REVIEWS

Human monogenic disorders — a source of novel drug targets

Ryan R. Brinkman, Marie-Pierre Dubé‡, Guy A. Rouleau§, Andrew C. Orr‖ and Mark E. Samuels¶*

Abstract | The decrease in new drug applications and approvals over the past several years results from an underlying crisis in drug target identification and validation. Model organisms are being used to address this problem, in combination with novel approaches such as the International HapMap Project. What has been underappreciated is that discovery of new drug targets can also be revived by traditional Mendelian genetics. A large fraction of the human gene repertoire remains phenotypically uncharacterized, and is likely to encode many unanticipated and novel phenotypes that will be of interest to pharmaceutical and biotechnological drug developers.

Haplotype mapping
A technique that involves the use of combinations of 'common' DNA polymorphisms to find blocks of association with phenotypic traits.

**British Columbia Cancer Research Centre, University of British Columbia, Vancouver, British Columbia V5Z 1C3, Canada.*
‡Montreal Heart Institute Research Centre, University of Montreal, Montreal, Quebec H1T 1C8, Canada.
§Department of Medicine, University of Montreal.
‖Department of Ophthalmology & Visual Sciences, Dalhousie University, Halifax, Nova Scotia B3H 2Y9, Canada.
¶Departments of Ophthalmology & Visual Sciences, Pediatrics and Psychiatry, Dalhousie University.
Correspondence to M.E.S.
e-mail: Mark.Samuels@iwk.nshealth.ca
doi:10.1038/nrg1828
Published online
14 March 2006

Recent discussions about the state of the pharmaceutical industry have decried annual decreases in new drug applications and new drugs coming to market[1,2]. Lack of novel drugs and drug candidates in the pipeline has become a concern for industry and medical professionals alike. There are many reasons for this trend, among them the use of inadequately validated targets[3]. Target identification and validation is a crucial first step in developing a drug against a given disease[4,5]. Historically, many drugs were developed without a clear understanding of the molecular mechanisms or targets. Today this is seen as a highly inefficient (and expensive) way to screen chemical compounds for biological activities. The availability of the complete human genome sequence has now uncovered all possible drug targets in this organism (or at least all protein and nucleic acid targets), although for many (perhaps most) of these the biological functions remain to be determined. The question has therefore shifted to choosing from among all potential targets those that are most amenable to chemical intervention and those that are most likely to yield desired physiological consequences in the disease state.

A fundamental difficulty in drug design is that crucial steps of the process occur in a highly reductionist setting; for example, when large chemical libraries are screened for modulators of a biochemically purified enzymatic activity. But because the ultimate setting for the desired action of a drug is the living human milieu, a crucial challenge is to maximize the recovery of appropriate biological activities in the highly complex setting of the organism. There are currently three main approaches that address this problem: use of animal model systems, complex human genetic analysis and Mendelian human genetic analysis. The first involves large-scale mutagenesis in vertebrate model organisms (for example, mouse, fish or mammalian tissue culture cells), followed by the characterization of mutants and the orthologous human genes[6–11]. Such initiatives are complicated by high logistical costs, difficulties of phenotyping and the continuing need to extrapolate the phenotypes to human disease. Targeted mutagenesis of specific genes reduces the logistical costs, and is widely used, but the challenge of interpreting the outcomes with respect to human biology still remains. The second approach, traditionally involving genetic analysis of candidate genes that are selected for biological plausibility, has been generalized by the International HapMap Project, which aims to identify all common genetic variants in human populations. Some of these variants help to identify factors that predispose to common medically and commercially significant diseases, yielding insights into disease mechanisms of potential value in therapeutic development[12]. This model has the potential to provide human biological validation for any given gene of interest for a particular disease condition. Although haplotype mapping has yielded intriguing results, significant theoretical and technical obstacles remain before the fruits of this approach can be fully realized[13].

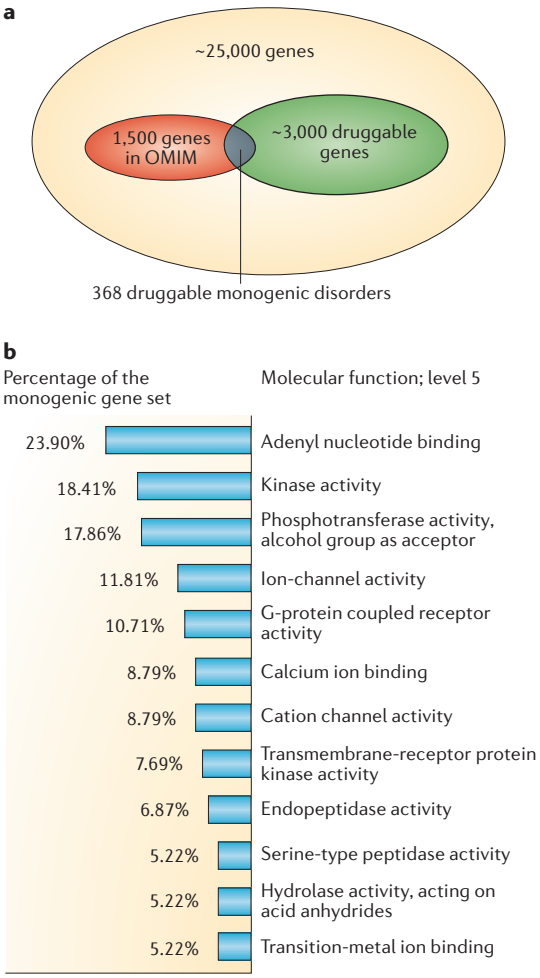
Less attention has been given to studying monogenic (that is, Mendelian) human disorders as a source of drug targets. Monogenic human genetics provides an ideal opportunity for target validation. When strong inferences can be made about the causal effects of a single genetic

Box 1 | **The druggable human genome**

Recent reviews have suggested that there are approximately 3,000 classically 'druggable human genes' in the entire annotated genome[104,105]. These druggable genes were defined on the basis of classes of proteins that are responsive to the small-molecule drugs currently in use, and so have been validated for pharmaceutical development using established biochemical methods in the industry. The list of the druggable human genes includes well-characterized classes of targets, such as G-protein coupled receptors, voltage or ligand gated-ion channels, proteases, kinases and phosphatases. Not all of these genes are necessarily legitimate potential drug targets, as issues of specificity, side-effects and bioavailability could present insuperable obstacles in particular cases. Moreover there are already drugs on the market with targets that do not conform to these 'druggable' rules, such as humanized antibodies, protein therapeutics and peptidomimetics.

The intersection of the two gene sets — those with allelic variants that lead to monogenic phenotypes (FIG. 1) together with those that contain classically druggable functional domains — provides a group of 368 genes that we define as the druggable monogenic genome (panel **a**; see also Supplementary information S1–S3 (figure and tables)). To define and characterize these genes, the non-overlapping complete set of gene entries with curated allelic variants was extracted from the OMIM database and compared with the set used by Orth and co-workers of druggable genes in the human genome[105] to yield a preliminary set of 418 genes[105]. In this study, causal mutations were identified by manual inspection, reducing the set to 368 genes, and functional protein domains were documented using the Gene Ontology (GO) system. The count of these genes will grow as new gene annotations and new monogenic phenotypes continue to be identified.

The 368 genes were queried for GO annotations using Fatigo[106]. The monogenic druggable gene set includes well-understood chemical activities such as adenylate binding (24%), protein kinase (18%), ion channel (12% distributed between anionic and cationic, voltage-gated and ligand-gated ion channels), and G-protein coupled receptors (11%). Overall, 24% of identified genes in monogenic disorders have classically druggable motifs (368/1,500 = 24%); interestingly, this fraction is greater than the fraction of all annotated genes that contain such motifs (2,935/25,000 = 10.5%). Panel **b** shows these results for GO level-5 domains that occur in at least 5% of genes; for a complete listing see Supplementary information S1 (figure). Other GO levels can be queried using the gene list in Supplementary information S3 (table). The percentages for the receptors, kinases and proteases that are quoted in the text are with respect to the 368 genes. These percentages are non-exclusive as the same gene product can be annotated with multiple GO functions.



variant one can ask detailed questions about the consequences of altering the activity of the associated gene product on the biology of the individual. Such information is invaluable to drug developers in evaluating potential novel targets. Because mutations alter the level of activity of gene products, they can be thought of as surrogates for perfectly targeted drugs, the job of which is to antagonize or agonize a given gene product.

Here we review the evidence that monogenic disorders can contribute significantly to the target discovery and validation process. We suggest that rare high-penetrance mutations still comprise the most easily interpretable component of human genetics, that these mutations identify potential rate-limiting steps in disease processes, that monogenic disease states occur with immediate relevance to major therapeutic programmes and that the majority of the human genome remains to be explored using monogenic genetics. We show that among monogenic disorders with understood molecular bases, many of the causal genes are 'druggable' using classical pharmaceutical methodologies (BOX 1). We define the concept of the druggable monogenic genome as the intersection between the set of genes with identified monogenic disorders and the set of genes that contain druggable chemical domains or functions.

The human phenome

The human phenome can be thought of as the set of all phenotypes that are attributable to sequence variation in the human genome[14]. The complete phenome therefore results from many gene variants that are found in the human population around the world. These variants have a wide range of allele frequencies and phenotypic consequences (FIG. 1; see also Supplementary information S2 (table)).

Much of contemporary human genetics is geared towards defining common genomic variation that contributes to the aetiology of common disease states[15]. The HapMap project can be thought of as an attempt to help understand the component of the phenome that depends on relatively common sequence variants, generally with relatively low phenotypic penetrance[12,13,16]. Complementing these efforts, clinical geneticists aim to define the component of the phenome that results from rare and deleterious sequence variants with relatively high penetrance (that is, mutations)[17–19]. The phenotypes

Humanized antibodies
Antibodies in which only the parts of antibody variable regions that mediate the contact to antigens have been grafted onto a human antibody framework by means of genetic engineering techniques.

Peptidomimetics
Compounds that are derived from peptides and proteins by structural modification using, for example, unnatural amino acids, conformational restraints, isosteric replacement and cyclization.

that result from such mutations are often referred to as monogenic or Mendelian. Defined in this way, monogenic variants represent a tiny subset of all the sequence variation that is present in humans worldwide, but they have the most clear genotype–phenotype correlation and functional consequences. This is partly because mutations typically have obvious effects on gene function through the disruption of ORFs or the alteration of highly conserved residues. Moreover, different, independent mutations among unrelated affected individuals can often be identified in the same gene, causing the same or closely related phenotypes.

Using family-based linkage mapping to identify the genes that underlie high-penetrance disorders is reliable and reproducible, facilitated by the availability of the human genome sequence with substantial functional gene annotation[20–22]. This is in contrast to the association of common genetic variants with more common diseases in unrelated individuals, which is less obvious and faces many challenges. Rapidly growing numbers of genetic associations are being published for common diseases, but the necessary replication reports are much slower to follow. Genetic variants that predispose to common diseases such as diabetes[23–26], prostate cancer[27–31], osteoporosis[32], schizophrenia[33,34], stroke[35] and age-related macular degeneration[36–39] have been suggested, although most of them await definitive biological validation.

The monogenic human phenome. The Online Mendelian Inheritance in Man (OMIM) database comprises the most thoroughly curated assemblage of clinically defined phenotypes with associated mutations[40]. As of June 2005, OMIM had 1,748 independent gene entries with described allelic variants (Supplementary information S2 (table)). Of these, about 10% consist of SNPs with suggestive (but not necessarily definitive) phenotypic consequences. Therefore, the 'monogenic genome' currently consists of approximately 1,500 genes, or 6% of the estimated 25,000 protein-coding transcription units in the genome[41]. The associated phenome differs in size because some genes mutate to yield different phenotypes, and in some cases mutations in several genes give rise to the same phenotype.

Monogenic disorders and therapeutics on the market. The phenotypes of rare monogenic disorders might be dismissed as being of little relevance to much of medicine. Most monogenic disorders are extremely rare. Even cystic fibrosis, one of the most common recessive disorders, affects only 0.03% of the Caucasian population and a smaller fraction of non-Caucasian populations. Moreover, many monogenic disorders are either highly pleiotropic (for example, Marfan syndrome, which is caused by mutations in fibrillin 1 (*FBN1*)) or excessively specific (for example, phenylketonuria, which is caused by mutations in phenylalanine hydroxylase (PAH)). By contrast, diseases that are of interest to biotechnology and pharmaceutical companies preferentially include common conditions such as atherosclerosis, obesity, diabetes, asthma, migraine, depression, bipolar disorder, autism, schizophrenia, multiple sclerosis, psoriasis, cancers and pathogenic infections.

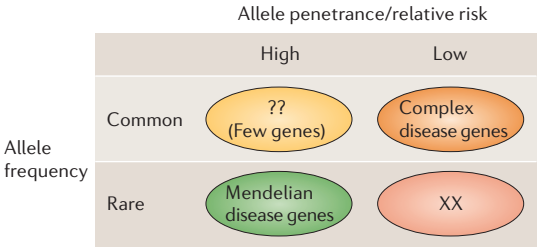


Figure 1 | **Characteristics of the human phenome.** Although genetic variants in the human population occur in an effective continuum of minor allele frequencies (MAF) and penetrance, here we categorize variants as common (MAF > 1%) or rare (MAF < 1%). The severity of the mutation (penetrance) can be expressed as the relative risk (RR) to minor allele carriers and can also be dichotomized as high (RR > 5) as in cystic fibrosis, or low (RR < 5) as for SNPs that are defined through large case–control association studies. With these simplifications, Mendelian or monogenic disorders generally occupy the high-penetrance, rare MAF quadrant. Variants predisposing to complex, potentially oligogenic common diseases generally occupy the low-penetrance, common MAF quadrant. Variants in the low-penetrance, rare MAF quadrant (XX) are intrinsically difficult to ascertain or validate. The high-penetrance, common MAF quadrant probably contains few genes, although founder effects or balanced selection might account for some genes falling into this category (for example, globin and HFE genes, in which certain common mutations lead to sickle-cell anaemia or hereditary haemochromatosis, respectively).

Pharmaceutical development programmes that are directed towards these major therapeutic markets might pay insufficient attention to the diversity of identified monogenic phenotypes (with the exception of drugs targeting rare orphan conditions that explicitly favour monogenic disorders, for example, enzyme replacement therapy with α-galactosidase A for the treatment of Fabry disease). In fact, mutations that cause monogenic disorders have been identified in the genes that encode 12 out of 43 protein targets of the top-selling 100 drugs in 2003 (REF. 42) (TABLES 1,2). Five out of the top ten individual drugs sold in 2004 (Lipitor, Zocor, Procrit, Plavix, Zyprexa) (REF. 43) have monogenic phenotypic correlates.

Close observation of the clinical phenotypes in TABLES 1,2 gene rally provides consistent validation of the mechanisms by which these drugs act on their targets. The sulphonylureas provide an example *par excellence*. These drugs function antagonistically through the sulphonylurea receptor SUR1 complex. Loss-of-function mutations in either of the two genes that encode components of the complex (*ABCA8* and *KCNJ11*) cause the rare genetic disorder persistent hyperinsulinaemic hypoglycaemia of infancy (PHHI)[44]. The phenotype of PHHI, which is hypoglycaemia resulting from increased insulin secretion, is directly mimicked by the action of the sulphonylureas (FIG. 2). Although these

Pleiotropic
The phenomenon in which a single gene is responsible for several distinct and seemingly unrelated phenotypic effects.

Minor allele frequency
The frequency of the less common allele of a polymorphic locus. It has a value that lies between 0 and 0.5, and can vary between populations.

Endometriosis
A common medical condition in which the tissue lining the uterus (the endometrium) is found outside the uterus, typically affecting other organs in the pelvis.

drugs were identified serendipitously years before the molecular basis for PHHI was determined, the similarity between the action of the drugs and the phenotype of the mutations highlights the point that rare, high-penetrance genetic variation in these genes identifies them as points for potential therapeutic intervention in a physiologically relevant disease process.

Similarly, mutations of the low-density lipoprotein (LDL) receptor have long been associated with familial hypercholesterolaemia[45]. Statins function indirectly through this protein by inhibiting endogenous cholesterol biosynthesis at the rate-limiting step that is mediated by HMG CoA (hydroxymethyl glutaryl co-enzyme A) reductase (leading to upregulation of the LDL receptor and increased LDL clearance from plasma)[46]. Oestrogen replacement or agonist therapy (using Premarin and Evista), which is commonly used to treat menopausal complications or to reduce the risk of osteoporosis or breast cancer, is validated by the phenotype of mutations in the oestrogen receptor, which lead to low bone-mineral density, among other clinical symptoms[47]. Drugs that target the gonadotropin-releasing hormone receptor (Lupron, Leuplin and Zoladex) are used to treat prostate cancer and endometriosis, and mutations of this receptor are associated with reduced gonadal development[48]. Inhibition of clotting factor X (using Lovenox) mimics the effect of mutations in this gene, which increase clotting time[49]. Mutations in the P2Y12 receptor, which is targeted by Plavix, also delay clotting[50]. Recently, drugs such as Xenical and Avandia have come on the market, which target PNLIP (pancreatic lipase) and PPARG (peroxisome proliferator-activated receptor-γ) for the treatment of obesity and insulin resistance respectively[51,52]. Mutations in *PNLIP* lead to reduced dietary fat uptake through a mechanism similar to that of the drug, and mutations in PPARG are associated with phenotypic obesity. Rare gain-of-function mutations in the erythropoietin receptor lead to increased mass of red blood cells (erythrocytosis), and this receptor is the target of recombinant erythropoietin hormone (Procrit and Epogen)[53,54]. Cardiac arrythmias are a component of Timothy syndrome, which is caused by mutations in a subunit of the cardiac/smooth muscle L-type calcium channel (CACNA1C), a channel that is targeted by the drugs Norvasc and Adalat for the treatment of arrhythmia[55]. The thiazide class of drugs are diuretics that are used to treat hypertension, and mutations in the thiazide target, solute transporter SLC12A3, cause the pleiotropic phenotypes of Gitelman syndrome, including various effects of metabolic ion misregulation[56]. Finally, prostatic hyperplasia (often called benign, although there are serious non-cancer related disease symptoms) is routinely treated with Proscar, which targets the steroid 5α-reductase 2 enzyme[57]. This gene is mutated in male pseudohermaphroditism, which includes reduced prostate size among other phenotypes.

Overall, the general mechanism of drug action (agonist versus antagonist) is consistent with the consequence of gene mutations (loss-of-function (LOF) versus gain-of-function (GOF)). For example, anti-clotting drugs function as protease inhibitors, and the LOF phenotypes of the targets of these drugs typically involve excessive bleeding. LOF alleles of the sulphonylurea receptor lower glucose levels, and the drugs function as inhibitors of the receptor. LOF alleles of steroid 5α–reductase 2 function

Table 1 | **Monogenic disorders of druggable genes among targets of top-selling drugs**

Drug(s)	Target (agonist/ antagonist)	Disease(s) treated	Genetic model (gene)	OMIM gene ID	Phenotype (effective allele)
Norvasc, Adalat	Cardiac L-type calcium channel α1C (antagonist)	Hypertension, arrythmia	Timothy syndrome (*CACNA1C*)	114205	Cardiac arrhythmia (LOF)
Zyprexa, Risperdal, Seroquel	Dopamine receptors (antagonist)	Schizophrenia, bipolar disorder	Myoclonus dystonia syndrome (*DRD2*)	126450	Neurological and psychiatric disease traits (LOF?)
Premarin, Evista	Oestrogen receptor (agonist)	Menopause, osteoporosis, breast cancer	Oestrogen resistance (*ESR1*)	133430	Early atherosclerosis, low bone mineral density (LOF)
Lovenox	Factor X (antagonist)	Thrombosis	Factor X deficiency (*F10*)	227600	Clotting delayed (LOF)
Flonase, Flovent, Advair, Pulmicort	Glucocorticoid receptor (agonist)	Asthma	Glucocorticoid resistance (*GCCR*)	138040	Severe hypertension (LOF)
Lupron, Leuplin, Zoladex	Gonadotropin releasing hormone receptor (agonist)	Prostate cancer, endometriosis, central precocious puberty	Hypogonadotropic hypogonadism (*GNRHR*)	148850	Reduced gonadal development (LOF?)
Humulin, Humalog	Insulin receptor (agonist)	Diabetes	Insulin-resistant diabetes, leprechaunism (*INSR*)	147670	Defects in growth and glucose metabolism (LOF)
Plavix	Purinergic receptor P2Y (antagonist)	Atherosclerosis, stroke	Congenital bleeding (*P2RY12*)	600515	Clotting delayed (LOF)
Xenical	Lipases (antagonist)	Obesity	Pancreatic lipase deficiency (*PNLIP*)	246600	Reduced dietary fat uptake (LOF)
Avandia	Peroxisome proliferator-activated receptor-γ (agonist)	Insulin resistance syndrome, diabetes	Severe obesity (*PPARG*)	601487	Obesity (LOF)

LOF, loss of function.

similarly to antagonistic drugs that target this enzyme. GOF alleles of the erythropoietin receptor cause erythrocytosis, a phenotype that is positively correlated with the action of the drugs that function as receptor agonists, whereas oestrogen, insulin and PPARG-receptor agonists produce biological effects that are effectively opposite to those of LOF mutations in the receptor. In only a few cases, such as the cardiac L-type calcium channel, the gonadotropin-releasing hormone receptor and dopamine receptors, is it difficult to directly correlate the action of the drug (for example, Norvasc, Lupron and Risperdal, in corresponding order) with the specific phenotypes of the mutant alleles.

Perhaps surprisingly, mutant alleles of genes that encode many other common drug targets have not yet been observed. There could be several reasons for this. First, it is possible that such mutations exist but have not yet been detected through the screening of phenotypically relevant individuals. Second, it is likely that for at least some of these genes, high-penetrance mutant alleles show additional phenotypic consequences that either mask or overcome the medically relevant phenotype. For example, although hundreds of mutations in the LDL receptor are known to cause hypercholesterolaemia, no relevant mutation has been detected in the gene encoding HMG CoA reductase, the primary target of statins. Mutations in this gene, at least when homozygous, probably show other phenotypes that relate to steroid metabolism (although some mutations in HMG CoA reductase could cause hypocholesterolaemia, which might be unobserved in the general population). Finally, as discussed in more detail below, some mutant phenotypes might be 'hidden' among more common environmentally caused or genetically complex disease states (such as leptin mutations that cause rare early-onset obesity).

Monogenic disorders and novel potential therapeutic targets. These examples support the argument that mutations can identify potential targets for important medical conditions. Historically, the development of drugs has usually preceded the human genetic analysis. But this has been changing as a result of methodological advancements and the availability of the human genome sequence. In the future human molecular genetics could lead the way towards more optimal therapeutic targets, rather than following in the wake of drug development. A selection of genetic disorders, most of which have recently been molecularly characterized, offers representative potential drug targets (TABLE 3). Foremost among these is cholesterol ester transfer protein (CETP). Mutations in *CETP* cause elevated levels of high-density lipoprotein (HDL) cholesterol in humans, which in principle should be cardioprotective[58]. This protein is the target of major drug discovery efforts for the treatment of hypoalphalipoproteinaemia (low HDL cholesterol) as a risk factor for heart disease, and small-molecule inhibitors of this protein have been developed that are now under scrutiny[59–61].

Mutations in the ATP-binding cassette transporter A1 (*ABCA1*) cause Tangier disease and low HDL cholesterol[62–64]; this is another potential target for treatment of low HDL[65]. Erythermalgia, a condition of spontaneously induced inflammatory pain, is caused by mutations of the sodium channel *NAV1.7* that lead to hyperexcitability

Table 2 | **Additional selected drugs with Mendelian disease correlates**

Drug(s)	Target (agonist/ antagonist)	Disease/condition	Genetic model (gene)	OMIM gene ID	Phenotype (GOF/LOF allele)
Statins (including Lipitor, Zocor and Mevacor)	HMG CoA-reductase (antagonist)	Hypercholesterolaemia	Familial hypercholesterolaemia (*LDLR*)	606945	High LDL cholesterol (LOF)
Recombinant erythropoietin (Procrit, Epogen)	Erythropoietin receptor (agonist)	Anaemia	Familial erythrocytosis (*EPOR*)	133171	Increased red blood cell mass (GOF)
Sulphonylureas (including Glucotrol, Micronase and Glynase)	Sulphonylurea receptor 1 (antagonist)	Diabetes	PHHI (*ABCC8*, *KCNJ11*); neonatal diabetes	600509, 600937	Severe hypoglycaemia (LOF); diabetes (GOF)
Thiazides (many examples)	Solute transporter 12A3 (antagonist)	Hypertension	Gitelman syndrome (*SLC12A3*)	600968	Pleiotropic disorder including muscle weakness and seizures, which is caused by metabolic ion misregulation (LOF)
Finasteride (Proscar)	Steroid 5α-reductase 2 (antagonist)	Benign prostatic hyperplasia	Male pseudohermaphroditism (*SRD5A2*)	607306	Ambiguous male genitalia, small prostate (LOF)
Imatinib (Gleevec)	Abelson kinase (antagonist)	Chronic myelogenous leukaemia	BCR–ABL (*ABL1*)	189980	Genetic chronic myelogenous leukaemia (GOF)
Abciximab, eptifibatide, tirofiban	Integrins GpIIb/IIIa (antagonist)	Acute cardiovascular care	Glanzmann thrombasthenia (*ITGA2B*, *ITGB3*)	173470, 607759	Pleiotropic disorder related to internal bleeding (LOF)

BCR–ABL, breakpoint cluster region–Abelson murine leukaemia; GOF, gain of function; GpIIb/IIIa, platelet glycoproteins IIb/IIIa; HMG CoA, hydroxymethyl glutaryl co-enzyme A; LDL, low-density lipoprotein; LOF, loss of function; PHHI, persistent hyperinsulinaemic hypoglycaemia of infancy.

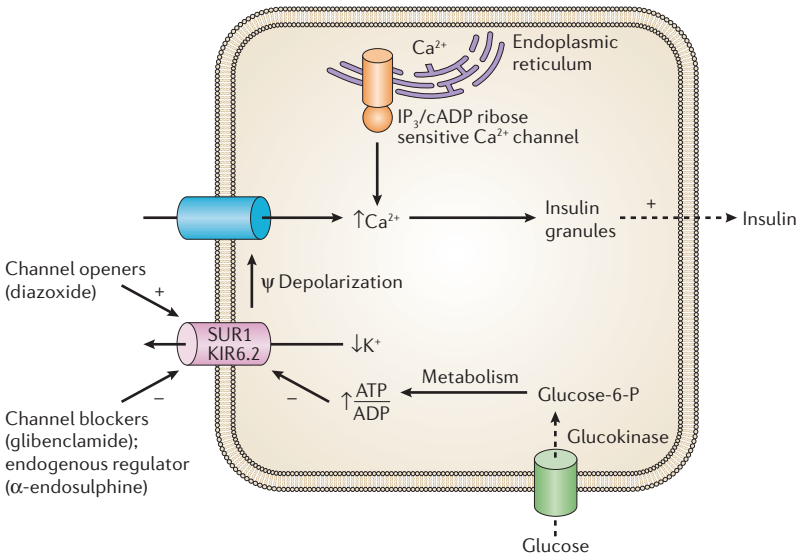


Figure 2 | **Regulation of insulin secretion by the SUR1 receptor complex in pancreatic islet cells.** The SUR1 receptor consists of two polypeptides, a potassium channel that is encoded by the KIR6.2 gene (*KCNJ11*) and an ATP-cassette binding protein that is encoded by the actual SUR1 gene (*ABCC8*). The SUR1 channel maintains a potassium current, generating cell-membrane polarization. Increased system levels of glucose lead to intracellular changes in the ATP:ADP ratio; consequently, the SUR1 receptor channel closes, depolarizing the membrane and triggering calcium influx. This in turn leads to the release of insulin from pre-packaged secretory granules. Insulin functions systemically to stimulate peripheral metabolism of glucose. The SUR1 channel can be modulated pharmacologically by channel blockers (for example, the sulphonylurea drugs) or by openers (for example, diazoxide). Blockers lead to constitutive membrane depolarization and insulin secretion, aiding in the control of diabetic hyperglycaemia as long as the peripheral tissues can still respond normally. Openers reduce insulin secretion. Similarly, both loss- and gain-of-function mutations are known in the SUR1 receptor genes. Loss-of-function mutations in either gene, each of which is usually recessive, results in failure of channel activity and causes hyperinsulinaemia, and ultimately hypoglycaemia — similar to the action of pharmacological channel blocking. Unusual gain-of-function alleles of *KCNJ11*, which are typically dominant, that constitutively cause channel opening lead to the alternative disease state of neonatal diabetes through the inability of the system to secrete insulin normally. Figure courtesy of Kimberly Watson, University of Reading, UK.

of peripheral neurons[66–69], which indicates that this gene could be a target for the treatment of generalized inflammatory pain. Mutations of the frizzled 4 gene (*FZD4*) cause a rare condition called familial exudative vitreoretinopathy (FEVR)[70]. The phenotype of FEVR closely resembles the more common retinopathy of prematurity (ROP), a complication that arises in some prematurely born babies and places them at a severely increased risk for blindness. This genetic discovery implicates the involvement of a frizzled pathway in retinal angiogenesis and suggests a novel approach to the rescue of infants with incipient ROP.

The G-protein coupled receptor (GPCR) superfamily of receptors is of particular interest to drug developers. A substantial fraction of drugs on the market target members of this class of proteins, which has historically proved particularly amenable to the design of both agonists and antagonists that have favourable physiological properties[71]. Apart from a large subgroup of olfactory receptors, which are generally not considered to be likely drug targets, the human genome might have as many as 371 GPCRs (REF. 72). Of these, 35 have associated known Mendelian phenotypes (see Supplementary information S4 (table)). Among them are targets of some marketed drugs as already noted in TABLE 1 (GNRHR for prostate cancer and P2RY12 for clotting), as well as targets that are being investigated in drug development programmes (melanocortin 4 receptor for obesity[73], CC chemokine receptor 5 for HIV resistance[74]) and other potential novel targets (for example, *FZD4* as already noted).

Novel genetic phenotypes yet to be discovered. With ~1,500 current OMIM entries, could monogenic human genetics as a source for novel drug targets be near exhaustion? This could be the case for three reasons: there are no more mutations segregating in the human population; mutations in the remaining genes lead to *in utero* lethality; and mutations of the remaining genes lead to no discernable phenotype.

There are several reasons for believing that these scenarios are unlikely. First, it is known that large numbers of high-penetrance alleles arise regularly in the human genome. Through a meta-analysis of dominant monogenic disorders (excluding conditions that have atypical molecular mechanisms, such as Huntington disease), Kondrashov estimated an overall spontaneous mutation rate of 1.8×10^{-8} per nucleotide per generation[75]. In principle, this value includes deleterious as well as neutral changes. Using Kondrashov's data from table 4 in REF. 75, we calculated a rate for highly deleterious mutations using the per gene per generation rate of mutations that have a major effect, m(major). After eliminating three outlier genes (*IL2RG*, *F8* and *DMD*) and using the remaining 17 gene values, we calculated $m(\text{major})_{\text{mean}} = 8.1 \times 10^{-7}$ per gene per generation. Assuming that there are 25,000 genes in the human genome, this yields a deleterious mutation rate per genome per generation (U) of 0.02. This number is conservative with respect to previously estimated values of U (REFS 76,77), but might be more appropriate for highly penetrant severe mutations. Given a worldwide annual birth rate of 133,000,000 (see the Eco-Economy Indicators web site), and assuming a generation time of 25 years, this value of U translates to 106,400 new mutations per year worldwide, or 8 per 10,000 births. It seems likely therefore that the human genome should be fully saturated with new deleterious alleles on a regular basis worldwide. This estimate is consistent with observations that have been made in founder populations[78–81].

Although such high numbers for mutation load might initially seem surprising, they are supported by reports of multiple mutant alleles in monogenic disorders that routinely follow the initial gene discovery. For example, since the first publication of two mutations in the *FZD4* gene that cause familial exudative vitreoretinopathy[70], more than a dozen independent and different causal mutations have been reported in the literature by other groups[82–87]. For well-ascertained and historically well-studied genetic disorders, such as dominant familial hypercholesterolaemia, many hundreds of different causal mutations have been reported in the LDL receptor

Founder populations
Populations that that have been derived from a limited pool of individuals within the last 100 or fewer generations.

(see the LDLR Locus web site)[88]. Even for a recessive disorder such as cystic fibrosis, over 1,300 different causal mutations in the *CFTR* gene have been documented (see the Cystic Fibrosis Mutation Database), despite the requirement for homozygosity or compound heterozygosity to produce the phenotype, and despite the fact that one particular mutation (delF508) alone accounts for over half the alleles recovered by diagnostic screening worldwide[89]. The genes with identified monogenic variants are not obviously different in structure from the remaining 94% of genes in the genome, so deleterious mutations should regularly arise in all protein-coding genes. Where are the remaining LOF phenotypes?

It is difficult to evaluate the phenotypic potential of mutations in the genes that remain genetically uncharacterized. This is especially true for the fraction of genes that lead to prenatal lethality as a result of either dominant or recessive LOF mutations. Estimates of the numbers of lethal mutations that occur in model organisms such as *D. melanogaster* are difficult to extrapolate to humans given the large evolutionary distances that are involved. In the mouse, one analysis indicated that two-thirds of targeted knockout strains (presumably equivalent to severe LOF alleles) yield homozygous viable and fertile animals[90]. In a collaborative mouse programme to knockout genes that correspond to major drug targets, all but one (insulin, a duplicated gene) were reported to have discernable phenotypes[42]. Only 5 strains had an embryonic lethal phenotype. A more comprehensive study by the same group reported a 19% embryonic lethality rate for a large collection of knockout strains[91].

In light of these results, it seems very unlikely that prenatal lethality represents the LOF phenotype of the entire genetically uncharacterized repertoire of human genes. Even LOF mutations in basic pattern formation genes may be viable when in a homozygous (for example, *HOXA1* in the case of Bosley–Salih–Alorainy syndrome) or heterozygous state (for example, *HOXA13* in the case of hand–foot–genital syndrome). Data on clinically documented human pregnancy losses have been interpreted to suggest that as many as two-thirds of such incidents result from cytologically observable chromosomal abnormalities, especially in the first trimester[92]. Many of the remainder probably result from immunological or other general mother–child incompatibilities. It is unclear what fraction of spontaneous miscarriages result from single gene mutations (either dominant or recessive), and this is an important area for further research.

It is unlikely that mutations in most genes yield no detectable phenotype, particularly in the homozygous or compound heterozygous LOF state. Among 317 mouse knockouts that are summarized in the Frontiers in Bioscience Database of Gene Knockouts, only 13 are claimed to show no observable phenotype.

Importantly, there is little evidence that the pace of discovery of genes that underlie monogenic disorders in humans has abated, or that the number of such genes has reached a plateau. A previous analysis of OMIM that we carried out in July 2003 identified 1,100 genes with allelic variants, indicating an increase of curated entries of 50% in the ensuing 2 years. Similarly, interrogation of PubMed indicates that reports of monogenic discoveries continue to be published frequently in high-impact journals.

In addition to phenotypes that have identified causal allelic variants, the OMIM database also curates reported putative monogenic phenotypes that do not have a known molecular basis. OMIM contained 1,498 such entries in February 2006. These range from entries that have linkage mapping data to specific chromosomal loci (for example, OMIM number 609306), to phenotypes that have weaker mapping data (for example, OMIM number 609261), to phenotypes that have no clear mapping data but for which there is a tentative exclusion of known genes (for example, OMIM number 600361). These entries support the existence of a large pool of novel monogenic phenotypes that could be explored through molecular genetic analysis.

Non-classical drug targets. Although we have focused on the intersection of the monogenic and classically druggable genomes, new approaches promise that a larger fraction of the genome will ultimately be accessible for therapeutic intervention. The RGD tripeptide motif of platelet integrin ITGB3/ITGA2B has been targeted with a customized antibody (Abciximab) for use in acute coronary care[93]. ERBB2 has been targeted with Herceptin in some breast cancers[94,95]. Peptide (eptifibatide) and non-peptidic (tirofiban) agents are also on the market against ITGB3/ITGA2B, and similar approaches are now being tested to target the integrin ITGAV/ITGB3 in certain cancer conditions[96].

Protein–protein interactions have long been considered undruggable with small organic molecules, as these interactions typically involve extensive surfaces of protein contact rather than localized molecular binding pockets. Recently some of these types of interaction have begun to yield to informed chemical attack[96–99]. In addition to the work on integrin RGD motifs, chemicals are under development that target other protein–protein interactions that have been genetically validated, such as interleukin 2 (IL2) and its receptors IL2RA and IL2RG, erythropoietin and its receptor EPOR, growth hormone and its receptor GHR, HIV and the CC chemokine receptor 5 (CCR5), and transthyretin aggregation. In the future the druggable target universe will hopefully expand to include many important genetically defined phenotypes that are underpinned by genes with products that engage in protein–protein interactions.

Anti-disease genetics. Most mutations cause LOF rather than GOF. For this reason 'anti-disease' states should be especially useful to identify points of therapeutic intervention, because pharmaceutical antagonists are also more generally tractable than agonists. Such anti-disease conditions include low blood sugar as the anti-disease of diabetes, high bone-mineral density as the anti-disease of osteoporosis, and leanness as the anti-disease of obesity. Mutations that cause anti-disease states potentially provide high-grade therapeutic targets if the genes fall in a chemically druggable class. The sulphonylureas discussed above provide an example of this reasoning,

Compound heterozygosity
A situation in which an inidividual is heterozygous for two different mutations at the same locus.

Proband
A subject that is ascertained on the basis of phenotype; they are often used to identify affected families for genetic studies.

as does *CETP* with the associated anti-disease state hyperalphalipoproteinaemia. The hypocholesterolaemia phenotype of LOF mutations in *PCSK9* (proprotein convertase subtilisin/kexin type 9) (TABLE 3) indicates that this gene might be a druggable anti-disease target for hypercholesterolaemia. Ascertainment of anti-disease states is challenging, except in cases where the anti-disease phenotype is sufficiently severe as to create a novel genetic disorder itself (as with mutations of the sulphonylurea receptor and *PHHI*).

Mining the untapped phenome

Given that high-penetrance, predominantly LOF alleles remain to be identified for most of the human gene repertoire, how might the associated phenotypes and, in the current context, the most medically significant phenotypes be ascertained?

As noted previously, the OMIM database already includes more than 1,000 familial disorders of potential genetic aetiology, for which the molecular basis remains undetermined. This set of disorders could provide an immediate pool of material for analysis, if the challenges of re-ascertainment, patient consent and sample collection can be met. It is likely that many of these disorders have remained uncharacterized owing to small family sizes with insufficient statistical power for whole-genome mapping. New methods, such as the use of dense SNP panels, and reduced costs of gene re-sequencing can mitigate this restriction to some extent. Review of the clinical genetics literature also strongly supports the contention that the monogenic phenome is nowhere near completion, and that new phenotypic descriptions are presented routinely in case reports, many of which are likely to result from single gene mutations.

Many characterized monogenic disorders show substantial genetic heterogeneity. For example, although there are currently nine molecularly identified genes that mutate to cause Bardet–Biedl syndrome, mutations in these genes fail to account for all ascertained cases of the disorder[100]. It is certain that some fraction of mutations in previously uncharacterized genes will yield phenotypes that are related to those that are already characterized.

In the case of anti-disease states, it seems likely that many monogenic disorders remain to be identified. In the examples cited above, the anti-disease state in fact generates a disease state itself (such as PHHI versus diabetes). It is difficult, although not impossible, to develop a clinical ascertainment scheme for anti-disease states that will lead to improved health rather than disease. Efforts in this direction might be successful through random screening of biologically relevant candidate genes (for example, the observation of LOF alleles of *PCSK9* in healthy hypocholesterolaemic individuals[44]). However this process is cumbersome and has not been extensively pursued, and monogenic phenotypes remain to be ascertained in this way.

The most efficient approach to ascertaining novel phenotypes probably lies in extensive clinical genetic studies among founder populations[101]. Such populations offer the advantages of reduced genetic heterogeneity, increased rates of homozygosity and often increased family sizes (although in developed countries family sizes are declining even among rural population isolates). Genotyping and positional cloning methods have significantly simplified mapping in these populations, and have lowered the costs of gene discovery (FIG. 3). Ultimately, if sequencing costs can be reduced sufficiently, exon-focused, whole-genome mutation detection might become cost-effective for screening individual probands with clinical phenotypes of interest and some evidence of family history[102]. Historically, the major challenge to this approach has been the unpredictable rate of ascertainment of novel phenotypes. The key to overcoming this challenge is close collaboration between molecular geneticists, clinical geneticists and clinicians who work in the various medical disciplines. Very small patient clusters with interesting rare phenotypes might not automatically be recognized as having a potential genetic aetiology unless the primary and specialist caregivers are sensitized to this possibility and are motivated to participate in the ensuing research. Our observation is that such caregivers, functioning in a clinical context, might not realize the long-term medical benefits that can accrue. By contrast, patients and families who

Table 3 | **Potential novel drug targets with Mendelian disease correlates**

Targets	Disease/condition (agonist/antagonist)	Genetic model (gene)	OMIM gene ID	Phenotype (GOF/LOF)
Cholesterol ester transfer protein	Low HDL cholesterol (antagonist)	Hyperalphalipoproteinaemia (*CETP*)	118470	High HDL cholesterol (LOF)
ATP-binding cassette transporter A1	Low HDL cholesterol (agonist)	Tangier disease (*ABCA1*)	600046	Low HDL cholesterol (LOF)
Proprotein convertase, subtilisin/kexin-type 9	High LDL cholesterol (antagonist)	Hypo- and hypercholesterolaemia (*PCSK9*)	607786	High/low LDL cholesterol (GOF/LOF respectively)
Sodium channel 9A	Inflammatory pain (antagonist)	Erythermalgia (*SCN9A*; also known as *NAV1.7*)	603415	Pain and inflammation of peripheral skin surfaces (GOF)
Frizzled 4, lipoprotein receptor related protein 5, Norrie disease protein	Retinopathy of prematurity (agonist)	Familial exudative vitreoretinopathy (*FZD4*, *LRP5*, *NDP*)	604579, 603506, 310600	Incomplete retinal angiogenesis, secondary neovascularization, retinal detachment, blindness (LOF)

GOF, gain of function; HDL, high-density lipoprotein; LDL, low-density lipoprotein; LOF, loss of function.

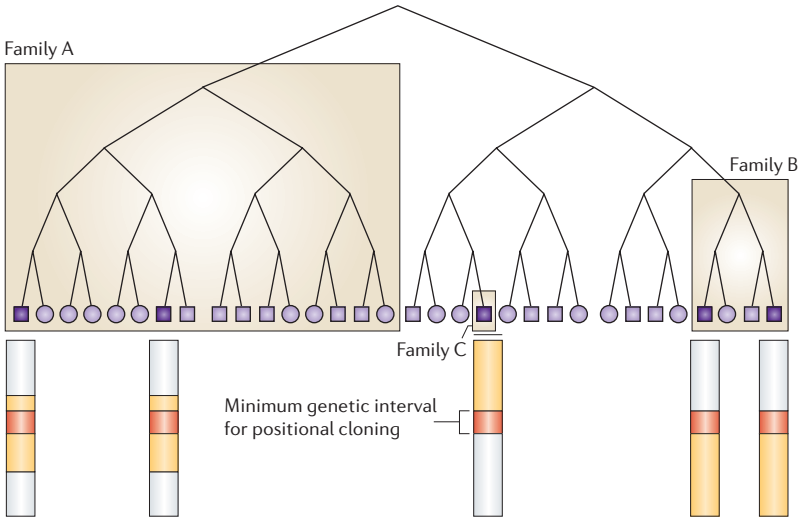


Figure 3 | **Mapping rare genetic disorders in founder or isolated populations.** Shared chromosomal segments (haplotypes, shown in yellow) that surround causal sequence variants might descend from a common ancestor to clinically ascertained families and probands that are unrelated by a documented genealogy (such as families A, B and C). If common ancestry for a mutation (either homozygous or heterozygous) is more likely than introductions of different causal alleles into such families, the overlapping shared haplotype segments can be identified in these different families or probands within the population, allowing the definition of small intervals for mutation detection (shown in red). Minimum intervals might be on the order of one to two-million base pairs that contain 10–20 genes.

suffer from severe monogenic disorders are frequently highly motivated to assist in research that clarifies the nature of their disease. Increased awareness of the value of genetics should be inculcated among medical practitioners and the general public.

Among novel monogenic phenotypes there might be rare forms of common diseases that represent significant unmet medical needs. It seems improbable that large, highly penetrant, dominant pedigrees have been regularly missed for important medical conditions, although dominant alleles can sometimes be masked; for example, in the case of incomplete penetrance of mutations in the melanocortin 4 receptor (MC4R) that causes obesity. Nevertheless, recessive transmission seems likely to predominate for common phenotypes, as has been documented in monogenic forms of obesity, hypoglycaemia and hypertension. Further examples of rare recessive monogenic disorders, which mimic common disease states, could easily be hidden among the more common complex oligogenic or environmentally caused forms of important diseases. It would be easiest, and perhaps only possible, to ascertain recessive monogenic forms of otherwise common diseases through the study of founder populations or population isolates.

Conclusions

The sequencing of the human genome and its continuing annotation provide an unprecedented opportunity to investigate the biological consequences of natural sequence variation. The subset of the complete human phenome that corresponds to monogenic or Mendelian disorders is incompletely defined, and study of these disorders is highly consistent with novel drug development for important unmet medical needs.

Certain classes of disease might not have monogenic equivalents. Surprisingly, however, monogenic forms exist even for behavioural phenotypes, which are particularly difficult to replicate *ab initio* in animal models. Neuropsychiatric conditions such as Alzheimer disease and Parkinson disease have rare Mendelian versions that have provided important clues about the underlying disease pathways that function in the general population, but are not necessarily direct drug targets themselves (for example, presenilin 1 and presenilin 2, amyloid-β A4 precursor, α-synuclein, parkin, ubiquitin carboxyl-terminase esterase L1, leucine-rich repeat kinase 2, oncogene DJ1, PTEN-induced kinase, and microtubule-associated protein tau). Monogenic disorders occur for even more subtle behavioural traits, including circadian rhythm and language disorders.

One potential problem with making inferences from human genetics about drug function stems from the fact that germline mutations function throughout development, and might not be predictive of the effects of targeting the corresponding gene product with a drug later in life. This is an important concern, particularly for genes that are involved in growth, pattern formation or organogenesis. However, for genes that regulate or participate in ongoing metabolic processes, such as energy metabolism, blood-pressure homeostasis and neuronal transmission, changes in gene activity might have similar dynamic consequences at various stages of life, even if system-wide phenotypes become more severe with increasing age. This concern seems equally as relevant for interpreting low-penetrance (complex) as for interpreting high-penetrance (Mendelian) genetic variants in the validation of potential drug targets. Following human genetic studies with conditional knockout or knockin experiments in mice can potentially address this problem, as can the observation of milder phenotypes that arise from hypomorphic or heterozygous alleles (for example, some maturity onset diabetes of the young (MODY) genes that are homozygous lethal as mouse knockouts yield heterozygous mutants with diabetes).

It is often assumed that genetic variation in a drug target is a necessary prerequisite for the successful action of the drug if the target has been defined through genetic analysis; the use of Gleevec in the treatment of patients with genetic chronic myelogenous leukaemia is one example[94]. However, in general, this expectation reflects a misunderstanding of the potential role of genetics in defining rate-limiting steps in pathways. The direct target of the statin drugs is the enzyme HMG CoA-reductase, in which inhibition leads to upregulation of the LDL receptor and increased plasma LDL clearance. However genetic variation in these two genes has not traditionally had a role in defining appropriate patient populations for drug therapy, despite the fact that the genetics of familial hypercholesterolaemia and the LDL receptor were important in validating this pathway for therapeutic intervention. The value of monogenic disorders in target discovery might be underappreciated if the

rarity of a genetic condition that is used to find a drug target is taken to imply a correspondingly small patient population for the therapeutic agent. The size of a potential therapeutic market has no obligate relationship to the size of the genetically affected population for a single gene disorder; it is the relevance of the rare phenotype to the more common disease state that determines this relationship.

Failure of novel therapeutics during the development process can arise for numerous reasons. Toxicity, a major cause of the failure of new drugs, might result from unanticipated consequences of modulating the desired drug target, from unanticipated modulation of other gene products that are not the desired target, or from such activities of the drug metabolites. To some extent, once the target has been chosen, the problem of minimizing toxicity is primarily one of medicinal chemistry. Nonetheless, the more appropriate the selected target is for the disease state, the less likely it is that toxicity issues will arise from the fundamental mechanism of drug action. For this reason, genetically defined targets are potentially superior to those that are selected from a more general study of the disease physiology, as side effects that are intrinsic to the target itself should be better appreciated in advance from phenotypic analysis of the associated monogenic disorder.

Having said that, the role of genetics in stratifying the patient pool for particular drugs, a field that has variously been known as pharmacogenetics or pharmacogenomics, is likely to become increasingly relevant in the future[103]. Even in this complex area there are examples of monogenic, high-penetrance genetic conditions, including the anaesthetic-induced apnoea that is caused by butyrylcholinesterase mutations, and the variable drug metabolism (with sometimes severe medical consequences) that is related to various mutations in cytochrome P450 2D6 (*CYP2D6*) and thiopurine S-methyltransferase (*TPMT*). Indeed, even some well-described life-threatening metabolic disorders such as phenylketonuria, favism and fructose intolerance can be thought of as single-gene pharmacogenetic conditions, in which a basic dietary component rather than a defined 'drug' provides the chemical risk factor.

Genetically defined targets have the potential to expand the pool of druggable targets under pharmaceutical development, or more accurately, to winnow the pool of all druggable genes to identify those with predictable outcomes when targeted. The current enthusiasm for combination therapies among pharmaceutical companies, although productive, probably results less from a paucity of novel single agent targets than from a plethora of such targets that have insufficient physiological data to choose between them or to justify each through hundred-million-dollar investments in new programmes.

A comprehensive description of the naturally occurring monogenic human phenome would in effect be equivalent to a saturation mutagenesis, defining the physiological consequences of gene knockouts, hypomorphs and also rarer GOF mutations in all genes. This endeavour is certain to provide many more surprises for students of human physiology and for scientists using genomics for the direct improvement of human health.

1. Frantz, S. FDA publishes analysis of the pipeline problem. *Nature Rev. Drug Discov.* **3**, 379 (2004).
2. Sams-Dodd, F. Target-based drug discovery: is something wrong? *Drug Discov. Today* **10**, 139–147 (2005).
The most recent of several reviews that discuss the challenges facing drug development pipelines and propose various genomic approaches to improve target selection and validation.
3. Butcher, S. P. Target discovery and validation in the post-genomic era. *Neurochem. Res.* **28**, 367–371 (2003).
4. Hardy, L. W. & Peet, N. P. The multiple orthogonal tools approach to define molecular causation in the validation of druggable targets. *Drug Discov. Today* **9**, 117–126 (2004).
5. Lindsay, M. A. Target discovery. *Nature Rev. Drug Discov.* **2**, 831–838 (2003).
6. Wienholds, E. *et al.* Efficient target-selected mutagenesis in zebrafish. *Genome Res.* **13**, 2700–2707 (2003).
7. 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.* **4**, 189–202 (1994).
8. Quwailid, M. M. *et al.* A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm. Genome.* **15**, 585–591 (2004).
9. Clark, A. T. *et al.* Implementing large-scale ENU mutagenesis screens in North America. *Genetica* **122**, 51–64 (2004).
10. Rastan, S. *et al.* Towards a mutant map of the mouse — new models of neurological, behavioural, deafness, bone, renal and blood disorders. *Genetica* **122**, 47–49 (2004).
11. Zambrowicz, B. P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
12. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
13. Cardon, L. R. & Abecasis, G. R. Using haplotype blocks to map human complex trait loci. *Trends Genet.* **19**, 135–140 (2003).
A thoughtful discussion of the methodological basis for the HapMap project, its potential, technical challenges and practical guidelines for experimental analysis of complex human traits including common diseases.
14. Freimer, N. & Sabatti, C. The human phenome project. *Nature Genet.* **34**, 15–21 (2003).
15. Allen, M. J. & Carrey, A. H. Target identification and validation through genetics. *Drug Discov. Today Targets* **3**, 183–190 (2004).
16. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Rev. Genet.* **6**, 95–108 (2005).
17. Leroi, A. M. *Mutants* (Viking Penguin Books, New York, 2003).
18. Brunner, H. G. & van Driel, M. A. From syndrome families to functional genomics. *Nature Rev. Genet.* **5**, 545–551 (2004).
An informative discussion of rare monogenic human disorders and their utility in dissecting basic physiological mechanisms of human development and disease.
19. Donnai, D. & Read, A. P. How clinicians add to knowledge of development. *Lancet* **362**, 477–484 (2003).
20. Samuels, M. E. & Dube, M. P. *Encyclopedia of Genetics, Genomics, Proteomics, and Bioinformatics* Genetics Vol. (John Wiley & Sons, 2005).
21. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nature Genet.* **33**, S228–S237 (2003).
An authoritative presentation of methods in human genetic discovery by key exponents in the field, with an historical discussion of past, present and future approaches in monogenic and complex trait analysis
22. Craig, D. W. & Stephan, D. A. Applications of whole-genome high-density SNP genotyping. *Expert Rev. Mol. Diagn.* **5**, 159–170 (2005).
23. Horikawa, Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genet.* **26**, 163–175 (2000).
24. Song, Y., Niu, T., Manson, J. E., Kwiatkowski, D. J. & Liu, S. Are variants in the *CAPN10* gene related to risk of type 2 diabetes? A quantitative assessment of population and family-based association studies. *Am. J. Hum. Genet.* **74**, 208–222 (2004).
25. Weedon, M. N. *et al.* Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility. *Am. J. Hum. Genet.* **73**, 1208–1212 (2003).
26. Fingerlin, T. E. *et al.* Variation in three single nucleotide polymorphisms in the calpain-10 gene not associated with type 2 diabetes in a large Finnish cohort. *Diabetes* **51**, 1644–1648 (2002).
27. Rokman, A. *et al.* Germline alterations of the *RNASEL* gene, a candidate *HPC1* gene at 1q25, in patients and families with prostate cancer. *Am. J. Hum. Genet.* **70**, 1299–1304 (2002).
28. Carpten, J. *et al.* Germline mutations in the ribonuclease L gene in families showing linkage with *HPC1*. *Nature Genet.* **30**, 181–184 (2002).
29. Tavtigian, S. V. *et al.* A candidate prostate cancer susceptibility gene at chromosome 17p. *Nature Genet.* **27**, 172–180 (2001).
30. Wang, L. *et al.* Role of *HPC2/ELAC2* in hereditary prostate cancer. *Cancer Res.* **61**, 6494–6499 (2001).
31. Camp, N. J. & Tavtigian, S. V. Meta-analysis of associations of the Ser217Leu and Ala541Thr variants in *ELAC2* (*HPC2*) and prostate cancer. *Am. J. Hum. Genet.* **71**, 1475–1478 (2002).

32. Styrkarsdottir, U. *et al.* Linkage of osteoporosis to chromosome 20p12 and association to *BMP2*. *PLoS Biol.* **1**, e69 (2003).
33. Stefansson, H. *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am. J. Hum. Genet.* **71**, 877–892 (2002).
34. Tosato, S., Dazzan, P. & Collier, D. Association between the neuregulin 1 gene and schizophrenia: a systematic review. *Schizophr. Bull.* **31**, 613–617 (2005).
35. Gretarsdottir, S. *et al.* The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nature Genet.* **35**, 131–138 (2003).
36. Hageman, G. S. *et al.* A common haplotype in the complement regulatory gene factor H (*HF1/CFH*) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* **102**, 7227–7232 (2005).
One of several reports of the most consistent mapping of a complex genetic trait through high-density SNP-based linkage disequilibrium analysis (see following references for replication and validation of the results).
37. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
38. Edwards, A. O. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421–424 (2005).
39. Haines, J. L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 (2005).
40. McKusick, V. A. *Online Mendelian Inheritance in Man, OMIM* (McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University and National Center for Biotechnology Information, National Library of Medicine, Baltimore and Bethesda, Maryland, 2000).
41. Southan, C. Has the yo-yo stopped? An assessment of human protein-coding gene number. *Proteomics* **4**, 1712–1726 (2004).
42. Zambrowicz, B. P. & Sands, A. T. Knockouts model the 100 best-selling drugs — will they model the next 100? *Nature Rev. Drug Discov.* **2**, 38–51 (2003).
Phenotypic description of engineered mouse knockouts in gene targets of 100 drugs that are available on the market, documenting the relevance of model-system genetics for human pharmaceutical development.
43. IMS Health. 2004 year-end U.S. prescription and sales information and commentary. *IMS Press Room web site* [online], < http://www.imshealth.com/ims/portal/front/articleC/0,2777,6652_3665_69890098,00.html > (2005).
44. Cohen, M. M. Jr. Persistent hyperinsulinemic hypoglycemia of infancy. *Am. J. Med. Genet. A* **122**, 351–353 (2003).
45. Austin, M. A., Hutter, C. M., Zimmern, R. L. & Humphries, S. E. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am. J. Epidemiol.* **160**, 407–420 (2004).
46. Reihner, E. *et al.* Influence of pravastatin, a specific inhibitor of HMG-CoA reductase, on hepatic metabolism of cholesterol. *N. Engl. J. Med.* **323**, 224–228 (1990).
47. Smith, E. P. *et al.* Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N. Engl. J. Med.* **331**, 1056–1061 (1994).
48. Yen, S. S. Clinical applications of gonadotropin-releasing hormone and gonadotropin-releasing hormone analogs. *Fertil. Steril.* **39**, 257–266 (1983).
49. Spencer, F. A. *et al.* Enoxaparin, a low molecular weight heparin, inhibits platelet-dependent prothrombinase assembly and activity by factor-Xa neutralization. *J. Thromb. Thrombolysis* **9**, 223–228 (2000).
50. Storey, R. F. The P2Y12 receptor as a therapeutic target in cardiovascular disease. *Platelets* **12**, 197–209 (2001).
51. Martens, F. M., Visseren, F. L., Lemay, J., de Koning, E. J. & Rabelink, T. J. Metabolic and additional vascular effects of thiazolidinediones. *Drugs* **62**, 1463–1480 (2002).
52. Hauptman, J. B., Jeunet, F. S. & Hartmann, D. Initial studies in humans with the novel gastrointestinal lipase inhibitor Ro 18–0647 (tetrahydrolipstatin). *Am. J. Clin. Nutr.* **55**, S309–S313 (1992).
53. Glaspy, J. The impact of epoetin alfa on quality of life during cancer chemotherapy: a fresh look at an old problem. *Semin. Hematol.* **34**, 20–26 (1997).
54. Mohini, R. Clinical efficacy of recombinant human erythropoietin in hemodialysis patients. *Semin. Nephrol.* **9**, 16–21 (1989).
55. Kass, R. S., Arena, J. P. & Chin, S. Cellular electrophysiology of amlodipine: probing the cardiac L-type calcium channel. *Am. J. Cardiol.* **64**, 35I–41I; discussion 41I–42I (1989).
56. Simon, D. B. *et al.* Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na–Cl cotransporter. *Nature Genet.* **12**, 24–30 (1996).
57. The MK-906 (Finasteride) Study Group. One-year experience in the treatment of benign prostatic hyperplasia with finasteride. *J. Androl.* **12**, 372–375 (1991).
58. de Grooth, G. J. *et al.* A review of CETP and its relation to atherosclerosis. *J. Lipid Res.* **45**, 1967–1974 (2004).
59. Forrester, J. S., Makkar, R. & Shah, P. K. Increasing high-density lipoprotein cholesterol in dyslipidemia by cholesteryl ester transfer protein inhibition: an update for clinicians. *Circulation* **111**, 1847–1854 (2005).
60. Hirano, K., Yamashita, S. & Matsuzawa, Y. Pros and cons of inhibiting cholesteryl ester transfer protein. *Curr. Opin. Lipidol.* **11**, 589–596 (2000).
61. Brousseau, M. E. *et al.* Effects of an inhibitor of cholesteryl ester transfer protein on HDL cholesterol. *N. Engl. J. Med.* **350**, 1505–1515 (2004).
62. Brooks-Wilson, A. *et al.* Mutations in *ABC1* in Tangier disease and familial high-density lipoprotein deficiency. *Nature Genet.* **22**, 336–345 (1999).
63. Rust, S. *et al.* Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. *Nature Genet.* **22**, 352–355 (1999).
64. Lawn, R. M. *et al.* The Tangier disease gene product ABC1 controls the cellular apolipoprotein-mediated lipid removal pathway. *J. Clin. Invest.* **104**, R25–R31 (1999).
65. Singaraja, R. R. *et al.* Increased *ABCA1* activity protects against atherosclerosis. *J. Clin. Invest.* **110**, 35–42 (2002).
66. Nassar, M. A. *et al.* Nociceptor-specific gene deletion reveals a major role for $Na_v1.7$ (PN1) in acute and inflammatory pain. *Proc. Natl Acad. Sci. USA* **101**, 12706–12711 (2004).
67. Yang, Y. *et al.* Mutations in *SCN9A*, encoding a sodium channel α-subunit, in patients with primary erythermalgia. *J. Med. Genet.* **41**, 171–174 (2004).
68. Dib-Hajj, S. D. *et al.* Gain-of-function mutation in $Na_v1.7$ in familial erythromelalgia induces bursting of sensory neurons. *Brain* **128**, 1847–1854 (2005).
69. Drenth, J. P. *et al. SCN9A* mutations define primary erythermalgia as a neuropathic disorder of voltage gated sodium channels. *J. Invest. Dermatol.* **124**, 1333–1338 (2005).
70. Robitaille, J. *et al.* Mutant frizzled-4 disrupts retinal angiogenesis in familial exudative vitreoretinopathy. *Nature Genet.* **32**, 326–330 (2002).
71. Robas, N., O'Reilly, M., Katugampola, S. & Fidock, M. Maximizing serendipity: strategies for identifying ligands for orphan G-protein-coupled receptors. *Curr. Opin. Pharmacol.* **3**, 121–126 (2003).
72. Fredriksson, R. & Schioth, H. B. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.* **67**, 1414–1425 (2005).
73. Irani, B. G. *et al.* Progress in the development of melanocortin receptor selective ligands. *Curr. Pharm. Des.* **10**, 3443–3479 (2004).
74. Kazmierski, W. *et al.* Recent progress in discovery of small-molecule CCR5 chemokine receptor ligands as HIV-1 inhibitors. *Bioorg. Med. Chem.* **11**, 2663–2676 (2003).
75. Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* **21**, 12–27 (2003).
A meta-analysis of spontaneously arising monogenic human disease states, leading to the most clinically relevant measure of disease-causing mutation rates.
76. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
77. Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
78. 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.* **92**, 343–345 (2000).
79. de la Chapelle, A. & Wright, F. A. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc. Natl Acad. Sci. USA* **95**, 12416–12423 (1998).
Discussion of methodological approaches to human genetic analysis in founder populations, which focused on Finland as one of the best-studied examples.
80. Scriver, C. R. Human genetics: lessons from Quebec populations. *Annu. Rev. Genomics Hum. Genet.* **2**, 69–101 (2001).
81. Rahman, P. *et al.* The Newfoundland population: a unique resource for genetic investigation of complex diseases. *Hum. Mol. Genet.* **12**, R167–R172 (2003).
82. Kondo, H., Hayashi, H., Oshima, K., Tahira, T. & Hayashi, K. Frizzled 4 gene (*FZD4*) mutations in patients with familial exudative vitreoretinopathy with variable expressivity. *Br. J. Ophthalmol.* **87**, 1291–1295 (2003).
83. Omoto, S., Hayashi, T., Kitahara, K., Takeuchi, T. & Ueoka, Y. Autosomal dominant familial exudative vitreoretinopathy in two Japanese families with *FZD4* mutations (H69Y and C181R). *Ophthalmic Genet.* **25**, 81–90 (2004).
84. Toomes, C. *et al.* Mutations in *LRP5* or *FZD4* underlie the common familial exudative vitreoretinopathy locus on chromosome 11q. *Am. J. Hum. Genet.* **74**, 721–730 (2004).
85. Toomes, C. *et al.* Spectrum and frequency of *FZD4* mutations in familial exudative vitreoretinopathy. *Invest. Ophthalmol. Vis. Sci.* **45**, 2083–2090 (2004).
86. Yoshida, S. *et al.* Novel mutation in *FZD4* gene in a Japanese pedigree with familial exudative vitreoretinopathy. *Am. J. Ophthalmol.* **138**, 670–671 (2004).
87. Qin, M. *et al.* Complexity of the genotype–phenotype correlation in familial exudative vitreoretinopathy with mutations in the *LRP5* and/or *FZD4* genes. *Hum. Mutat.* **26**, 104–112 (2005).
88. Wilson, D. J. *et al.* A World Wide Web site for low-density lipoprotein receptor gene mutations in familial hypercholesterolemia: sequence-based, tabular, and direct submission data handling. *Am. J. Cardiol.* **81**, 1509–1511 (1998).
89. Bobadilla, J. L., Macek, M. Jr, Fine, J. P. & Farrell, P. M. Cystic fibrosis: a worldwide analysis of *CFTR* mutations — correlation with incidence data and application to screening. *Hum. Mutat.* **19**, 575–606 (2002).
90. Hurst, L. D. & Smith, N. G. Do essential genes evolve slowly? *Curr. Biol.* **9**, 747–750 (1999).
91. Zambrowicz, B. & Sands, A. T. Modelling drug action in the mouse with knockouts and RNA interference. *Drug Discov. Today* **3**, 198–207 (2004).
92. Gardner, R. M. & Sutherland, G. R. *Chromosome Abnormalities and Genetic Counselling* (Oxford Univ. Press, New York, 2004).
93. Madan, M., Berkowitz, S. D. & Tcheng, J. E. Glycoprotein IIb/IIIa integrin blockade. *Circulation* **98**, 2629–2635 (1998).
94. Workman, P. New drug targets for genomic cancer therapy: successes, limitations, opportunities and future challenges. *Curr. Cancer Drug Targets* **1**, 33–47 (2001).
95. Emens, L. A. Trastuzumab: targeted therapy for the management of HER-2/neu-overexpressing metastatic breast cancer. *Am. J. Ther.* **12**, 243–253 (2005).
96. Berg, T. Modulation of protein-protein interactions with small organic molecules. *Angew Chem. Int. Ed. Engl.* **42**, 2462–2481 (2003).
97. Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nature Rev. Drug Discov.* **3**, 301–317 (2004).
98. Pagliaro, L. *et al.* Emerging classes of protein–protein interaction inhibitors and new tools for their development. *Curr. Opin. Chem. Biol.* **8**, 442–449 (2004).
99. Zhao, L. & Chmielewski, J. Inhibiting protein–protein interactions using designed molecules. *Curr. Opin. Struct. Biol.* **15**, 31–34 (2005).
100. Nishimura, D. Y. *et al.* Comparative genomics and gene expression analysis identifies *BBS9*, a new Bardet–Biedl syndrome gene. *Am. J. Hum. Genet.* **77**, 1021–1033 (2005).
101. Peltonen, L., Palotie, A. & Lange, K. Use of population isolates for mapping complex traits. *Nature Rev. Genet.* **1**, 182–190 (2000).
102. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
103. Goldstein, D. B., Tate, S. K. & Sisodiya, S. M. Pharmacogenetics goes genomic. *Nature Rev. Genet.* **4**, 937–947 (2003).

104. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature Rev. Drug Discov.* **1**, 727–730 (2002).
The most-often cited reference for the component of the human genome that encodes classically druggable protein products, which is based on functional domains found in targets of drugs already on the market.
105. Orth, A. P., Batalov, S., Perrone, M. & Chanda, S. K. The promise of genomics to identify novel therapeutic targets. *Expert Opin. Ther. Targets* **8**, 587–596 (2004).
Follow-up reference that updates the set of druggable gene products of the human genome.
106. Al-Shahrour, F., Diaz-Uriarte, R. & Dopazo, J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **20**, 578–580 (2004).
A simple but powerful web application that can be used to extract Gene Ontology (GO) terms that are differentially represented in sets of genes within the context of genome-scale experiments; part of the Babelomics suite.

Acknowledgements
Thanks to V. McKusick and J. Amberger for data from the Online Mendelian Inheritance in Man. Thanks to M. Ludman, S. Dyack and D. Skidmore for valuable discussions of clinical genetics phenotypes. Thanks to P. Goldberg, R. Sherrington, S. Pimstone for discussions of therapeutic targets. Thanks to S. Chanda for logistic assistance. R.R.B. is supported by a Michael Smith Foundation for Health Research (MSFHR) Research Unit Infrastructure award. M.E.S. is supported by the IWK Health Centre, Capital District Health Authority and Dalhousie University Faculty of Medicine.

Competing interests statement
The authors declare no competing financial interests.

DATABASES

The following terms in this article are linked online to:
OMIM: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
α-synuclein | amyloid-β A4 precursor | butyrylcholinesterase | CCR5 | circadian rhythm | *CYP2D6* | diabetes | EPOR | ERBB2 | Fabry disease | favism | fibrillin 1 | fructose intolerance | GHR | *HOXA1* | *HOXA13* | hyperalphalipoproteinaemia | hypertension | hypoglycaemia | IL2RA | IL2RG | ITGA2B | language disorders | leucine-rich repeat kinase 2 | MC4R | microtubule-associated protein tau | obesity | oncogene DJ1 | parkin | phenylketonuria | presenilin 1 | presenilin 2 | PTEN-induced kinase | *TPMT* | transthyretin | ubiquitin carboxylterminase esterase L1

FURTHER INFORMATION

Babelomics: http://www.babelomics.org
Cystic Fibrosis Mutation Database: http://www.genet.sickkids.on.ca/cftr
Eco-Economy Indicators web site: http://www.earth-policy.org/Indicators/Pop/2004.htm
Frontiers in Bioscience Database of Gene Knockouts: http://www.bioscience.org/knockout/knochome.htm
International HapMap Project: http://www.hapmap.org
LDLR Locus web site: http://www.ucl.ac.uk/fh
Online Mendelian Inheritance in Man database: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
PubMed: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

SUPPLEMENTARY INFORMATION

See online article: S1 (figure) | S2 (table) | S3 (table) | S4 (table)
Access to this links box is available online.