per live birth include both new cases and familial cases, but by restricting to new cases alone, the median and mean total mutation rates for this set of genes can be calculated as 7 and 14×10^{-6} per live birth respectively. Assuming 20,000 protein coding genes, these are equivalent to $U = 0.14-0.28$ per genome per birth, substantially lower than the rates estimated by evolutionary comparison. A likely explanation for the discrepancy is again that evolutionary comparisons probably involve functional effects too modest to yield a clinically ascertainable phenotype in individuals. Although a rate of $U \sim 0.2$ seems low, given a worldwide 2009 birth rate of approximately 139,000,000 [166], this represents almost 28 million new pathogenic mutations, or 1400 new mutations per gene worldwide. Even Canada, with a modest total population of about 34 million, has an annual birth rate of 378,000, and could have as many as 75,000 new pathogenic mutations per year nationwide. A significant caveat is that these rates may reflect a bias to genes with unusually high mutation rates. A number of X-linked recessive mutating genes have lower mutation rates, such that a value of $U = 0.02$ may be a much more conservative genome-average estimate [167]. Even so, the entire human gene repertoire is clearly saturated annually worldwide with new pathogenic alleles. There is no shortage of high penetrance mutations in every gene segregating in human populations. The only reason why high penetrance monogenic disorders are relatively rare is that most mutations act recessively.

DOMINANCE VERSUS RECESSIVITY, GAIN VERSUS LOSS OF FUNCTION

The proportion of monogenic disorders transmitted as dominant versus recessive traits in families is hard to determine from the curated databases. Similarly, it is difficult to assess the proportion of gain-of-function (gof) versus loss-of-function (lof) alleles, although this is not purely an issue of curation since for many alleles the functional effect remains untested biochemically. Complete gene deletions represent the gold standard for a loss-of-function allele, however few clean deletions of individual genes have been associated with monogenic conditions (whereas many contiguous deletion syndromes encompassing multiple genes are well documented). Geneticists generally assume that major protein truncations arising through frameshift or premature stop codon mutations are pathogenic, but it would be naive to presume that these all represent complete loss-of-function since amino terminal protein fragments could easily encode partial or aberrant activity. Wilkie provides a useful compendium of molecular mechanisms that can lead to genetic dominance (haploinsufficiency, excessive gene expression, novel toxic protein function, etc) with examples from model organisms and the human genetics literature [168]. For individual genes, the best way to distinguish between gain and loss of function phenotypes is to observe the array of alleles associated with a particular phenotype. For example, mutations of BRCA1 which increase the risk of early-onset breast cancer include a wide variety of missense, truncating and deletion mutations [169]. It is reasonable to presume that most of these represent loss of function alleles. Similarly, many independent alleles causing Huntington disease all involve expansion of the triplet repeat, while causal mutations elsewhere in the gene are not known for this phenotype, supportive of the interpretation that these are gain of function alleles. In contrast, Friedreich ataxia is typically caused by a triplet repeat expansion in the frataxin gene, however a small fraction of patients result from point alleles, both missense and truncating [170]. Conceivably, the ataxia in general could represent a loss of function phenotype, and the frequency of triplet repeat expansion alleles could result simply from an easily accessed mutational mechanism versus a specific novel function of the mutated protein. Alternatively, some specific point alleles could cause an aberrant novel protein function akin to that of the triplet repeat expansion. One glaring example is the gene encoding superoxide dismutase (SOD), mutations in which are causal for a fraction of cases of amyotrophic lateral sclerosis (ALS, or Lou Gehrig disease). There are many different causal mutations in SOD, almost all of which are missense variants [171]. However, a mouse gene knockout does not generate the ALS phenotype, whereas overexpressing equivalent missense variants recapitulates some aspects of the human disease [172]. The consensus in the field is that many or most of these missense variants, which arise in almost every residue of this small tightly folded protein, are somehow gain-of-function, although the exact biochemical mechanisms are still being elucidated. Thus it cannot be assumed that phenotypes arising from many different mutations in a gene invariably represent loss of function, although they probably do in most cases.

The clinical presentation of different mutant alleles can sometimes clarify the interpretation. Hypercholesterolemia can be caused by one of a small number of missense mutations in the gene encoding protease PCSK9. However, a much larger number of mutations including obligate protein truncations in the same gene are associated with hypocholesterolemia. The hyper alleles are typically heterozygous, whereas the hypo alleles are typically homozygous in the patients [173-177]. These observations strongly suggest that the hyper alleles are gain of function, and the hypo alleles are loss of function. The rarity of the gof alleles is consistent that the map of genotype to the hyper phenotype is more highly constrained in this gene than the map from genotype to the hypo phenotype. Similarly, many presumptive lof alleles of the gene encoding apolipoprotein B exist, and lead to hypobetalipoproteinemia (with low LDL-cholesterol) due to lack of this major structural protein component of VLDL and LDL particles. In contrast, only a small number of specific amino acid missense variants are known in the ApoB gene causing hyperbetalipoproteinemia (with high LDL-cholesterol), with the mechanism believed to be failure of recognition of mutant ApoB protein by the LDL receptor, and thus inability of the body to remove LDL particles from the plasma [178-180].

Although most mutations are probably recessive, the ratio of dominant to recessive monogenic disorders is probably significantly skewed in human genetics in favor of dominants. High penetrance mutations segregating as heterozygotes have the potential to generate large multigenerational pedigrees transmitting the physiological trait. If the trait is a medical disorder, such as Huntington disease or Alagille Syndrome, such large families are relatively obvious to astute clinicians, particularly since the advent of effective molecular mapping in the 1980s. In contrast, recessive disorders often involve only a single affected in a nuclear family, since only $\frac{1}{4}$ of offspring on average segregate the trait (for an autosomal disorder). Ascertainment of recessive traits by clinicians typically involves unusually large sibships with recurrence of the phenotype in multiple children. In small sibships with only one or a few affected children, environmental or stochastic etiologies are more likely, and the hypothesis of recessive segregation is less easy to demonstrate except by a genome-wide mapping experiment.

There is a strong practical reason to believe that most loss of function alleles function as recessive genetic traits. Measured rates of mutation are consistent with very high rates of occurrence of novel pathogenic alleles in humans, perhaps as many as one per meiosis. If many new mutations are pathogenic, then severe effects on fitness of entire populations would ensue if most mutations were dominant with pathogenic effects in heterozygotes. Hence, most mutations are reasonably anticipated to have recessive effects on physiological traits (even if they have formally co-dominant effects on their immediate biochemical function through simple gene dosage effects). Indeed, protection from high mutation rates is likely one of the evolutionary forces driving diploidy and hence sex, although most evolutionary biologists tend to focus on sex as a mechanism for generating rather than reducing diversity.

The recessive nature of most mutations was appreciated early in the 20th century, and theoretical geneticists devel-

oped several alternative models to explain this [181, 182]. Wright suggested that two doses of genes provided near-saturating levels of activity for the ultimate physiological trait, such that loss of one dose had a non-linear and only weak effect on the trait (Fig. 2) [183]. Fisher suggested instead that mutant alleles evolved in concert with genetic modifiers elsewhere in the genome. Wright's argument received strong theoretical support from the metabolic control rate theory of Kacser and Burns [184, 185]. These theoreticians sought to understand why rates of production of pathway end-products in engineered microorganisms were difficult to influence by changing amounts of individual enzymes in the reaction pathways, even when those enzymes were known to be 'rate-limiting' based on biochemical kinetic studies. They showed mathematically that control of flux rates through complex metabolic pathways *in vivo* was more likely to be distributed than focused on one gene product (Fig. 3) [184]. The result would be that modest (even two-fold) changes in dosage of a gene product would be unlikely to affect overall flux through a metabolic pathway, although this point has been argued [186]. Assuming that some kind of biochemically quantitative pathways underlie all physiological traits, even those of embryonic development or neuropsychiatric function, metabolic control theory suggests at least in principle that gene dose response could readily be nonlinear, and that loss of even a full copy of gene function might have only partial effects in non-enzymatic systems [187].

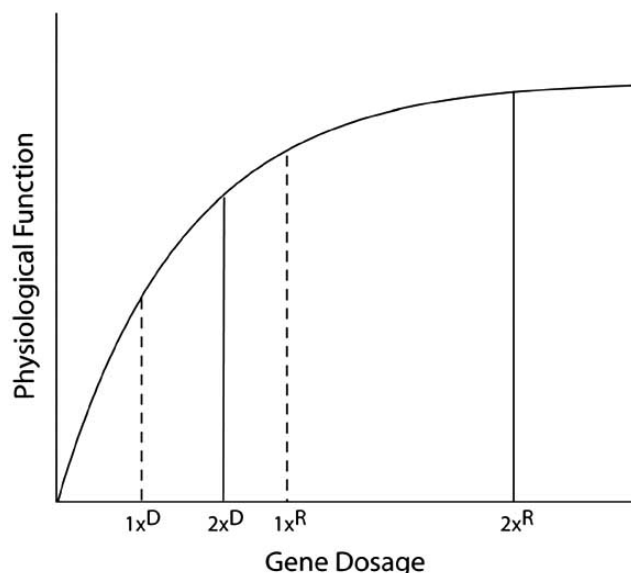


Fig. (2). Saturation kinetics and genetic dominance.

As suggested by Wright, if two doses of a wild type gene in a diploid organism provide saturating levels of physiological function ($2x^R$), then loss of one copy through mutation would still provide most functionality ($1x^R$). Such a situation would lead to recessive genetic behavior, whereby a heterozygote *lof* mutation would have little biological impact. Wright hypothesized that this would be the case for the majority of genes. In some cases, two wild type gene doses might provide function in the subsaturating range ($2x^D$), such that loss of one dose in heterozygous *lof* mutation carriers would have substantial biological effect ($1x^D$), leading to observed dominance. Dominance caused by unusual gain-of-function alleles is not an element of this model.

Genetic studies in model organisms tend to support these arguments [168]. Although spontaneous or mutagen induced mouse mutants can span the full range of mutation types, engineered knockouts almost always involve deletions of large amounts of protein-coding potential, and are likely to function as *lof* alleles in most cases. Anecdotal evidence from mouse geneticists suggests that most such knockouts have no detectable phenotype in heterozygous state. A caveat to these observations of course is that mice cannot tell us what is wrong with them, thus some heterozygous knockout mice could have physiological phenotypes undetected by experimentalists. Similar results are seen in fly mutagenesis experiments, in which mutagenized flies are made heterozygous against a strain carrying a defined chromosomal deletion, with breakpoints defined either cytologically or now by DNA sequencing. In such experiments, many independent alleles can be obtained for most genes covered by the deletion, and their phenotypes compared against the deletion or when homozygosed individually or combined as compound heterozygotes against each other. The general result of such studies is that most alleles with a detectable phenotype in most genes act recessively, with no obvious heterozygous effect.

In a unique analysis to test the Wright hypothesis directly, the mutational spectrum was reviewed for the microorganism *Chlamydomonas*, which normally grows in a haploid state. When semi-synthetic diploids are generated, mutations which originally arose in haploids prove nonetheless to be almost exclusively recessive [188]. This result strongly argues against recessivity arising from coordinate evolution of oligogenic modifiers in the diploid state, and supports Wright's argument that recessivity of mutations is a consequence of the way in which metabolic pathways have evolved. Efforts to extend the Kacser and Burns argument to other, nonmetabolic developmental pathways have shown that this is in principle feasible [181]. Nonetheless, it remains conceivable that certain types of developmental programs might not be so buffered in higher eukaryotes. One example might be the human brain. Copy number studies have documented a significantly higher incidence of CNVs, mostly deletions, in autistic than in control patients, presumably mostly in the heterozygous state. If so, then loss of function mutations might cause frequent dominant phenotypes such as schizophrenia or autism which both appear to have a high rate of sporadic etiology consistent with high mutation rates and recent evolutionary events.

Given that mutations in most genes probably act recessively, the rate of monogenic disorders ascertained clinically is substantially lower than the rate of appearance of new alleles. It depends on a combination of new mutation rates and population demographics. In outbred large populations, individuals will rarely be born carrying two non-complementing mutations in *trans* in the same gene (either as homozygotes or compound heterozygotes). In contrast, in population isolates experiencing significant inbreeding due to either geographic or cultural factors, rates of recessive disorders may be substantially higher than otherwise expected. In addition, due to founder effects different genetic disorders occur at increased frequency in different populations. This is well-documented by medical geneticists, in diverse groups such as Finns, French Canadians, Ashkenazi Jews and Alberta

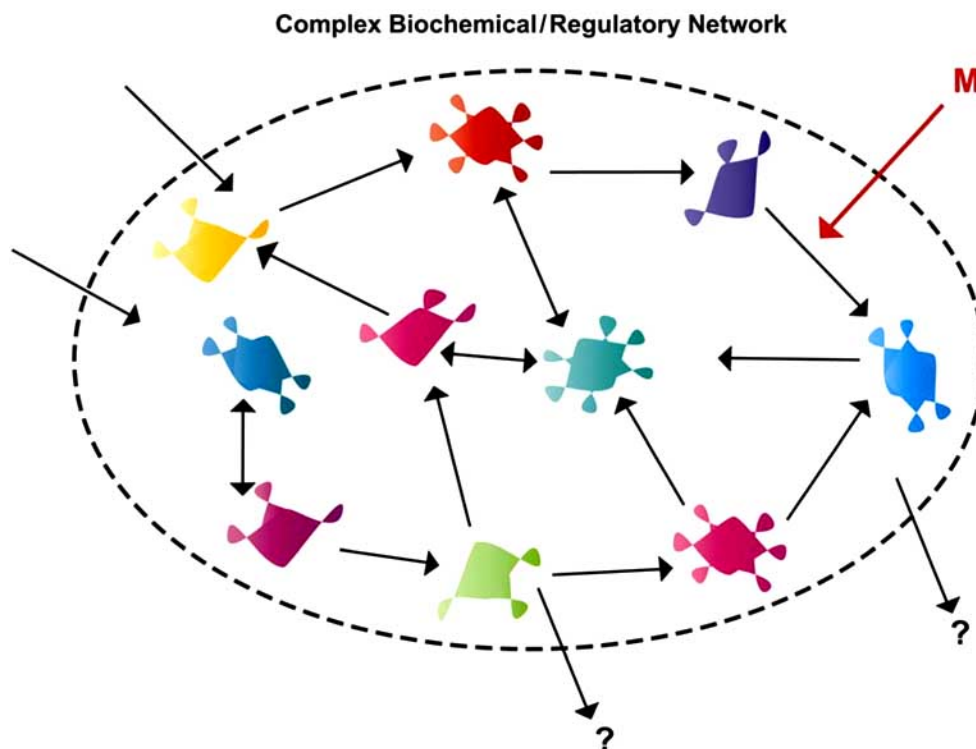


Fig. (3). Metabolic control theory and gene networks.

As suggested by Kacser and Burns [184], measures of enzymatic rates in purified *in vitro* systems do not necessarily reflect true *in vivo* situations. In reality, metabolic networks are complex, often with alternative routes from substrate inputs to catabolite outputs. In such cases, changing the rate of one specific step (indicated by arrow M), either *via* genetic mutation or through a targeted chemical antagonist or agonist (i.e. a drug) might lead to only modest, or unpredictable effects on total flux through the system and on the equilibrium levels of intermediate metabolites. This could explain both dominance as well as functional redundancy, whereby mutation of most genes even in homozygotes is not lethal, even though the genes are “required” in a mechanistic sense. Similar situations could be the case for developmental regulatory networks, where the flow of information rather than chemical metabolites is concerned.

Hutterites [167, 189-192]. In one town in Israel, geneticists identified 19 different recessive monogenic disorders arising in a population of less than 10,000 [193].

DEFINING THE REST OF THE PHENOME

If only a small fraction of protein-coding genes have a molecularly characterized monogenic phenotype in humans as of now, what does this mean for the complete phenome? If the complete phenome comprises all possible phenotypes of all genes in the genome, then the phenome is effectively infinite. However, if most loss-of-function alleles in a given gene cause similar phenotypes, then the lof phenome is much smaller, probably finite, and only slightly larger than the number of genes. Further, if gof alleles are much rarer than lof alleles, then most of the gof phenome will never be observed given the total human population size. In contrast, as described previously, the lof phenome probably exists across the world near completion now.

All these points together strongly suggest the following scenario: lof alleles of all genes in the genome arise regularly around the world; most of these represent recessive alleles; clinical ascertainment of homozygotes for these alleles will be individually rare except in founder or other population types with higher rates of consanguinity than expected for complete panmixis; compound heterozygotes will arise at

low rates essentially everywhere, with little or no increased rate in founder populations. The actual rate of appearance of recessive cases that might be clinically ascertained is very difficult to estimate, since it depends on current allele frequencies for deleterious alleles, which in turn depend partly on details of local population demographic history which for most populations are poorly known. However, even in presumptive outbred or non-founder populations there can be increased frequency of recessive deleterious alleles through unappreciated low levels of population mobility – the reduced number of observed versus potential haplotypes found in non-African populations provides an extreme example of this.

The inescapable conclusion is that recessive cases for lof alleles of most genes do occur regularly. Where then are these missing phenotypes? Three possibilities remain: either many genes mutate to a neutral lof phenotype, or else they mutate to an embryonic lethal phenotype, or else they mutate to phenotypes not yet clinically defined or molecularly characterized as genetic conditions. Half of all spontaneous miscarriages show cytogenetically detectable chromosomal abnormalities [194]. With the advent of high density hybridization arrays, smaller rearrangements especially deletions can readily be detected, well below the resolution of light microscopy [195, 196]. It seems likely that an additional fraction of miscarriages, even earlier in fetal development, entail

mutation of critical embryonic genes. For example, viable human mutation phenotypes are known for six of the 39 genes encoded by the four HOX gene clusters [197]. These are among the extreme rostrally and caudally expressed genes HOXA1 [198], A2 [199], A11 [200], A13 [201, 202], D10 [203, 204] and D13 [205]. Mutations in the majority of more centrally expressed HOX genes may mutate to recessive embryonic lethality early in development, as probably do other early expressed pattern-forming genes. OMIM currently includes several thousand phenotype entries with suggestion of a genetic etiology, many of which probably do represent recessive cases arising in families too small to map a genetic locus.

The advent of dense SNP marker panels has greatly facilitated the mapping of recessive genetic loci, especially in population isolates with not only increased recessivity but also homozygosity. Formal linkage analysis is relatively inefficient for recessive disorders unless there are an unusually large number of affected individuals in a given pedigree or related population (*i.e.* founder effects). The availability of new technologies is about to change this paradigm. In the last two years multiple individual human genomes have been sequenced to high coverage using next-generation technologies [206-213]. To reduce the cost of whole genome analysis, hybridization capture methods have been developed by which selected subsets of the genome, such as protein-coding exons, can be sequenced [214-217]. Individual genomes have millions of differences from the human consensus sequence as assembled by the Human Genome Project (current version hg19, also called GRCh37). Of the 3-4 million such differences in individual genomes, over 90% can be found in dbSNP v130, which includes a large number of variants discovered through the 1000 Genomes Project [218-220]. The large number of remaining variants is still too many to define pathogenic causality in general cases, although several recent reports demonstrate the feasibility of array-capture exome resequencing to define likely pathogenic variants in affected individuals or families with rare clinical phenotypes under ideal conditions [221-229]. Operationally, assigning genotype/phenotype causality unambiguously requires either multiple patients mutated in the same gene with a rare monogenic phenotype, or alternatively strong validation of a candidate gene through animal model studies. Nonetheless, whole genome or exome sequencing is poised to revolutionize medical genetics, allowing the identification of causal genes for many more rare phenotypes, including some (such as prenatal lethality, or dominant disease with reduced viability or fertility) for which family studies are intrinsically impossible.

This review has focused primarily on protein coding genes. There are some examples of genes producing non-protein coding RNAs for which high penetrance mutational phenotypes occur (the H19 gene in mice [230], and the XIST gene in humans) [231-233]. In general such assignments are challenging since we have little ability to define functional changes in non-coding RNAs except when they are extensively deleted. It remains to be seen whether naturally occurring variation in other such RNAs, such as microRNAs, can lead to monogenic phenotypes. One exception is in the mitochondrial genome. Well documented mutations in mitochondrial genes, including tRNA and rRNA coding genes, lead to

clinical disorders of the mitochondrion, which although complex physiologically are generally definable by experienced clinicians [234, 235].

PRACTICAL APPLICATIONS OF PHENOME SATURATION

Beyond the purely scientific value of determining high penetrance mutational phenotypes for as many genes as possible, there is also practical value in this program. The human genetics community has focused much of its effort for the past decade on defining risk variants for complex, often adult-onset degenerative diseases of great social and commercial importance. This work has mostly employed the analytical paradigm of high throughput SNP genotyping and case/control comparison using large patient cohorts (the GWAS approach). GWAS studies have succeeded in identifying many such risk factors, but in general the variants found individually add only small incremental risk of disease, moreover only a modest fraction of total genetic risk in populations is explainable summing over all such common alleles [236]. The emerging consensus is that finding the "missing heritability" will require high throughput genomic resequencing in larger and larger cohorts. If one of the goals is to identify novel therapeutic targets for pharmaceutical intervention, this will be a long-term and monumentally expensive proposition.

In contrast, there are many good examples of high penetrance monogenic traits of direct relevance to common disease. Elsewhere I have discussed in depth the use of "monogenic disorders" as a tool to define novel therapeutic targets [237]. These can be found either by traditional family-based linkage studies [238-240], or for quantitative traits such as body mass, plasma glucose or cholesterol, by sequencing population outliers looking for excess variation in specific genes [241-245]. Essentially, some component of the missing heritability for medically important traits lies in a large aggregate number of individually rare alleles. Such genes provide "low hanging fruit" for drug development, something desperately needed by the biotechnology and pharmaceutical industries, which are universally acknowledged to be suffering a dearth of good new products. For example, the human genome encodes roughly 370 G-protein coupled receptors, not including olfactory receptors. Many of these are targets of drugs already on the market, as GPCRs have several major advantages for new drug development. As with the human genome in general, only about 10% of GPCRs have documented human monogenic phenotypes. Many of these human phenotypes are directly relevant to important disease models. Thus the GPCR phenome represents a fertile area for genetic analysis, requiring only the resequencing of this set of genes in clinically defined cases of medical interest.

Setting aside the current hype surrounding personalized genomics, there are several ways in which phenome saturation will have immediate diagnostic benefit. In pediatric hospital settings, a substantial fraction of admissions involve genetic disorders [246, 247]. Some of these, such as congenital heart malformations and neural tube defects, are seemingly genetically complex. But others involve genetically simple disorders. For example, pediatric hospitals often have

dedicated cystic fibrosis units, as CF is one of the most commonly diagnosed pediatric genetic disorders with a known single gene basis. Given that there are already nearly 2500 molecular characterized genetic disorders, many of which are regularly ascertained in any large pediatric care setting, genetic analysis is critical in pediatrics even now. The definitive diagnosis of such conditions, either through molecular analyses or simply by phenotypic description, has high value to patients and families, and potentially to the health care system as well. Further definition of all remaining pediatric monogenic conditions will prove similarly useful, if only to eliminate non-genetic causes of disease in individual patients, ending otherwise time-consuming and expensive diagnostic odysseys.

CONCLUSION

I have attempted to document that most of the human phenome remains to be ascertained and molecularly characterized. This is primarily due to challenges in clinical ascertainment and gene mapping, not to any theoretical issues with intrinsic deleterious mutation rates or the occurrence of previously uncharacterized phenotypes. Mutation rates in the human population are sufficiently high that pathogenic variants must exist in all functional genes at the present time. Most of these are likely to cause recessive phenotypes, thus ascertainment depends on identifying homozygous or compound heterozygous mutation carriers from among all medically identified patients. Historically, causal gene identification has depended on the identification of multiplex families, followed by mapping with panels of anonymous DNA polymorphisms, position cloning and resequencing. With the advent of next-generation sequencing technologies, it is now feasible to sequence whole genomes or exomes of patients without the need for families or linkage mapping. However, the large amount of individual genome variation raises the challenge of identifying causal variants in these patients. This will require collaboration among medical geneticists dispersed around the world, possibly with development of large genotype/phenotype databases, together with the use of model organisms to validate gene function and/or mutation pathogenicity.

With these ideas in mind, this is my own estimate of the lof human phenome. Time will tell how accurate these predictions are.

Known genes with phenotypes (2200-2400)

Embryonic lethals (5000)

Intellectual disability or other high penetrance neuropsychiatric condition (1000)

No clinical phenotype (3000)

Phenotypes to be defined or characterized (8-10000)

WEB SITES

Main NCBI entry site <http://www.ncbi.nlm.nih.gov/>

UCSC genome browser <http://genome.ucsc.edu/>

HGMD curated mutation database
<http://www.hgmd.cf.ac.uk/ac/index.php>

CCDS gene annotation database

<http://www.ncbi.nlm.nih.gov/projects/CCDS/CcidsBrowse.cgi>

Mammalian Gene Collection database
<http://mgc.nci.nih.gov/>

Mitochondrial proteome database
<http://www.mitop.de:8080/mitop2/>

Canadian rare genetic diseases <http://raredisorders.ca/>

Comparative genome annotations
<http://ecrbrowser.dcode.org/>

GeneTests molecular diagnostic database
<http://www.ncbi.nlm.nih.gov/sites/GeneTests/?db=GeneTests>

BioGPS gene expression database
<http://biogps.gnf.org/#goto=welcome>

Budding yeast mutation database
http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html

Fission yeast mutation database
http://www.sanger.ac.uk/Projects/S_pombe/

Statistics Canada <http://www.statcan.gc.ca>

ACKNOWLEDGEMENTS

Many thanks to Mark Ludman, Christie Riddell, David Skidmore, Sarah Dyack, Andrew Orr, Grant Mitchell, Jacques Michaud, Gregor Andelfinger and Guy Rouleau for sharing insights gained from their wide clinical genetics experience. Thanks to Gerard Battail, and to Bill Gelbart, Pierre Drapeau, Louis Saint-Amant and Christian Neri for their insights into model organism developmental genetics. Thanks to D. Samuels for providing illustrations. MES is supported by Genome Canada, Genome Atlantic, Nova Scotia Health Research Foundation, Nova Scotia Research and Innovation Trust, IWK Health Centre Foundation, Dalhousie Medical Research Fund and Capital Health Research Fund, March of Dimes and the Centre de Recherche du CHU Ste-Justine.

REFERENCES

- [1] Freimer, N.; Sabatti, C. The human phenome project. *Nat. Genet.*, **2003**, *34*, 15-21.
- [2] Oti, M.; Huynen, M.A.; Brunner, H.G. Phenome connections. *Trends Genet.*, **2008**, *24*, 103-106.
- [3] Oti, M.; Huynen, M.A.; Brunner, H.G. The biological coherence of human phenome databases. *Am. J. Hum. Genet.*, **2009**, *85*, 801-808.
- [4] Sriver, C.R. After the genome--the phenome? *J. Inher. Metab. Dis.*, **2004**, *27*, 305-317.
- [5] Paulson, H.L. The spinocerebellar ataxias. *J. Neuroophthalmol.*, **2009**, *29*, 227-237.
- [6] Dror, A.A.; Avraham, K.B. Hearing loss: mechanisms revealed by genetics and cell biology. *Annu. Rev. Genet.*, **2009**, *43*, 411-437.
- [7] Matsunaga, T. Value of genetic testing in the otological approach for sensorineural hearing loss. *Keio J. Med.*, **2009**, *58*, 216-222.
- [8] Hilgert, N.; Smith, R.J.; Van Camp, G. Forty-six genes causing nonsyndromic hearing impairment: which ones should be analyzed in DNA diagnostics? *Mutat. Res.*, **2009**, *681*, 189-196.
- [9] Berger, W.; Kloeckener-Gruissem, B.; Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.*, **2010**, *29*(5), 335-375.
- [10] Phelan, J.K.; Bok, D. A brief review of retinitis pigmentosa and the identified retinitis pigmentosa genes. *Mol. Vis.*, **2000**, *6*, 116-124.
- [11] van Soest, S.; Westerveld, A.; de Jong, P.T.; Bleecker-Wagemakers, E.M.; Bergen, A.A. Retinitis pigmentosa: defined from a molecular point of view. *Surv. Ophthalmol.*, **1999**, *43*, 321-334.

- [12] Banchs, I.; Casanovas, C.; Alberti, A.; De Jorge, L.; Povedano, M.; Montero, J.; Martinez-Matos, J.A.; Volpini, V. Diagnosis of Charcot-Marie-Tooth disease. *J. Biomed. Biotechnol.*, **2009**, *2009*, 985415.
- [13] Barisic, N.; Claeys, K.G.; Sirotkovic-Skerlev, M.; Lofgren, A.; Nelis, E.; De Jonghe, P.; Timmerman, V. Charcot-Marie-Tooth disease: a clinico-genetic confrontation. *Ann. Hum. Genet.*, **2008**, *72*, 416-441.
- [14] Casanovas, C.; Cano, L.M.; Alberti, A.; Cespedes, M.; Rigo, G. Charcot-Marie-tooth disease. *Foot Ankle Spec.*, **2008**, *1*, 350-354.
- [15] Jani-Acsadi, A.; Krajewski, K.; Shy, M.E. Charcot-Marie-Tooth neuropathies: diagnosis and management. *Semin. Neurol.*, **2008**, *28*, 185-194.
- [16] Pareyson, D.; Marchesi, C.; Salsano, E. Hereditary predominantly motor neuropathies. *Curr. Opin. Neurol.*, **2009**, *22*, 451-459.
- [17] Parman, Y. Hereditary neuropathies. *Curr. Opin. Neurol.*, **2007**, *20*, 542-547.
- [18] Reilly, M.M. Sorting out the inherited neuropathies. *Pract. Neurol.*, **2007**, *7*, 93-105.
- [19] Szigeti, K.; Nelis, E.; Lupski, J.R. Molecular diagnostics of Charcot-Marie-Tooth disease and related peripheral neuropathies. *Neuromolecular Med.*, **2006**, *8*, 243-254.
- [20] Dragatsis, I.; Efstratiadis, A.; Zeitlin, S. Mouse mutant embryos lacking huntingtin are rescued from lethality by wild-type extraembryonic tissues. *Development*, **1998**, *125*, 1529-1539.
- [21] Kajiwara, K.; Berson, E.L.; Dryja, T.P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science*, **1994**, *264*, 1604-1608.
- [22] Dryja, T.P.; Hahn, L.B.; Kajiwara, K.; Berson, E.L. Dominant and digenic mutations in the peripherin/RDS and ROM1 genes in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, **1997**, *38*, 1972-1982.
- [23] Goldberg, A.F.; Molday, R.S. Defective subunit assembly underlies a digenic form of retinitis pigmentosa linked to mutations in peripherin/rds and rom-1. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 13726-13730.
- [24] Jin, Z.B.; Mandai, M.; Yokota, T.; Higuchi, K.; Ohmori, K.; Ohtsuki, F.; Takakura, S.; Itabashi, T.; Wada, Y.; Akimoto, M.; Ooto, S.; Suzuki, T.; Hirami, Y.; Ikeda, H.; Kawagoe, N.; Oishi, A.; Ichiyama, S.; Takahashi, M.; Yoshimura, N.; Kosugi, S. Identifying pathogenic genetic background of simplex or multiplex retinitis pigmentosa patients: a large scale mutation screening study. *J. Med. Genet.*, **2008**, *45*, 465-472.
- [25] Kedzierski, W.; Nusinowitz, S.; Birch, D.; Clarke, G.; McInnes, R.R.; Bok, D.; Travis, G.H. Deficiency of rds/peripherin causes photoreceptor death in mouse models of digenic and dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 7718-7723.
- [26] Keen, T.J.; Inglehearn, C.F. Mutations and polymorphisms in the human peripherin-RDS gene and their involvement in inherited retinal degeneration. *Hum. Mutat.*, **1996**, *8*, 297-303.
- [27] Loewen, C.J.; Moritz, O.L.; Molday, R.S. Molecular characterization of peripherin-2 and rom-1 mutants responsible for digenic retinitis pigmentosa. *J. Biol. Chem.*, **2001**, *276*, 22388-22396.
- [28] Sullivan, L.S.; Bowne, S.J.; Birch, D.G.; Hughbanks-Wheaton, D.; Heckenlively, J.R.; Lewis, R.A.; Garcia, C.A.; Ruiz, R.S.; Blanton, S.H.; Northrup, H.; Gire, A.L.; Seaman, R.; Duzkale, H.; Spellacy, C.J.; Zhu, J.; Shankar, S.P.; Daiger, S.P. Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. *Invest. Ophthalmol. Vis. Sci.*, **2006**, *47*, 3052-3064.
- [29] Jacolot, S.; Le Gac, G.; Scotet, V.; Quere, I.; Mura, C.; Ferec, C. HAMP as a modifier gene that increases the phenotypic expression of the HFE pC282Y homozygous genotype. *Blood*, **2004**, *103*, 2835-2840.
- [30] Merryweather-Clarke, A.T.; Cadet, E.; Bomford, A.; Capron, D.; Viprakasit, V.; Miller, A.; McHugh, P.J.; Chapman, R.W.; Pointon, J.J.; Wimbhurst, V.L.; Livesey, K.J.; Tanphachitr, V.; Rochette, J.; Robson, K.J. Digenic inheritance of mutations in HAMP and HFE results in different types of haemochromatosis. *Hum. Mol. Genet.*, **2003**, *12*, 2241-2247.
- [31] Badano, J.L.; Kim, J.C.; Hoskins, B.E.; Lewis, R.A.; Ansley, S.J.; Cutler, D.J.; Castellani, C.; Beales, P.L.; Leroux, M.R.; Katsanis, N. Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistatic effect on Bardet-Biedl patients with two mutations at a second BBS locus. *Hum. Mol. Genet.*, **2003**, *12*, 1651-1659.
- [32] Badano, J.L.; Leitch, C.C.; Ansley, S.J.; May-Simera, H.; Lawson, S.; Lewis, R.A.; Beales, P.L.; Dietz, H.C.; Fisher, S.; Katsanis, N. Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature*, **2006**, *439*, 326-330.
- [33] Beales, P.L.; Badano, J.L.; Ross, A.J.; Ansley, S.J.; Hoskins, B.E.; Kirsten, B.; Mein, C.A.; Froguel, P.; Scambler, P.J.; Lewis, R.A.; Lupski, J.R.; Katsanis, N. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am. J. Hum. Genet.*, **2003**, *72*, 1187-1199.
- [34] Eichers, E.R.; Lewis, R.A.; Katsanis, N.; Lupski, J.R. Triallelic inheritance: a bridge between Mendelian and multifactorial traits. *Ann. Med.*, **2004**, *36*, 262-272.
- [35] Fauser, S.; Munz, M.; Besch, D. Further support for digenic inheritance in Bardet-Biedl syndrome. *J. Med. Genet.*, **2003**, *40*, e104.
- [36] Katsanis, N. The oligogenic properties of Bardet-Biedl syndrome. *Hum. Mol. Genet.*, **2004**, *13 Spec No 1*, R65-71.
- [37] Adato, A.; Kalinski, H.; Weil, D.; Chaib, H.; Korostishevsky, M.; Bonne-Tamir, B. Possible interaction between USH1B and USH3 gene products as implied by apparent digenic deafness inheritance. *Am. J. Hum. Genet.*, **1999**, *65*, 261-265.
- [38] Cama, E.; Melchionda, S.; Palladino, T.; Carella, M.; Santarelli, R.; Genovese, E.; Benetazzo, F.; Zelante, L.; Arslan, E. Hearing loss features in GJB2 biallelic mutations and GJB2/GJB6 digenic inheritance in a large Italian cohort. *Int. J. Audiol.*, **2009**, *48*, 12-17.
- [39] Chen, A.H.; Fukushima, K.; McGuirt, W.T.; Smith, R.J. DFNB15: autosomal recessive non-syndromic hearing loss gene-chromosome 3q, 19p or digenic recessive inheritance? *Adv. Otorhinolaryngol.*, **2000**, *56*, 171-175.
- [40] Liu, X.Z.; Yuan, Y.; Yan, D.; Ding, E.H.; Ouyang, X.M.; Fei, Y.; Tang, W.; Yuan, H.; Chang, Q.; Du, L.L.; Zhang, X.; Wang, G.; Ahmad, S.; Kang, D.Y.; Lin, X.; Dai, P. Digenic inheritance of non-syndromic deafness caused by mutations at the gap junction proteins Cx26 and Cx31. *Hum. Genet.*, **2009**, *125*, 53-62.
- [41] Morell, R.; Spritz, R.A.; Ho, L.; Pierpont, J.; Guo, W.; Friedman, T.B.; Asher, J.H., Jr. Apparent digenic inheritance of Waardenburg syndrome type 2 (WS2) and autosomal recessive ocular albinism (AROA). *Hum. Mol. Genet.*, **1997**, *6*, 659-664.
- [42] Nozu, K.; Inagaki, T.; Fu, X.J.; Nozu, Y.; Kaito, H.; Kanda, K.; Sekine, T.; Igarashi, T.; Nakanishi, K.; Yoshikawa, N.; Iijima, K.; Matsuo, M. Molecular analysis of digenic inheritance in Bartter syndrome with sensorineural deafness. *J. Med. Genet.*, **2008**, *45*, 182-186.
- [43] Pallares-Ruiz, N.; Blanchet, P.; Mondain, M.; Claustres, M.; Roux, A.F. A large deletion including most of GJB6 in recessive non syndromic deafness: a digenic effect? *Eur. J. Hum. Genet.*, **2002**, *10*, 72-76.
- [44] Stevenson, V.A.; Ito, M.; Milunsky, J.M. Connexin-30 deletion analysis in connexin-26 heterozygotes. *Genet. Test.*, **2003**, *7*, 151-154.
- [45] Yang, T.; Gurrola, J.G., 2nd; Wu, H.; Chiu, S.M.; Wangemann, P.; Snyder, P.M.; Smith, R.J. Mutations of KCNJ10 together with mutations of SLC26A4 cause digenic nonsyndromic hearing loss associated with enlarged vestibular aqueduct syndrome. *Am. J. Hum. Genet.*, **2009**, *84*, 651-657.
- [46] Zheng, Q.Y.; Yan, D.; Ouyang, X.M.; Du, L.L.; Yu, H.; Chang, B.; Johnson, K.R.; Liu, X.Z. Digenic inheritance of deafness caused by mutations in genes encoding cadherin 23 and protocadherin 15 in mice and humans. *Hum. Mol. Genet.*, **2005**, *14*, 103-111.
- [47] Weatherall, D.J. Single gene disorders or complex traits: lessons from the thalassaemias and other monogenic diseases. *BMJ*, **2000**, *321*, 1117-1120.
- [48] Weatherall, D.J. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat. Rev. Genet.*, **2001**, *2*, 245-255.
- [49] Scriver, C.R.; Waters, P.J. Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet.*, **1999**, *15*, 267-272.
- [50] Charlton, R.; Bothwell, T. Current clinical practice: hereditary haemochromatosis: recent advances in biology and management. *Hematology*, **1999**, *4*, 185-194.
- [51] Weatherall, M.W.; Higgs, D.R.; Weiss, H.; Weatherall, D.J.; Searjeant, G.R. Phenotype/genotype relationships in sickle cell disease: a pilot twin study. *Clin. Lab. Haematol.*, **2005**, *27*, 384-390.
- [52] Dipple, K.M.; McCabe, E.R. Modifier genes convert "simple" Mendelian disorders to complex traits. *Mol. Genet. Metab.*, **2000**, *71*, 43-50.

- [53] Dipple, K.M.; McCabe, E.R. Phenotypes of patients with "simple" Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. *Am. J. Hum. Genet.*, **2000**, *66*, 1729-1735.
- [54] Boyle, M.P. Strategies for identifying modifier genes in cystic fibrosis. *Proc. Am. Thorac. Soc.*, **2007**, *4*, 52-57.
- [55] Collaco, J.M.; Cutting, G.R. Update on gene modifiers in cystic fibrosis. *Curr. Opin. Pulm. Med.*, **2008**, *14*, 559-566.
- [56] Merlo, C.A.; Boyle, M.P. Modifier genes in cystic fibrosis lung disease. *J. Lab. Clin. Med.*, **2003**, *141*, 237-241.
- [57] Sliker, M.G.; Sanders, E.A.; Rijkers, G.T.; Ruven, H.J.; van der Ent, C.K. Disease modifying genes in cystic fibrosis. *J. Cyst. Fibros.*, **2005**, *4*(Suppl 2), 7-13.
- [58] Roy, C.N.; Andrews, N.C. Recent advances in disorders of iron metabolism: mutations, mechanisms and modifiers. *Hum. Mol. Genet.*, **2001**, *10*, 2181-2186.
- [59] Wood, M.J.; Powell, L.W.; Ramm, G.A. Environmental and genetic modifiers of the progression to fibrosis and cirrhosis in hemochromatosis. *Blood*, **2008**, *111*, 4456-4462.
- [60] Edwards, A.O.; Ritter, R., 3rd; Abel, K.J.; Manning, A.; Panhuysen, C.; Farrer, L.A. Complement factor H polymorphism and age-related macular degeneration. *Science*, **2005**, *308*, 421-424.
- [61] Hageman, G.S.; Anderson, D.H.; Johnson, L.V.; Hancox, L.S.; Taiber, A.J.; Hardisty, L.I.; Hageman, J.L.; Stockman, H.A.; Borchart, J.D.; Gehrs, K.M.; Smith, R.J.; Silvestri, G.; Russell, S.R.; Klaver, C.C.; Barbazetto, I.; Chang, S.; Yannuzzi, L.A.; Barile, G.R.; Merriam, J.C.; Smith, R.T.; Olsh, A.K.; Bergeron, J.; Zernant, J.; Merriam, J.E.; Gold, B.; Dean, M.; Allikmets, R. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 7227-7232.
- [62] Haines, J.L.; Hauser, M.A.; Schmidt, S.; Scott, W.K.; Olson, L.M.; Gallins, P.; Spencer, K.L.; Kwan, S.Y.; Noureddine, M.; Gilbert, J.R.; Schnetz-Boutaud, N.; Agarwal, A.; Postel, E.A.; Pericak-Vance, M.A. Complement factor H variant increases the risk of age-related macular degeneration. *Science*, **2005**, *308*, 419-421.
- [63] Klein, R.J.; Zeiss, C.; Chew, E.Y.; Tsai, J.Y.; Sackler, R.S.; Haynes, C.; Henning, A.K.; SanGiovanni, J.P.; Mane, S.M.; Mayne, S.T.; Bracken, M.B.; Ferris, F.L.; Ott, J.; Barnstable, C.; Hoh, J. Complement factor H polymorphism in age-related macular degeneration. *Science*, **2005**, *308*, 385-389.
- [64] Drayna, D.; Coon, H.; Kim, U.K.; Elsner, T.; Cromer, K.; Otterud, B.; Baird, L.; Peiffer, A.P.; Leppert, M. Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. *Hum. Genet.*, **2003**, *112*, 567-572.
- [65] Wooding, S.; Kim, U.K.; Bamshad, M.J.; Larsen, J.; Jorde, L.B.; Drayna, D. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am. J. Hum. Genet.*, **2004**, *74*, 637-646.
- [66] Eriksson, N.; Macpherson, J.M.; Tung, J.Y.; Hon, L.S.; Naughton, B.; Saxonov, S.; Avey, L.; Wojcicki, A.; Pe'er, I.; Mountain, J. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.*, **2010**, *6*, e1000993.
- [67] Weedon, M.N.; Frayling, T.M. Reaching new heights: insights into the genetics of human stature. *Trends Genet.*, **2008**, *24*, 595-603.
- [68] Sturm, R.A. Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.*, **2009**, *18*, R9-17.
- [69] Sturm, R.A.; Duffy, D.L.; Box, N.F.; Newton, R.A.; Shepherd, A.G.; Chen, W.; Marks, L.H.; Leonard, J.H.; Martin, N.G. Genetic association and cellular function of MC1R variant alleles in human pigmentation. *Ann. N. Y. Acad. Sci.*, **2003**, *994*, 348-358.
- [70] Sturm, R.A.; Larsson, M. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res.*, **2009**, *22*, 544-562.
- [71] Antequera, F.; Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 11995-11999.
- [72] Roest Crolius, H.; Jaillon, O.; Bernot, A.; Dasilva, C.; Bouneau, L.; Fischer, C.; Fizames, C.; Wincker, P.; Brottier, P.; Quetier, F.; Saurin, W.; Weissenbach, J. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat. Genet.*, **2000**, *25*, 235-238.
- [73] Imanishi, T.; Itoh, T.; Suzuki, Y.; O'Donovan, C.; Fukuchi, S.; Koyanagi, K.O.; Barrero, R.A.; Tamura, T.; Yamaguchi-Kabata, Y.; Tanino, M.; Yura, K.; Miyazaki, S.; Ikeo, K.; Homma, K.; Kasprzyk, A.; Nishikawa, T.; Hirakawa, M.; Thierry-Mieg, J.; Thierry-Mieg, D.; Ashurst, J.; Jia, L.; Nakao, M.; Thomas, M.A.; Mulder, N.; Karavidopoulou, Y.; Jin, L.; Kim, S.; Yasuda, T.; Lenhard, B.; Eveno, E.; Suzuki, Y.; Yamasaki, C.; Takeda, J.; Gough, C.; Hilton, P.; Fujii, Y.; Sakai, H.; Tanaka, S.; Amid, C.; Bellgard, M.; Bonaldo Mde, F.; Bono, H.; Bromberg, S.K.; Brookes, A.J.; Bruford, E.; Carninci, P.; Chelala, C.; Couillault, C.; de Souza, S.J.; Debily, M.A.; Devignes, M.D.; Dubchak, I.; Endo, T.; Estreicher, A.; Eyras, E.; Fukami-Kobayashi, K.; Gopinath, G.R.; Graudens, E.; Hahn, Y.; Han, M.; Han, Z.G.; Hanada, K.; Hanaoka, H.; Harada, E.; Hashimoto, K.; Hinz, U.; Hirai, M.; Hishiki, T.; Hopkinson, I.; Imbeaud, S.; Inoko, H.; Kanapin, A.; Kaneko, Y.; Kasukawa, T.; Kelso, J.; Kersey, P.; Kikuno, R.; Kimura, K.; Korn, B.; Kuryshev, V.; Makalowska, I.; Makino, T.; Mano, S.; Mariage-Samson, R.; Mashima, J.; Matsuda, H.; Mewes, H.W.; Minoshima, S.; Nagai, K.; Nagasaki, H.; Nagata, N.; Nigam, R.; Ogasawara, O.; Ohara, O.; Ohtsubo, M.; Okada, N.; Okido, T.; Oota, S.; Ota, M.; Ota, T.; Otsuki, T.; Piatier-Tonneau, D.; Poustka, A.; Ren, S.X.; Saitou, N.; Sakai, K.; Sakamoto, S.; Sakate, R.; Schupp, I.; Servant, F.; Sherry, S.; Shiba, R.; Shimizu, N.; Shimoyama, M.; Simpson, A.J.; Soares, B.; Steward, C.; Suwa, M.; Quackenbush, J.; Okazaki, T.; Tamiya, G.; Tanaka, H.; Taylor, T.; Terwilliger, J.D.; Unneberg, P.; Veeramachaneni, V.; Watanabe, S.; Wilming, L.; Yasuda, N.; Yoo, H.S.; Stodolsky, M.; Makalowski, W.; Go, M.; Nakai, K.; Takagi, T.; Kanehisa, M.; Sakaki, Y.; Quackenbush, J.; Okazaki, Y.; Hayashizaki, Y.; Hide, W.; Chakraborty, R.; Nishikawa, K.; Sugawara, H.; Tateno, Y.; Chen, Z.; Oishi, M.; Tonellato, P.; Apweiler, R.; Okubo, K.; Wagner, L.; Wiemann, S.; Strausberg, R.L.; Isogai, T.; Auffray, C.; Nomura, N.; Gojobori, T.; Sugano, S. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.*, **2004**, *2*, e162.
- [74] Yamasaki, C.; Koyanagi, K.O.; Fujii, Y.; Itoh, T.; Barrero, R.; Tamura, T.; Yamaguchi-Kabata, Y.; Tanino, M.; Takeda, J.; Fukuchi, S.; Miyazaki, S.; Nomura, N.; Sugano, S.; Imanishi, T.; Gojobori, T. Investigation of protein functions through data-mining on integrated human transcriptome database, H-Invitational database (H-InvDB). *Gene*, **2005**, *364*, 99-107.
- [75] Clamp, M.; Fry, B.; Kamal, M.; Xie, X.; Cuff, J.; Lin, M.F.; Kellis, M.; Lindblad-Toh, K.; Lander, E.S. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 19428-19433.
- [76] Ashurst, J.L.; Chen, C.K.; Gilbert, J.G.; Jekosch, K.; Keenan, S.; Meidl, P.; Searle, S.M.; Stalker, J.; Storey, R.; Trevanion, S.; Wilming, L.; Hubbard, T. The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res.*, **2005**, *33*, D459-65.
- [77] Wilming, L.G.; Gilbert, J.G.; Howe, K.; Trevanion, S.; Hubbard, T.; Harrow, J.L. The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.*, **2008**, *36*, D753-60.
- [78] Pruitt, K.D.; Harrow, J.; Harte, R.A.; Wallin, C.; Diekhans, M.; Maglott, D.R.; Searle, S.; Farrell, C.M.; Loveland, J.E.; Ruef, B.J.; Hart, E.; Suner, M.M.; Landrum, M.J.; Aken, B.; Ayling, S.; Baertsch, R.; Fernandez-Banet, J.; Cherry, J.L.; Curwen, V.; DiCuccio, M.; Kellis, M.; Lee, J.; Lin, M.F.; Schuster, M.; Shkeda, A.; Amid, C.; Brown, G.; Dukhanina, O.; Frankish, A.; Hart, J.; Maidak, B.L.; Mudge, J.; Murphy, M.R.; Murphy, T.; Rajan, J.; Rajput, B.; Riddick, L.D.; Snow, C.; Steward, C.; Webb, D.; Weber, J.A.; Wilming, L.; Wu, W.; Birney, E.; Haussler, D.; Hubbard, T.; Ostell, J.; Durbin, R.; Lipman, D. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.*, **2009**, *19*, 1316-1323.
- [79] Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **2008**, *40*, 1413-1415.
- [80] Sultan, M.; Schulz, M.H.; Richard, H.; Magen, A.; Klingenhoff, A.; Scherf, M.; Seifert, M.; Borodina, T.; Soldatov, A.; Parkhomchuk, D.; Schmidt, D.; O'Keefe, S.; Haas, S.; Vingron, M.; Lehrach, H.; Yaspo, M.L. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, **2008**, *321*, 956-960.
- [81] Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.; Coulson, A.R.; Drouin, J.; Eperon, I.C.; Nierlich, D.P.; Roe, B.A.; Sanger, F.; Schreier, P.H.; Smith, A.J.; Staden, R.; Young, I.G. Sequence and organization of the human mitochondrial genome. *Nature*, **1981**, *290*, 457-465.
- [82] Elstner, M.; Andreoli, C.; Ahting, U.; Tetko, I.; Klopstock, T.; Meitinger, T.; Prokisch, H. MitoP2: an integrative tool for the

- analysis of the mitochondrial proteome. *Mol. Biotechnol.*, **2008**, *40*, 306-315.
- [83] Elstner, M.; Andreoli, C.; Klopstock, T.; Meitinger, T.; Prokisch, H. The mitochondrial proteome database: MitoP2. *Methods Enzymol.*, **2009**, *457*, 3-20.
- [84] Miklos, G.L.; Rubin, G.M. The role of the genome project in determining gene function: insights from model organisms. *Cell*, **1996**, *86*, 521-529.
- [85] Kelly, D.E.; Lamb, D.C.; Kelly, S.L. Genome-wide generation of yeast gene deletion strains. *Comp. Funct. Genomics*, **2001**, *2*, 236-242.
- [86] Que, Q.Q.; Winzeler, E.A. Large-scale mutagenesis and functional genomics in yeast. *Funct. Integr. Genomics*, **2002**, *2*, 193-198.
- [87] Giaever, G.; Chu, A.M.; Ni, L.; Connelly, C.; Riles, L.; Veronneau, S.; Dow, S.; Luciau-Danila, A.; Anderson, K.; Andre, B.; Arkin, A.P.; Astromoff, A.; El-Bakkoury, M.; Bangham, R.; Benito, R.; Brachar, S.; Campanaro, S.; Curtiss, M.; Davis, K.; Deutschbauer, A.; Entian, K.D.; Flaherty, P.; Foury, F.; Garfinkel, D.J.; Gerstein, M.; Gotte, D.; Guldener, U.; Hegemann, J.H.; Hempel, S.; Herman, Z.; Jaramillo, D.F.; Kelly, D.E.; Kelly, S.L.; Kotter, P.; LaBonte, D.; Lamb, D.C.; Lan, N.; Liang, H.; Liao, H.; Liu, L.; Luo, C.; Lussier, M.; Mao, R.; Menard, P.; Ooi, S.L.; Revuelta, J.L.; Roberts, C.J.; Rose, M.; Ross-Macdonald, P.; Scherens, B.; Schimmack, G.; Shafer, B.; Shoemaker, D.D.; Sookhai-Mahadeo, S.; Storms, R.K.; Strathern, J.N.; Valle, G.; Voet, M.; Volckaert, G.; Wang, C.Y.; Ward, T.R.; Wilhelmly, J.; Winzeler, E.A.; Yang, Y.; Yen, G.; Youngman, E.; Yu, K.; Bussey, H.; Boeke, J.D.; Snyder, M.; Philippsen, P.; Davis, R.W.; Johnston, M. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **2002**, *418*, 387-391.
- [88] Costanzo, M.; Baryshnikova, A.; Bellay, J.; Kim, Y.; Spear, E.D.; Sevier, C.S.; Ding, H.; Koh, J.L.; Toufighi, K.; Mostafavi, S.; Prinz, J.; St Onge, R.P.; VanderSluis, B.; Makhnevych, T.; Vizeacoumar, F.J.; Alizadeh, S.; Bahr, S.; Brost, R.L.; Chen, Y.; Cokol, M.; Deshpande, R.; Li, Z.; Lin, Z.Y.; Liang, W.; Marback, M.; Paw, J.; San Luis, B.J.; Shuteriqi, E.; Tong, A.H.; van Dyk, N.; Wallace, I.M.; Whitney, J.A.; Weirauch, M.T.; Zhong, G.; Zhu, H.; Houry, W.A.; Brudno, M.; Ragibizadeh, S.; Papp, B.; Pal, C.; Roth, F.P.; Giaever, G.; Nislow, C.; Troyanskaya, O.G.; Bussey, H.; Bader, G.D.; Gingras, A.C.; Morris, Q.D.; Kim, P.M.; Kaiser, C.A.; Myers, C.L.; Andrews, B.J.; Boone, C. The genetic landscape of a cell. *Science*, **2010**, *327*, 425-431.
- [89] Adams, M.D.; Celniker, S.E.; Holt, R.A.; Evans, C.A.; Gocayne, J.D.; Amanatides, P.G.; Scherer, S.E.; Li, P.W.; Hoskins, R.A.; Galle, R.F.; George, R.A.; Lewis, S.E.; Richards, S.; Ashburner, M.; Henderson, S.N.; Sutton, G.G.; Wortman, J.R.; Yandell, M.D.; Zhang, Q.; Chen, L.X.; Brandon, R.C.; Rogers, Y.H.; Blazej, R.G.; Champe, M.; Pfeiffer, B.D.; Wan, K.H.; Doyle, C.; Baxter, E.G.; Helt, G.; Nelson, C.R.; Gabor, G.L.; Abril, J.F.; Agbayani, A.; An, H.J.; Andrews-Pfannkoch, C.; Baldwin, D.; Ballew, R.M.; Basu, A.; Baxendale, J.; Bayraktaroglu, L.; Beasley, E.M.; Beeson, K.Y.; Benos, P.V.; Berman, B.P.; Bhandari, D.; Bolshakov, S.; Borkova, D.; Botchan, M.R.; Boucek, J.; Brokstein, P.; Brottier, P.; Burtis, K.C.; Busam, D.A.; Butler, H.; Cadieu, E.; Center, A.; Chandra, I.; Cherry, J.M.; Cawley, S.; Dahlke, C.; Davenport, L.B.; Davies, P.; de Pablos, B.; Delcher, A.; Deng, Z.; Mays, A.D.; Dew, I.; Dietz, S.M.; Dodson, K.; Doup, L.E.; Downes, M.; Dugan-Rocha, S.; Dunkov, B.C.; Dunn, P.; Durbin, K.J.; Evangelista, C.C.; Ferraz, C.; Ferriera, S.; Fleischmann, W.; Fosler, C.; Gabrielian, A.E.; Garg, N.S.; Gelbart, W.M.; Glasser, K.; Glodek, A.; Gong, F.; Gorrell, J.H.; Gu, Z.; Guan, P.; Harris, M.; Harris, N.L.; Harvey, D.; Heiman, T.J.; Hernandez, J.R.; Houck, J.; Hostin, D.; Houston, K.A.; Howland, T.J.; Wei, M.H.; Ibegwam, C.; Jalali, M.; Kalush, F.; Karpen, G.H.; Ke, Z.; Kennison, J.A.; Ketchum, K.A.; Kimmel, B.E.; Kodira, C.D.; Kraft, C.; Kravitz, S.; Kulp, D.; Lai, Z.; Lasko, P.; Lei, Y.; Levitsky, A.A.; Li, J.; Li, Z.; Liang, Y.; Lin, X.; Liu, X.; Mattei, B.; McIntosh, T.C.; McLeod, M.P.; McPherson, D.; Merkulov, G.; Milshina, N.V.; Mobarry, C.; Morris, J.; Moshrefi, A.; Mout, S.M.; Moy, M.; Murphy, B.; Murphy, L.; Muzny, D.M.; Nelson, D.L.; Nelson, D.R.; Nelson, K.A.; Nixon, K.; Nusskern, D.R.; Pacleb, J.M.; Palazzolo, M.; Pittman, G.S.; Pan, S.; Pollard, J.; Puri, V.; Reese, M.G.; Reinert, K.; Remington, K.; Saunders, R.D.; Scheeler, F.; Shen, H.; Shue, B.C.; Siden-Kiamos, I.; Simpson, M.; Skupski, M.P.; Smith, T.; Spier, E.; Spradling, A.C.; Stapleton, M.; Strong, R.; Sun, E.; Svirskaas, R.; Tector, C.; Turner, R.; Venter, E.; Wang, A.H.; Wang, X.; Wang, Z.Y.; Wasarman, D.A.; Weinstock, G.M.; Weissenbach, J.; Williams, S.M.; Woodage, T.; Worley, K.C.; Wu, D.; Yang, S.; Yao, Q.A.; Ye, J.; Yeh, R.F.; Zaveri, J.S.; Zhan, M.; Zhang, G.; Zhao, Q.; Zheng, L.; Zheng, X.H.; Zhong, F.N.; Zhong, W.; Zhou, X.; Zhu, S.; Zhu, X.; Smith, H.O.; Gibbs, R.A.; Myers, E.W.; Rubin, G.M.; Venter, J.C. The genome sequence of *Drosophila melanogaster*. *Science*, **2000**, *287*, 2185-2195.
- [90] Liu, C.P.; Lim, J.K. Complementation analysis of methyl methane-sulfonate-induced recessive lethal mutations in the zeste-white region of the X chromosome of *Drosophila melanogaster*. *Genetics*, **1975**, *79*, 601-611.
- [91] Lim, J.K.; Snyder, L.A. Cytogenetic and complementation analyses of recessive lethal mutations induced in the X chromosome of *Drosophila* by three alkylating agents. *Genet. Res.*, **1974**, *24*, 1-10.
- [92] Ashburner, M.; Misra, S.; Roote, J.; Lewis, S.E.; Blazej, R.; Davis, T.; Doyle, C.; Galle, R.; George, R.; Harris, N.; Hartzell, G.; Harvey, D.; Hong, L.; Houston, K.; Hoskins, R.; Johnson, G.; Martin, C.; Moshrefi, A.; Palazzolo, M.; Reese, M.G.; Spradling, A.; Tsang, G.; Wan, K.; Whitelaw, K.; Celniker, S.; *et al.* An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the Adh region. *Genetics*, **1999**, *153*, 179-219.
- [93] Cooley, L.; Kelley, R.; Spradling, A. Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science*, **1988**, *239*, 1121-1128.
- [94] Myrick, K.V.; Huet, F.; Mohr, S.E.; Alvarez-Garcia, I.; Lu, J.T.; Smith, M.A.; Crosby, M.A.; Gelbart, W.M. Large-scale functional annotation and expanded implementations of the P{wHy} hybrid transposon in the *Drosophila melanogaster* genome. *Genetics*, **2009**, *182*, 653-660.
- [95] Hacker, U.; Nystedt, S.; Barmchi, M.P.; Horn, C.; Wimmer, E.A. piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 7720-7725.
- [96] Thibault, S.T.; Singer, M.A.; Miyazaki, W.Y.; Milash, B.; Dompe, N.A.; Singh, C.M.; Buchholz, R.; Demsky, M.; Fawcett, R.; Francis-Lang, H.L.; Ryner, L.; Cheung, L.M.; Chong, A.; Erickson, C.; Fisher, W.W.; Greer, K.; Hartouni, S.R.; Howie, E.; Jakkula, L.; Joo, D.; Killpack, K.; Laufer, A.; Mazzotta, J.; Smith, R.D.; Stevens, L.M.; Stuber, C.; Tan, L.R.; Ventura, R.; Woo, A.; Zakrajsek, I.; Zhao, L.; Chen, F.; Swimmer, C.; Kopczyński, C.; Duyk, G.; Winberg, M.L.; Margolis, J. A complementary transposon tool kit for *Drosophila melanogaster* using P and piggyBac. *Nat. Genet.*, **2004**, *36*, 283-287.
- [97] St Johnston, D. The art and design of genetic screens: *Drosophila melanogaster*. *Nat. Rev. Genet.*, **2002**, *3*, 176-188.
- [98] Ryder, E.; Russell, S. Transposable elements as tools for genomics and genetics in *Drosophila*. *Brief Funct. Genomic Proteomic.*, **2003**, *2*, 57-71.
- [99] Sulston, J.E.; Albertson, D.G.; Thomson, J.N. The *Caenorhabditis elegans* male: postembryonic development of nongonadal structures. *Dev. Biol.*, **1980**, *78*, 542-576.
- [100] Sulston, J.E.; Schierenberg, E.; White, J.G.; Thomson, J.N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.*, **1983**, *100*, 64-119.
- [101] Hillier, L.W.; Coulson, A.; Murray, J.I.; Bao, Z.; Sulston, J.E.; Waterston, R.H. Genomics in *C. elegans*: so many genes, such a little worm. *Genome Res.*, **2005**, *15*, 1651-1660.
- [102] Granger, L.; Martin, E.; Segalat, L. Mos as a tool for genome-wide insertional mutagenesis in *Caenorhabditis elegans*: results of a pilot study. *Nucleic Acids Res.*, **2004**, *32*, e117.
- [103] Montgomery, M.K.; Fire, A. Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet.*, **1998**, *14*, 255-258.
- [104] Fraser, A.G.; Kamath, R.S.; Zipperlen, P.; Martinez-Campos, M.; Sohrmann, M.; Ahringer, J. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature*, **2000**, *408*, 325-330.
- [105] Kamath, R.S.; Fraser, A.G.; Dong, Y.; Poulin, G.; Durbin, R.; Gotta, M.; Kanapin, A.; Le Bot, N.; Moreno, S.; Sohrmann, M.; Welchman, D.P.; Zipperlen, P.; Ahringer, J. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature*, **2003**, *421*, 231-237.
- [106] Alessandrini, F.; Jakob, T.; Wolf, A.; Wolf, E.; Balling, R.; Hrabe de Angelis, M.; Ring, J.; Behrendt, H. Enu mouse mutagenesis: generation of mouse mutants with aberrant plasma IgE levels. *Int. Arch. Allergy Immunol.*, **2001**, *124*, 25-28.

- [107] Barbaric, I.; Wells, S.; Russ, A.; Dear, T.N. Spectrum of ENU-induced mutations in phenotype-driven and gene-driven screens in the mouse. *Environ. Mol. Mutagen.*, **2007**, *48*, 124-142.
- [108] Boles, M.K.; Wilkinson, B.M.; Wilming, L.G.; Liu, B.; Probst, F.J.; Harrow, J.; Grafham, D.; Hentges, K.E.; Woodward, L.P.; Maxwell, A.; Mitchell, K.; Risley, M.D.; Johnson, R.; Hirsch, K.; Lupski, J.R.; Funato, Y.; Miki, H.; Marin-Garcia, P.; Matthews, L.; Coffey, A.J.; Parker, A.; Hubbard, T.J.; Rogers, J.; Bradley, A.; Adams, D.J.; Justice, M.J. Discovery of candidate disease genes in ENU-induced mouse mutants by large-scale sequencing, including a splice-site mutation in nucleoredoxin. *PLoS Genet.*, **2009**, *5*, e1000759.
- [109] Brown, S.D.; Balling, R. Systematic approaches to mouse mutagenesis. *Curr. Opin. Genet. Dev.*, **2001**, *11*, 268-273.
- [110] Clark, A.T.; Goldowitz, D.; Takahashi, J.S.; Vitaterna, M.H.; Siepka, S.M.; Peters, L.L.; Frankel, W.N.; Carlson, G.A.; Rossant, J.; Nadeau, J.H.; Justice, M.J. Implementing large-scale ENU mutagenesis screens in North America. *Genetica*, **2004**, *122*, 51-64.
- [111] Cordes, S.P. N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol. Mol. Biol. Rev.*, **2005**, *69*, 426-439.
- [112] Hrabe de Angelis, M.; Balling, R. Large scale ENU screens in the mouse: genetics meets genomics. *Mutat. Res.*, **1998**, *400*, 25-32.
- [113] Hrabe de Angelis, M.; Strivens, M. Large-scale production of mouse phenotypes: the search for animal models for inherited diseases in humans. *Brief Bioinform.*, **2001**, *2*, 170-180.
- [114] Hrabe de Angelis, M.H.; Flawinkel, H.; Fuchs, H.; Rathkolb, B.; Soewarto, D.; Marschall, S.; Heffner, S.; Pargent, W.; Wuensch, K.; Jung, M.; Reis, A.; Richter, T.; Alessandrini, F.; Jakob, T.; Fuchs, E.; Kolb, H.; Kremmer, E.; Schaeble, K.; Rollinski, B.; Roscher, A.; Peters, C.; Meitinger, T.; Strom, T.; Steckler, T.; Holsboer, F.; Klopstock, T.; Gekeler, F.; Schindewolf, C.; Jung, T.; Avraham, K.; Behrendt, H.; Ring, J.; Zimmer, A.; Schughart, K.; Pfeffer, K.; Wolf, E.; Balling, R. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet.*, **2000**, *25*, 444-447.
- [115] Nishimura, I.; Drake, T.A.; Lusic, A.J.; Lyons, K.M.; Nadeau, J.H.; Zernik, J. ENU large-scale mutagenesis and quantitative trait linkage (QTL) analysis in mice: novel technologies for searching polygenic determinants of craniofacial abnormalities. *Crit. Rev. Oral Biol. Med.*, **2003**, *14*, 320-330.
- [116] Nolan, P.M.; Peters, J.; Vizer, L.; Strivens, M.; Washbourne, R.; Hough, T.; Wells, C.; Glenister, P.; Thornton, C.; Martin, J.; Fisher, E.; Rogers, D.; Hagan, J.; Reavill, C.; Gray, I.; Wood, J.; Spurr, N.; Browne, M.; Rastan, S.; Hunter, J.; Brown, S.D. Implementation of a large-scale ENU mutagenesis program: towards increasing the mouse mutant resource. *Mamm. Genome*, **2000**, *11*, 500-506.
- [117] Rathkolb, B.; Fuchs, E.; Kolb, H.J.; Renner-Muller, I.; Krebs, O.; Balling, R.; Hrabe de Angelis, M.; Wolf, E. Large-scale N-ethyl-N-nitrosourea mutagenesis of mice--from phenotypes to genes. *Exp. Physiol.*, **2000**, *85*, 635-644.
- [118] Sakuraba, Y.; Sezutsu, H.; Takahashi, K.R.; Tsuchihashi, K.; Ichikawa, R.; Fujimoto, N.; Kaneko, S.; Nakai, Y.; Uchiyama, M.; Goda, N.; Motoi, R.; Ikeda, A.; Karashima, Y.; Inoue, M.; Kaneda, H.; Masuya, H.; Minowa, O.; Noguchi, H.; Toyoda, A.; Sakaki, Y.; Wakana, S.; Noda, T.; Shiroishi, T.; Gondo, Y. Molecular characterization of ENU mouse mutagenesis and archives. *Biochem. Biophys. Res. Commun.*, **2005**, *336*, 609-616.
- [119] Soewarto, D.; Fella, C.; Teubner, A.; Rathkolb, B.; Pargent, W.; Heffner, S.; Marschall, S.; Wolf, E.; Balling, R.; Hrabe de Angelis, M. The large-scale Munich ENU-mouse-mutagenesis screen. *Mamm. Genome*, **2000**, *11*, 507-510.
- [120] Stanford, W.L.; Cohn, J.B.; Cordes, S.P. Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.*, **2001**, *2*, 756-768.
- [121] Araki, M.; Araki, K.; Yamamura, K. International Gene Trap Project: towards gene-driven saturation mutagenesis in mice. *Curr. Pharm. Biotechnol.*, **2009**, *10*, 221-229.
- [122] Guan, C.; Ye, C.; Yang, X.; Gao, J. A review of current large-scale mouse knockout efforts. *Genesis*, **2010**, *48*, 73-85.
- [123] Nord, A.S.; Chang, P.J.; Conklin, B.R.; Cox, A.V.; Harper, C.A.; Hicks, G.G.; Huang, C.C.; Johns, S.J.; Kawamoto, M.; Liu, S.; Meng, E.C.; Morris, J.H.; Rossant, J.; Ruiz, P.; Skarnes, W.C.; Soriano, P.; Stanford, W.L.; Stryke, D.; von Melchner, H.; Wurst, W.; Yamamura, K.; Young, S.G.; Babbitt, P.C.; Ferrin, T.E. The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res.*, **2006**, *34*, D642-8.
- [124] Skarnes, W.C.; von Melchner, H.; Wurst, W.; Hicks, G.; Nord, A.S.; Cox, T.; Young, S.G.; Ruiz, P.; Soriano, P.; Tessier-Lavigne, M.; Conklin, B.R.; Stanford, W.L.; Rossant, J. A public gene trap resource for mouse functional genomics. *Nat. Genet.*, **2004**, *36*, 543-544.
- [125] Stanford, W.L.; Epp, T.; Reid, T.; Rossant, J. Gene trapping in embryonic stem cells. *Methods Enzymol.*, **2006**, *420*, 136-162.
- [126] Brandon, E.P.; Idzerda, R.L.; McKnight, G.S. Targeting the mouse genome: a compendium of knockouts (Part III). *Curr. Biol.*, **1995**, *5*, 873-881.
- [127] Brandon, E.P.; Idzerda, R.L.; McKnight, G.S. Targeting the mouse genome: a compendium of knockouts (Part II). *Curr. Biol.*, **1995**, *5*, 758-765.
- [128] Brandon, E.P.; Idzerda, R.L.; McKnight, G.S. Knockouts. Targeting the mouse genome: a compendium of knockouts (Part I). *Curr. Biol.*, **1995**, *5*, 625-634.
- [129] Hurst, L.D.; Smith, N.G. Do essential genes evolve slowly? *Curr. Biol.*, **1999**, *9*, 747-750.
- [130] Zambrowicz, B.P.; Sands, A.T. Knockouts model the 100 best-selling drugs--will they model the next 100? *Nat. Rev. Drug Discov.*, **2003**, *2*, 38-51.
- [131] Leroux, L.; Desbois, P.; Lamotte, L.; Duville, B.; Cordonnier, N.; Jackerott, M.; Jami, J.; Bucchini, D.; Joshi, R.L. Compensatory responses in mice carrying a null mutation for Ins1 or Ins2. *Diabetes*, **2001**, *50*(Suppl 1), S150-3.
- [132] Duville, B.; Cordonnier, N.; Deltour, L.; Dandoy-Dron, F.; Itier, J.M.; Monthieux, E.; Jami, J.; Joshi, R.L.; Bucchini, D. Phenotypic alterations in insulin-deficient mutant mice. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*, 5137-5140.
- [133] Joshi, R.L.; Lamothe, B.; Cordonnier, N.; Mesbah, K.; Monthieux, E.; Jami, J.; Bucchini, D. Targeted disruption of the insulin receptor gene in the mouse results in neonatal lethality. *EMBO J.*, **1996**, *15*, 1542-1547.
- [134] Accili, D.; Drago, J.; Lee, E.J.; Johnson, M.D.; Cool, M.H.; Salvatore, P.; Asico, L.D.; Jose, P.A.; Taylor, S.I.; Westphal, H. Early neonatal death in mice homozygous for a null allele of the insulin receptor gene. *Nat. Genet.*, **1996**, *12*, 106-109.
- [135] Mullins, M.C.; Hammerschmidt, M.; Haffter, P.; Nusslein-Volhard, C. Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr. Biol.*, **1994**, *4*, 189-202.
- [136] Haffter, P.; Nusslein-Volhard, C. Large scale genetics in a small vertebrate, the zebrafish. *Int. J. Dev. Biol.*, **1996**, *40*, 221-227.
- [137] Geisler, R.; Rauch, G.J.; Geiger-Rudolph, S.; Albrecht, A.; van Bebber, F.; Berger, A.; Busch-Nentwich, E.; Dahm, R.; Dekens, M.P.; Dooley, C.; Elli, A.F.; Gehring, I.; Geiger, H.; Geisler, M.; Glaser, S.; Holley, S.; Huber, M.; Kerr, A.; Kirm, A.; Knirsch, M.; Konantz, M.; Kuchler, A.M.; Maderspacher, F.; Neuhauss, S.C.; Nicolson, T.; Ober, E.A.; Praeg, E.; Ray, R.; Rentzsch, B.; Rick, J.M.; Rief, E.; Schauerte, H.E.; Schepp, C.P.; Schonberger, U.; Schonhaler, H.B.; Seiler, C.; Sidi, S.; Sollner, C.; Wehner, A.; Weiler, C.; Nusslein-Volhard, C. Large-scale mapping of mutations affecting zebrafish development. *BMC Genomics*, **2007**, *8*, 11.
- [138] de Bruijn, E.; Cuppen, E.; Feitsma, H. Highly Efficient ENU Mutagenesis in Zebrafish. *Methods Mol. Biol.*, **2009**, *546*, 3-12.
- [139] Draper, B.W.; McCallum, C.M.; Stout, J.L.; Slade, A.J.; Moens, C.B. A high-throughput method for identifying N-ethyl-N-nitrosourea (ENU)-induced point mutations in zebrafish. *Methods Cell Biol.*, **2004**, *77*, 91-112.
- [140] Grunwald, D.J.; Streisinger, G. Induction of recessive lethal and specific locus mutations in the zebrafish with ethyl nitrosourea. *Genet. Res.*, **1992**, *59*, 103-116.
- [141] Sood, R.; English, M.A.; Jones, M.; Mullikin, J.; Wang, D.M.; Anderson, M.; Wu, D.; Chandrasekharappa, S.C.; Yu, J.; Zhang, J.; Paul Liu, P. Methods for reverse genetic screening in zebrafish by resequencing and TILLING. *Methods*, **2006**, *39*, 220-227.
- [142] Henikoff, S.; Till, B.J.; Comai, L. TILLING. Traditional mutagenesis meets functional genomics. *Plant Physiol.*, **2004**, *135*, 630-636.
- [143] Moens, C.B.; Donn, T.M.; Wolf-Saxon, E.R.; Ma, T.P. Reverse genetics in zebrafish by TILLING. *Brief Funct. Genomic. Proteomic.*, **2008**, *7*, 454-459.
- [144] Nasevicius, A.; Ekker, S.C. Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.*, **2000**, *26*, 216-220.

- [145] George, R.A.; Smith, T.D.; Callaghan, S.; Hardman, L.; Pierides, C.; Horaitis, O.; Wouters, M.A.; Cotton, R.G. General mutation databases: analysis and review. *J. Med. Genet.*, **2008**, *45*, 65-70.
- [146] Hamosh, A.; Scott, A.F.; Amberger, J.; Valle, D.; McKusick, V.A. Online Mendelian Inheritance in Man (OMIM). *Hum. Mutat.*, **2000**, *15*, 57-61.
- [147] Amberger, J.; Bocchini, C.A.; Scott, A.F.; Hamosh, A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, **2009**, *37*, D793-6.
- [148] Stenson, P.D.; Mort, M.; Ball, E.V.; Howells, K.; Phillips, A.D.; Thomas, N.S.; Cooper, D.N. The Human Gene Mutation Database: 2008 update. *Genome Med.*, **2009**, *1*, 13.
- [149] Cooper, D.N.; Chen, J.M.; Ball, E.V.; Howells, K.; Mort, M.; Phillips, A.D.; Chuzhanova, N.; Krawczak, M.; Kehrer-Sawatzki, H.; Stenson, P.D. Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Hum. Mutat.*, **2010**, *31*, 631-655.
- [150] Ring, H.Z.; Kwok, P.Y.; Cotton, R.G. Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics*, **2006**, *7*, 969-972.
- [151] Fredman, D.; Munns, G.; Rios, D.; Sjöholm, F.; Siegfried, M.; Lenhard, B.; Lehvaslaiho, H.; Brookes, A.J. HGVBbase: a curated resource describing human DNA variation and phenotype relationships. *Nucleic Acids Res.*, **2004**, *32*, D516-9.
- [152] Drake, J.W.; Charlesworth, B.; Charlesworth, D.; Crow, J.F. Rates of spontaneous mutation. *Genetics*, **1998**, *148*, 1667-1686.
- [153] Nachman, M.W.; Crowell, S.L. Estimate of the mutation rate per nucleotide in humans. *Genetics*, **2000**, *156*, 297-304.
- [154] Xue, Y.; Wang, Q.; Long, Q.; Ng, B.L.; Swerdlow, H.; Burton, J.; Skuce, C.; Taylor, R.; Abdellah, Z.; Zhao, Y.; MacArthur, D.G.; Quail, M.A.; Carter, N.P.; Yang, H.; Tyler-Smith, C. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.*, **2009**, *19*, 1453-1457.
- [155] Roach, J.C.; Glusman, G.; Smit, A.F.; Huff, C.D.; Hubley, R.; Shannon, P.T.; Rowen, L.; Pant, K.P.; Goodman, N.; Bamshad, M.; Shendure, J.; Drmanac, R.; Jorde, L.B.; Hood, L.; Galas, D.J. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*, **2010**, *328*, 636-639.
- [156] Carson, A.R.; Feuk, L.; Mohammed, M.; Scherer, S.W. Strategies for the detection of copy number and other structural variants in the human genome. *Hum. Genomics*, **2006**, *2*, 403-414.
- [157] Feuk, L.; Carson, A.R.; Scherer, S.W. Structural variation in the human genome. *Nat. Rev. Genet.*, **2006**, *7*, 85-97.
- [158] Pinto, D.; Marshall, C.; Feuk, L.; Scherer, S.W. Copy-number variation in control population cohorts. *Hum. Mol. Genet.*, **2007**, *16 Spec No. 2*, R168-73.
- [159] Kumar, D. Disorders of the genome architecture: a review. *Genomic Med.*, **2008**, *2*, 69-76.
- [160] Kryukov, G.V.; Pennacchio, L.A.; Sunyaev, S.R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.*, **2007**, *80*, 727-739.
- [161] Sunyaev, S.; Ramensky, V.; Koch, I.; Lathe, W., 3rd; Kondrashov, A.S.; Bork, P. Prediction of deleterious human alleles. *Hum. Mol. Genet.*, **2001**, *10*, 591-597.
- [162] Yampolsky, L.Y.; Kondrashov, F.A.; Kondrashov, A.S. Distribution of the strength of selection against amino acid replacements in human proteins. *Hum. Mol. Genet.*, **2005**, *14*, 3191-3201.
- [163] Boyko, A.R.; Williamson, S.H.; Indap, A.R.; Degenhardt, J.D.; Hernandez, R.D.; Lohmueller, K.E.; Adams, M.D.; Schmidt, S.; Sninsky, J.J.; Sunyaev, S.R.; White, T.J.; Nielsen, R.; Clark, A.G.; Bustamante, C.D. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.*, **2008**, *4*, e1000083.
- [164] Eyre-Walker, A.; Keightley, P.D. High genomic deleterious mutation rates in hominids. *Nature*, **1999**, *397*, 344-347.
- [165] Kondrashov, A.S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.*, **2003**, *21*, 12-27.
- [166] Population Reference Bureau. World Population Data Sheet. **2009**.
- [167] Sankaranarayanan, K. Ionizing radiation and genetic risks IX. Estimates of the frequencies of mendelian diseases and spontaneous mutation rates in human populations: a 1998 perspective. *Mutat. Res.*, **1998**, *411*, 129-178.
- [168] Wilkie, A.O. The molecular basis of genetic dominance. *J. Med. Genet.*, **1994**, *31*, 89-98.
- [169] Frank, T.S.; Deffenbaugh, A.M.; Reid, J.E.; Hulick, M.; Ward, B.E.; Lingenfelter, B.; Gumper, K.L.; Scholl, T.; Tavtigian, S.V.; Pruss, D.R.; Critchfield, G.C. Clinical characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10,000 individuals. *J. Clin. Oncol.*, **2002**, *20*, 1480-1490.
- [170] Delatycki, M.B.; Williamson, R.; Forrest, S.M. Friedreich ataxia: an overview. *J. Med. Genet.*, **2000**, *37*, 1-8.
- [171] Vucic, S.; Kiernan, M.C. Pathophysiology of neurodegeneration in familial amyotrophic lateral sclerosis. *Curr. Mol. Med.*, **2009**, *9*, 255-272.
- [172] Julien, J.P.; Kriz, J. Transgenic mouse models of amyotrophic lateral sclerosis. *Biochim. Biophys. Acta*, **2006**, *1762*, 1013-1024.
- [173] Abifadel, M.; Rabes, J.P.; Devillers, M.; Munnich, A.; Erlich, D.; Junien, C.; Varret, M.; Boileau, C. Mutations and polymorphisms in the proprotein convertase subtilisin kexin 9 (PCSK9) gene in cholesterol metabolism and disease. *Hum. Mutat.*, **2009**, *30*, 520-529.
- [174] Soutar, A.K.; Naoumova, R.P. Mechanisms of disease: genetic causes of familial hypercholesterolemia. *Nat. Clin. Pract. Cardiovasc. Med.*, **2007**, *4*, 214-225.
- [175] Horton, J.D.; Cohen, J.C.; Hobbs, H.H. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem. Sci.*, **2007**, *32*, 71-77.
- [176] Rahalkar, A.R.; Hegele, R.A. Monogenic pediatric dyslipidemias: classification, genetics and clinical spectrum. *Mol. Genet. Metab.*, **2008**, *93*, 282-294.
- [177] Tarugi, P.; Averna, M.; Di Leo, E.; Cefalu, A.B.; Noto, D.; Magnolo, L.; Cattin, L.; Bertolini, S.; Calandra, S. Molecular diagnosis of hypobetalipoproteinemia: an ENID review. *Atherosclerosis*, **2007**, *195*, e19-27.
- [178] Whitfield, A.J.; Barrett, P.H.; van Bockxmeer, F.M.; Burnett, J.R. Lipid disorders and mutations in the APOB gene. *Clin. Chem.*, **2004**, *50*, 1725-1732.
- [179] Farese, R.V., Jr.; Linton, M.F.; Young, S.G. Apolipoprotein B gene mutations affecting cholesterol levels. *J. Intern. Med.*, **1992**, *231*, 643-652.
- [180] Varret, M.; Abifadel, M.; Rabes, J.P.; Boileau, C. Genetic heterogeneity of autosomal dominant hypercholesterolemia. *Clin. Genet.*, **2008**, *73*, 1-13.
- [181] Phadnis, N.; Fry, J.D. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics*, **2005**, *171*, 385-392.
- [182] Porteous, J.W. Dominance—one hundred and fifteen years after Mendel's paper. *J. Theor. Biol.*, **1996**, *182*, 223-232.
- [183] Wright, S. Physiological and evolutionary theories of dominance. *Am. Nat.*, **1934**, *68*, 24-53.
- [184] Kacser, H.; Burns, J.A. The control of flux. *Symp. Soc. Exp. Biol.*, **1973**, *27*, 65-104.
- [185] Kacser, H.; Burns, J.A. The molecular basis of dominance. *Genetics*, **1981**, *97*, 639-666.
- [186] Grossniklaus, U.; Madhusudhan, M.S.; Nanjundiah, V. Nonlinear enzyme kinetics can lead to high metabolic flux control coefficients: implications for the evolution of dominance. *J. Theor. Biol.*, **1996**, *182*, 299-302.
- [187] Bagheri, H.C.; Wagner, G.P. Evolution of dominance in metabolic pathways. *Genetics*, **2004**, *168*, 1713-1735.
- [188] Orr, H.A. A test of Fisher's theory of dominance. *Proc. Natl. Acad. Sci. USA*, **1991**, *88*, 11413-11415.
- [189] Charrow, J. Ashkenazi Jewish genetic disorders. *Fam. Cancer*, **2004**, *3*, 201-206.
- [190] Boycott, K.M.; Parboosingh, J.S.; Chodirker, B.N.; Lowry, R.B.; McLeod, D.R.; Morris, J.; Greenberg, C.R.; Chudley, A.E.; Bernier, F.P.; Midgley, J.; Moller, L.B.; Innes, A.M. Clinical genetics and the Hutterite population: a review of Mendelian disorders. *Am. J. Med. Genet. A*, **2008**, *146A*, 1088-1098.
- [191] Labege, A.M.; Michaud, J.; Richter, A.; Lemyre, E.; Lambert, M.; Brais, B.; Mitchell, G.A. Population history and its impact on medical genetics in Quebec. *Clin. Genet.*, **2005**, *68*, 287-301.
- [192] Peltonen, L.; Jalanko, A.; Varilo, T. Molecular genetics of the Finnish disease heritage. *Hum. Mol. Genet.*, **1999**, *8*, 1913-1923.
- [193] Zlotogora, J.; Shalev, S.; Habiballah, H.; Barjes, S. Genetic disorders among Palestinian Arabs: 3. Autosomal recessive disorders in a single village. *Am. J. Med. Genet.*, **2000**, *92*, 343-345.
- [194] Hassold, T.; Chen, N.; Funckhouser, J.; Jooss, T.; Manuel, B.; Matsuyama, J.; Matsuyama, A.; Wilson, C.; Yamane, J.A.; Jacobs, P.A.

- A cytogenetic study of 1000 spontaneous abortions. *Ann. Hum. Genet.*, **1980**, *44*, 151-178.
- [195] Philipp, T.; Philipp, K.; Reiner, A.; Beer, F.; Kalousek, D.K. Embryoscopic and cytogenetic analysis of 233 missed abortions: factors involved in the pathogenesis of developmental defects of early failed pregnancies. *Hum. Reprod.*, **2003**, *18*, 1724-1732.
- [196] Schaeffer, A.J.; Chung, J.; Heretis, K.; Wong, A.; Ledbetter, D.H.; Lese Martin, C. Comparative genomic hybridization-array analysis enhances the detection of aneuploidies and submicroscopic imbalances in spontaneous miscarriages. *Am. J. Hum. Genet.*, **2004**, *74*, 1168-1174.
- [197] Kosaki, K.; Kosaki, R.; Suzuki, T.; Yoshihashi, H.; Takahashi, T.; Sasaki, K.; Tomita, M.; McGinnis, W.; Matsuo, N. Complete mutation analysis panel of the 39 human HOX genes. *Teratology*, **2002**, *65*, 50-62.
- [198] Bosley, T.M.; Alorainy, I.A.; Salih, M.A.; Aldhalaan, H.M.; Abu-Amero, K.K.; Oystreck, D.T.; Tischfield, M.A.; Engle, E.C.; Erickson, R.P. The clinical spectrum of homozygous HOXA1 mutations. *Am. J. Med. Genet. A*, **2008**, *146A*, 1235-1240.
- [199] Alasti, F.; Sadeghi, A.; Sanati, M.H.; Farhadi, M.; Stollar, E.; Somers, T.; Van Camp, G. A mutation in HOXA2 is responsible for autosomal-recessive microtia in an Iranian family. *Am. J. Hum. Genet.*, **2008**, *82*, 982-991.
- [200] Horvat-Switzer, R.D.; Thompson, A.A. HOXA11 mutation in amegakaryocytic thrombocytopenia with radio-ulnar synostosis syndrome inhibits megakaryocytic differentiation *in vitro*. *Blood Cells Mol. Dis.*, **2006**, *37*, 55-63.
- [201] Goodman, F.R.; Bacchelli, C.; Brady, A.F.; Brueton, L.A.; Fryns, J.P.; Mortlock, D.P.; Innis, J.W.; Holmes, L.B.; Donnenfeld, A.E.; Feingold, M.; Beemer, F.A.; Hennekam, R.C.; Scambler, P.J. Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am. J. Hum. Genet.*, **2000**, *67*, 197-202.
- [202] Mortlock, D.P.; Innis, J.W. Mutation of HOXA13 in hand-foot-genital syndrome. *Nat. Genet.*, **1997**, *15*, 179-180.
- [203] Dobbs, M.B.; Gurnett, C.A.; Pierce, B.; Exner, G.U.; Robarge, J.; Morcuende, J.A.; Cole, W.G.; Templeton, P.A.; Foster, B.; Bowcock, A.M. HOXD10 M319K mutation in a family with isolated congenital vertical talus. *J. Orthop. Res.*, **2006**, *24*, 448-453.
- [204] Shrimpton, A.E.; Levinsohn, E.M.; Yozawitz, J.M.; Packard, D.S., Jr.; Cady, R.B.; Middleton, F.A.; Persico, A.M.; Hootnick, D.R. A HOX gene mutation in a family with isolated congenital vertical talus and Charcot-Marie-Tooth disease. *Am. J. Hum. Genet.*, **2004**, *75*, 92-96.
- [205] Muragaki, Y.; Mundlos, S.; Upton, J.; Olsen, B.R. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science*, **1996**, *272*, 548-551.
- [206] Bentley, D.R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.*, **2006**, *16*, 545-552.
- [207] Levy, S.; Sutton, G.; Ng, P.C.; Feuk, L.; Halpern, A.L.; Walenz, B.P.; Axelrod, N.; Huang, J.; Kirkness, E.F.; Denisov, G.; Lin, Y.; MacDonald, J.R.; Pang, A.W.; Shago, M.; Stockwell, T.B.; Tsismouri, A.; Bafna, V.; Bansal, V.; Kravitz, S.A.; Busam, D.A.; Beeson, K.Y.; McIntosh, T.C.; Remington, K.A.; Abril, J.F.; Gill, J.; Borman, J.; Rogers, Y.H.; Frazier, M.E.; Scherer, S.W.; Strausberg, R.L.; Venter, J.C. The diploid genome sequence of an individual human. *PLoS Biol.*, **2007**, *5*, e254.
- [208] Kim, J.I.; Ju, Y.S.; Park, H.; Kim, S.; Lee, S.; Yi, J.H.; Mudge, J.; Miller, N.A.; Hong, D.; Bell, C.J.; Kim, H.S.; Chung, I.S.; Lee, W.C.; Lee, J.S.; Seo, S.H.; Yun, J.Y.; Woo, H.N.; Lee, H.; Suh, D.; Lee, S.; Kim, H.J.; Yavartanoo, M.; Kwak, M.; Zheng, Y.; Lee, M.K.; Park, H.; Kim, J.Y.; Gokcumen, O.; Mills, R.E.; Zaranek, A.W.; Thakuria, J.; Wu, X.; Kim, R.W.; Huntley, J.J.; Luo, S.; Schroth, G.P.; Wu, T.D.; Kim, H.; Yang, K.S.; Park, W.Y.; Kim, H.; Church, G.M.; Lee, C.; Kingsmore, S.F.; Seo, J.S. A highly annotated whole-genome sequence of a Korean individual. *Nature*, **2009**, *460*, 1011-1015.
- [209] McKernan, K.J.; Peckham, H.E.; Costa, G.L.; McLaughlin, S.F.; Fu, Y.; Tsung, E.F.; Clouser, C.R.; Duncan, C.; Ichikawa, J.K.; Lee, C.C.; Zhang, Z.; Ranade, S.S.; Dimalanta, E.T.; Hyland, F.C.; Sokolsky, T.D.; Zhang, L.; Sheridan, A.; Fu, H.; Hendrickson, C.L.; Li, B.; Kotler, L.; Stuart, J.R.; Malek, J.A.; Manning, J.M.; Antipova, A.A.; Perez, D.S.; Moore, M.P.; Hayashibara, K.C.; Lyons, M.R.; Beaudoin, R.E.; Coleman, B.E.; Laptewicz, M.W.; Sannicandro, A.E.; Rhodes, M.D.; Gottimukkala, R.K.; Yang, S.; Bafna, V.; Bashir, A.; MacBride, A.; Alkan, C.; Kidd, J.M.; Eichler, E.E.; Reese, M.G.; De La Vega, F.M.; Blanchard, A.P. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.*, **2009**, *19*, 1527-1541.
- [210] Wang, J.; Wang, W.; Li, R.; Li, Y.; Tian, G.; Goodman, L.; Fan, W.; Zhang, J.; Li, J.; Zhang, J.; Guo, Y.; Feng, B.; Li, H.; Lu, Y.; Fang, X.; Liang, H.; Du, Z.; Li, D.; Zhao, Y.; Hu, Y.; Yang, Z.; Zheng, H.; Hellmann, I.; Inouye, M.; Pool, J.; Yi, X.; Zhao, J.; Duan, J.; Zhou, Y.; Qin, J.; Ma, L.; Li, G.; Yang, Z.; Zhang, G.; Yang, B.; Yu, C.; Liang, F.; Li, W.; Li, S.; Li, D.; Ni, P.; Ruan, J.; Li, Q.; Zhu, H.; Liu, D.; Lu, Z.; Li, N.; Guo, G.; Zhang, J.; Ye, J.; Fang, L.; Hao, Q.; Chen, Q.; Liang, Y.; Su, Y.; San, A.; Ping, C.; Yang, S.; Chen, F.; Li, L.; Zhou, K.; Zheng, H.; Ren, Y.; Yang, L.; Gao, Y.; Yang, G.; Li, Z.; Feng, X.; Kristiansen, K.; Wong, G.K.; Nielsen, R.; Durbin, R.; Bolund, L.; Zhang, X.; Li, S.; Yang, H.; Wang, J. The diploid genome sequence of an Asian individual. *Nature*, **2008**, *456*, 60-65.
- [211] Wheeler, D.A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y.J.; Makhijani, V.; Roth, G.T.; Gomes, X.; Tartaro, K.; Niazi, F.; Turcotte, C.L.; Irzyk, G.P.; Lupski, J.R.; Chinault, C.; Song, X.Z.; Liu, Y.; Yuan, Y.; Nazareth, L.; Qin, X.; Muzny, D.M.; Margulies, M.; Weinstock, G.M.; Gibbs, R.A.; Rothberg, J.M. The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **2008**, *452*, 872-876.
- [212] Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.; Brown, C.G.; Hall, K.P.; Evers, D.J.; Barnes, C.L.; Bignell, H.R.; Boutell, J.M.; Bryant, J.; Carter, R.J.; Keira Cheetham, R.; Cox, A.J.; Ellis, D.J.; Flatbush, M.R.; Gormley, N.A.; Humphray, S.J.; Irving, L.J.; Karmali, M.S.; Kirk, S.M.; Li, H.; Liu, X.; Maizinger, K.S.; Murray, L.J.; Obradovic, B.; Ost, T.; Parkinson, M.L.; Pratt, M.R.; Rasolonjatovo, I.M.; Reed, M.T.; Rigatti, R.; Rodighiero, C.; Ross, M.T.; Sabot, A.; Sankar, S.V.; Scally, A.; Schroth, G.P.; Smith, M.E.; Smith, V.P.; Spiridou, A.; Torrance, P.E.; Tzovev, S.S.; Vermaas, E.H.; Walter, K.; Wu, X.; Zhang, L.; Alam, M.D.; Anastasi, C.; Aniebo, I.C.; Bailey, D.M.; Bancarz, I.R.; Banerjee, S.; Barbour, S.G.; Baybayan, P.A.; Benoit, V.A.; Benson, K.F.; Bevis, C.; Black, P.J.; Boodhun, A.; Brennan, J.S.; Bridgman, J.A.; Brown, R.C.; Brown, A.A.; Buermann, D.H.; Bundu, A.A.; Burrows, J.C.; Carter, N.P.; Castillo, N.; Chiara, E.C.M.; Chang, S.; Neil Cooley, R.; Crake, N.R.; Dada, O.O.; Diakoumakos, K.D.; Dominguez-Fernandez, B.; Earnshaw, D.J.; Egbujor, U.C.; Elmore, D.W.; Etchin, S.S.; Ewan, M.R.; Fedurco, M.; Fraser, L.J.; Fuentes Fajardo, K.V.; Scott Furey, W.; George, D.; Gietzen, K.J.; Goddard, C.P.; Golda, G.S.; Granieri, P.A.; Green, D.E.; Gustafson, D.L.; Hansen, N.F.; Harnish, K.; Haudenschild, C.D.; Heyer, N.I.; Hims, M.M.; Ho, J.T.; Horgan, A.M.; Hoshler, K.; Hurwitz, S.; Ivanov, D.V.; Johnson, M.Q.; James, T.; Huw Jones, T.A.; Kang, G.D.; Kerelska, T.H.; Kersey, A.D.; Khrebtkova, I.; Kindwall, A.P.; Kingsbury, Z.; Kokko-Gonzales, P.I.; Kumar, A.; Laurent, M.A.; Lawley, C.T.; Lee, S.E.; Lee, X.; Liao, A.K.; Loch, J.A.; Lok, M.; Luo, S.; Mammen, R.M.; Martin, J.W.; McCauley, P.G.; McNitt, P.; Mehta, P.; Moon, K.W.; Mullens, J.W.; Newington, T.; Ning, Z.; Ling Ng, B.; Novo, S.M.; O'Neill, M.J.; Osborne, M.A.; Osnowski, A.; Ostadan, O.; Paraschos, L.L.; Pickering, L.; Pike, A.C.; Pike, A.C.; Chris Pinkard, D.; Pliskin, D.P.; Podhasky, J.; Quijano, V.J.; Raczy, C.; Rae, V.H.; Rawlings, S.R.; Chiva Rodriguez, A.; Roe, P.M.; Rogers, J.; Rogert Bacigalupo, M.C.; Romanov, N.; Romieu, A.; Roth, R.K.; Rourke, N.J.; Ruediger, S.T.; Rusman, E.; Sanches-Kuiper, R.M.; Schenker, M.R.; Seoane, J.M.; Shaw, R.J.; Shiver, M.K.; Short, S.W.; Sizto, N.L.; Sluis, J.P.; Smith, M.A.; Ernest Sohna, J.; Spence, E.J.; Stevens, K.; Sutton, N.; Szajkowski, L.; Tregidgo, C.L.; Turcatti, G.; Vandevondele, S.; Verhovskiy, Y.; Virk, S.M.; Wakelin, S.; Walcott, G.C.; Wang, J.; Worsley, G.J.; Yan, J.; Yau, L.; Zuerlein, M.; Rogers, J.; Mullikin, J.C.; Hurles, M.E.; McCooke, N.J.; West, J.S.; Oaks, F.L.; Lundberg, P.L.; Klennerman, D.; Durbin, R.; Smith, A.J. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **2008**, *456*, 53-59.
- [213] Ahn, S.M.; Kim, T.H.; Lee, S.; Kim, D.; Ghang, H.; Kim, D.S.; Kim, B.C.; Kim, S.Y.; Kim, W.Y.; Kim, C.; Park, D.; Lee, Y.S.; Kim, S.; Reja, R.; Jho, S.; Kim, C.G.; Cha, J.Y.; Kim, K.H.; Lee, B.; Bhak, J.; Kim, S.J. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res.*, **2009**, *19*, 1622-1629.
- [214] Bau, S.; Schracke, N.; Kranzle, M.; Wu, H.; Stahler, P.F.; Hoheisel, J.D.; Beier, M.; Summerer, D. Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume

- microfluidic DNA arrays. *Anal. Bioanal. Chem.*, **2009**, 393, 171-175.
- [215] Porreca, G.J.; Zhang, K.; Li, J.B.; Xie, B.; Austin, D.; Vassallo, S.L.; LeProust, E.M.; Peck, B.J.; Emig, C.J.; Dahl, F.; Gao, Y.; Church, G.M.; Shendure, J. Multiplex amplification of large sets of human exons. *Nat. Methods*, **2007**, 4, 931-936.
- [216] Hodges, E.; Rooks, M.; Xuan, Z.; Bhattacharjee, A.; Benjamin Gordon, D.; Brizuela, L.; Richard McCombie, W.; Hannon, G.J. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat. Protoc.*, **2009**, 4, 960-974.
- [217] Hodges, E.; Xuan, Z.; Balija, V.; Kramer, M.; Molla, M.N.; Smith, S.W.; Middle, C.M.; Rodesch, M.J.; Albert, T.J.; Hannon, G.J.; McCombie, W.R. Genome-wide in situ exon capture for selective resequencing. *Nat. Genet.*, **2007**, 39, 1522-1527.
- [218] Via, M.; Gignoux, C.; Burchard, E.G. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med.*, **2010**, 2, 3.
- [219] Siva, N. 1000 Genomes project. *Nat. Biotechnol.*, **2008**, 26, 256.
- [220] Kuehn, B.M. 1000 Genomes Project promises closer look at variation in human genome. *JAMA*, **2008**, 300, 2715.
- [221] Vasta, V.; Ng, S.B.; Turner, E.H.; Shendure, J.; Hahn, S.H. Next generation sequence analysis for mitochondrial disorders. *Genome Med.*, **2009**, 1, 100.
- [222] Hedges, D.J.; Burges, D.; Powell, E.; Almonte, C.; Huang, J.; Young, S.; Boese, B.; Schmidt, M.; Pericak-Vance, M.A.; Martin, E.; Zhang, X.; Harkins, T.T.; Zuchner, S. Exome sequencing of a multigenerational human pedigree. *PLoS One*, **2009**, 4, e8232.
- [223] Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; Shendure, J.; Bamshad, M.J. Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.*, **2010**, 42, 30-5.
- [224] Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; Bamshad, M.; Nickerson, D.A.; Shendure, J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, **2009**, 461, 272-276.
- [225] Choi, M.; Scholl, U.I.; Ji, W.; Liu, T.; Tikhonova, I.R.; Zumbo, P.; Nayir, A.; Bakkaloglu, A.; Ozen, S.; Sanjad, S.; Nelson-Williams, C.; Farhi, A.; Mane, S.; Lifton, R.P. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA*, **2009**, 106, 19096-19101.
- [226] Hoischen, A.; Gilissen, C.; Arts, P.; Wieskamp, N.; van der Vliet, W.; Vermeer, S.; Steehouwer, M.; de Vries, P.; Meijer, R.; Sequeiros, J.; Knoers, N.V.; Buckley, M.F.; Scheffer, H.; Veltman, J.A. Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum. Mutat.*, **2010**, 31, 494-499.
- [227] Raca, G.; Jackson, C.; Warman, B.; Bair, T.; Schimmenti, L.A. Next generation sequencing in research and diagnostics of ocular birth defects. *Mol. Genet. Metab.*, **2010**, 100(2), 184-192.
- [228] Biesecker, L.G. Exome sequencing makes medical genomics a reality. *Nat. Genet.*, **2010**, 42, 13-14.
- [229] Lalonde, E.; Albrecht, S.; Ha, K.C.; Jacob, K.; Bolduc, N.; Polychronakos, C.; Dechelotte, P.; Majewski, J.; Jabado, N. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. *Hum. Mutat.*, **2010**, [Epub ahead of print].
- [230] Gabory, A.; Ripoché, M.A.; Yoshimizu, T.; Dandolo, L. The H19 gene: regulation and function of a non-coding RNA. *Cytogenet. Genome Res.*, **2006**, 113, 188-193.
- [231] Brown, C.J.; Ballabio, A.; Rupert, J.L.; Lafreniere, R.G.; Groppe, M.; Tonlorenzi, R.; Willard, H.F. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **1991**, 349, 38-44.
- [232] Leeb, M.; Steffen, P.A.; Wutz, A. X chromosome inactivation sparked by non-coding RNAs. *RNA Biol.*, **2009**, 6, 94-99.
- [233] Senner, C.E.; Brockdorff, N. Xist gene regulation at the onset of X inactivation. *Curr. Opin. Genet. Dev.*, **2009**, 19, 122-126.
- [234] DiMauro, S. Lessons from mitochondrial DNA mutations. *Semin. Cell Dev. Biol.*, **2001**, 12, 397-405.
- [235] Tuppen, H.A.; Blakely, E.L.; Turnbull, D.M.; Taylor, R.W. Mitochondrial DNA mutations and human disease. *Biochim. Biophys. Acta*, **2010**, 1797, 113-128.
- [236] Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorf, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; Cho, J.H.; Guttmacher, A.E.; Kong, A.; Kruglyak, L.; Mardis, E.; Rotimi, C.N.; Slatkin, M.; Valle, D.; Whittemore, A.S.; Boehnke, M.; Clark, A.G.; Eichler, E.E.; Gibson, G.; Haines, J.L.; Mackay, T.F.; McCarroll, S.A.; Visscher, P.M. Finding the missing heritability of complex diseases. *Nature*, **2009**, 461, 747-753.
- [237] Brinkman, R.R.; Dube, M.P.; Rouleau, G.A.; Orr, A.C.; Samuels, M.E. Human monogenic disorders - a source of novel drug targets. *Nat. Rev. Genet.*, **2006**, 7, 249-260.
- [238] Shimkets, R.A.; Lifton, R.P. Recent advances in the molecular genetics of hypertension. *Curr. Opin. Nephrol. Hypertens.*, **1996**, 5, 162-165.
- [239] Porter, J.R.; Barrett, T.G. Monogenic syndromes of abnormal glucose homeostasis: clinical review and relevance to the understanding of the pathology of insulin resistance and beta cell failure. *J. Med. Genet.*, **2005**, 42, 893-902.
- [240] Lifton, R.P. Genetic dissection of human blood pressure variation: common pathways from rare phenotypes. *Harvey Lect.*, **2004**, 100, 71-101.
- [241] Ahituv, N.; Kavaslar, N.; Schackwitz, W.; Ustaszewska, A.; Martin, J.; Hebert, S.; Doelle, H.; Ersoy, B.; Kryukov, G.; Schmidt, S.; Yosef, N.; Ruppin, E.; Sharan, R.; Vaisse, C.; Sunyaev, S.; Dent, R.; Cohen, J.; McPherson, R.; Pennacchio, L.A. Medical sequencing at the extremes of human body mass. *Am. J. Hum. Genet.*, **2007**, 80, 779-791.
- [242] Cohen, J.C.; Kiss, R.S.; Pertsemlidis, A.; Marcel, Y.L.; McPherson, R.; Hobbs, H.H. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*, **2004**, 305, 869-872.
- [243] Cohen, J.C.; Pertsemlidis, A.; Fahmi, S.; Esmail, S.; Vega, G.L.; Grundy, S.M.; Hobbs, H.H. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc. Natl. Acad. Sci. USA*, **2006**, 103, 1810-1815.
- [244] Romeo, S.; Pennacchio, L.A.; Fu, Y.; Boerwinkle, E.; Tybjaerg-Hansen, A.; Hobbs, H.H.; Cohen, J.C. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat. Genet.*, **2007**, 39, 513-516.
- [245] Ji, W.; Foo, J.N.; O'Roak, B.J.; Zhao, H.; Larson, M.G.; Simon, D.B.; Newton-Cheh, C.; State, M.W.; Levy, D.; Lifton, R.P. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet.*, **2008**, 40, 592-599.
- [246] McCandless, S.E.; Brunger, J.W.; Cassidy, S.B. The burden of genetic disease on inpatient care in a children's hospital. *Am. J. Hum. Genet.*, **2004**, 74, 121-127.
- [247] Scriver, C.R.; Neal, J.L.; Saginur, R.; Clow, A. The frequency of genetic disease and congenital malformation among patients in a pediatric hospital. *Can. Med. Assoc. J.*, **1973**, 108, 1111-1115.