JMG Online First, published on September 3, 2014 as 10.1136/jmedgenet-2014-102437

Rare and low-frequency variants in human common diseases and other complex traits

Guillaume Lettre[1,2]

[1]Montreal Heart Institute, Montreal, Quebec, Canada
[2]Faculty of Medicine, Department of Medicine, Université de Montréal, Montreal, Quebec, Canada

Correspondence to
Dr Guillaume Lettre, Montreal Heart Institute, 5000 Bélanger Street, Montréal, Québec, Canada H1T 1C8; guillaume.lettre@umontreal.ca

Received 1 July 2014
Revised 14 August 2014
Accepted 16 August 2014

ABSTRACT

In humans, most of the genetic variation is rare and often population-specific. Whereas the role of rare genetic variants in familial monogenic diseases is firmly established, we are only now starting to explore the contribution of this class of genetic variation to human common diseases and other complex traits. Such large-scale experiments are possible due to the development of next-generation DNA sequencing. Early findings suggested that rare and low-frequency coding variation might have a large effect on human phenotypes (eg, *PCSK9* missense variants on low-density lipoprotein-cholesterol and coronary heart diseases). This observation sparked excitement in prognostic and diagnostic medicine, as well as in genetics-driven strategies to develop new drugs. In this review, I describe results and present initial conclusions regarding some of the recent rare and low-frequency variant discoveries. We can already assume that most phenotype-associated rare and low-frequency variants have modest-to-weak phenotypical effect. Thus, we will need large cohorts to identify them, as for common variants in genome-wide association studies. As we expand the list of associated rare and low-frequency variants, we can also better recognise the current limitations: we need to develop better statistical methods to optimally test association with rare variants, including non-coding variation, and to account for potential confounders such as population stratification.

INTRODUCTION

There is nothing as mysterious as the unknown. This is also true in genetics. For this reason, scientists sequenced the human genome more than a decade ago.[1 2] The aims of the Human Genome Project were to gain insights into the organisation of our genome, but also to understand the role of genetic variation in human diseases and other traits. We have made tremendous progress in assigning functions to each of the ~3.3 billion nucleotides that constitute our genetic code, although much work remains.[3] By comparing our genome sequence with the sequence of other species, we are also starting to learn why we, humans, are different. And by analysing the genome sequence of different human populations, we are beginning to unravel how our genome impacts our phenotypes, including our risk to develop diseases. In this article, I briefly review the types of segregating genetic variation detected in the human genome, with an emphasis on the characterisation of rare and low-frequency sequence variants (figure 1). I arbitrarily define variants with a minor allele frequency (MAF) <0.1% as rare, whereas low-frequency and common variants have MAF of 0.1%–1% and >1%, respectively. My main aim is to draw conclusions on our early successes in order to guide the design of better studies to find genetic associations between rare or low-frequency variants and human complex phenotypes. Although clearly important, I do not discuss the role of de novo or somatic mutations in human phenotypical variation, nor will I extensively describe the different statistical methods specific to the analysis of rare variants. These topics have been recently discussed in excellent review articles.[4–7]

THE HUMAN GENETIC VARIATION THAT WE (THINK WE) UNDERSTAND

Over the last 40 years, positional cloning, linkage studies and DNA sequencing allowed investigators to identify hundreds of mutations responsible for rare human diseases that follow Mendel's laws of inheritance. These mutations, along with the corresponding genotype–phenotype correlations, are extremely well documented in the National Center of Biotechnology Information's Online Mendelian Inheritance in Men database (OMIM; http://www.omim.org/). Thanks to the development of next-generation DNA sequencing (NGS) technologies,[8] Mendelian genetics continues to be in the front-line of research, with weekly reports of new genes mutated in rare human disorders or syndromes. In particular, whole-exome sequencing (WES) makes it possible to identify aetiological mutations for extremely rare diseases even in the absence of pedigrees, a major limitation for the linkage approach.[9]

Until recently, the genetic causes of common human diseases (eg, diabetes, myocardial infarction) and other complex traits (eg, height, blood cholesterol levels) also remained a mystery. The seminal theoretical work by Fisher, published in 1918, predicted what geneticists should be looking for: a large number of genetic variants, each with a very small effect on phenotypes.[10] But it took ~90 years before we could combine conclusions from ground-breaking work on the patterns of common genetic variation in the human genome[11–16] with new genome-wide genotyping technologies to tackle complex trait genetics. We have now identified genetic associations between thousands of 'common' bi-allelic SNPs and human phenotypes.[17 18] These genome-wide association studies (GWAS) have yielded new insights into human biology in health and diseases. Translating these GWAS discoveries is the next frontier. With novel tools (eg, TALEN and CRISPR/Cas9 genome editing methods[19 20]) and resources (eg, epigenomic data from the ENCODE and Roadmap Epigenomic Projects[21 22] and transcriptomic data from FANTOM5[23 24]) available, wet-lab experimentalists can now make significant progress to understand

To cite: Lettre G. *J Med Genet* Published Online First: [*please include* Day Month Year] doi:10.1136/jmedgenet-2014-102437

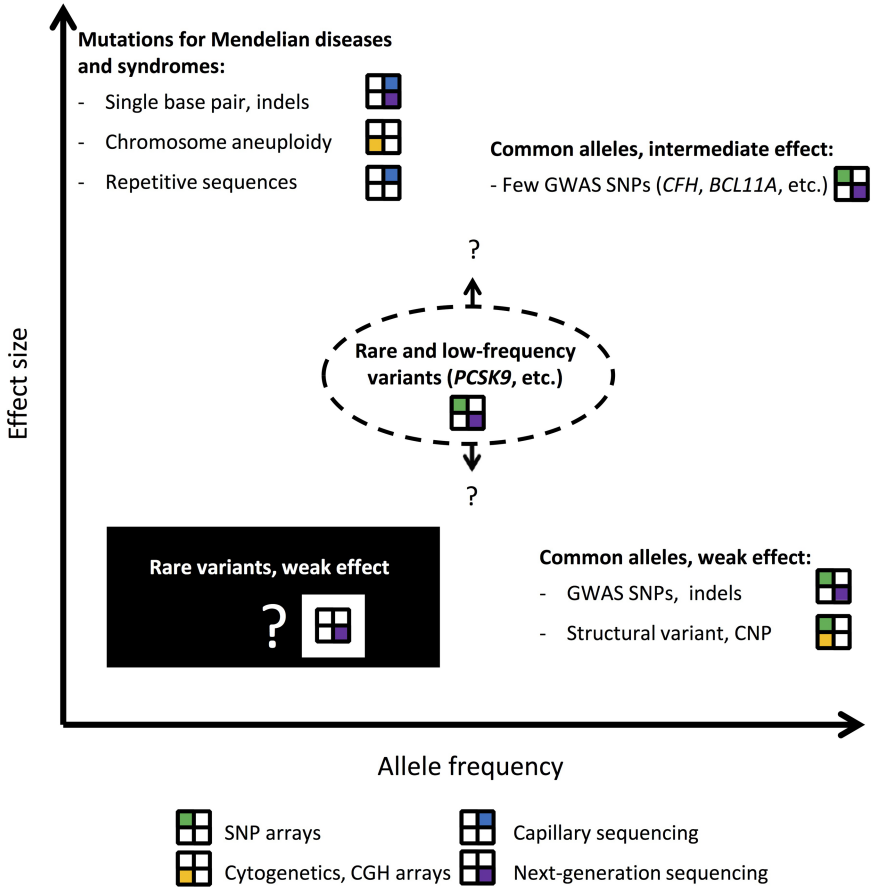

Figure 1 Human genetic variation, phenotypical effect and genomic technologies. A summary of some of the genetic variation in our genome that has been associated with human common diseases and complex traits. The role of repetitive sequence variation and weak effect rare variants in complex trait genetics is still unclear. The different technologies most often used to detect each class of genetic variation are shown. Indel, insertion–deletion; CNP, copy number polymorphism; CGH, comparative genomic hybridisation; GWAS, genome-wide association studies. Adapted from McCarthy *et al* [82].

the molecular mechanisms that drive human phenotypical variation.

When comparing two human genomes, most of the differences in terms of nucleotide changes reside not at SNPs but in large (>1 kilobase) structural variants, such as insertions, deletions and duplications.[25] Improvements in array and sequencing technologies helped to generate accurate, single base resolution maps of these copy number variants (CNVs).[25 26] The excitement and expectations were high regarding the potential influence of CNVs on human phenotypes. Investigators identified associations of CNVs with complex human diseases and traits, including neurocognitive disorders,[27–29] Crohn's disease[30] and body mass index,[31 32] but the number of such associations remained low. This is, in part, due to the technical difficulty in obtaining accurate CNV genotypes in large populations.[33] In a study that tested 3432 CNVs for association with eight common human diseases in ~19 000 participants, the Wellcome Trust Case Control Consortium did not report novel associations.[33] An important conclusion of that study, however, is that most common CNVs are in LD with SNPs normally surveyed by genotyping arrays.[33] Therefore, current large meta-analyses of GWAS results test indirectly the effect of a large subset of common CNV on human phenotypical variation. Although there is probably more than meets the eye, and this may change as we explore further our genome, the current role of common structural variants in complex human diseases and traits appear limited.

RARE AND LOW-FREQUENCY VARIANTS: WE KNOW THEY EXIST, BUT WE DON'T REALLY UNDERSTAND THEM (YET)

One of the main conclusions of the 1000 Genomes Project is that that most of the genetic variation in our genome is rare and private to the different human populations.[15 16] Despite remaining challenges (table 1), studying rare and low-frequency variants is the new *hype* in human genetics for at least three reasons. First, despite its success in finding thousands of SNP associations, the GWAS approach has not yet identified most of the genetic variation that contributes to disease risk of trait variation—the so-called missing heritability paradox.[34] Although theoretical and empirical analyses have determined that a large fraction of the heritability is not missing but, in fact ,hidden in GWAS results,[35 36] it is also true that rare and low-frequency variants, which are usually not tested by genome-wide genotyping arrays, could influence phenotypes. The identification of rare coding variants can also help pinpoint which genes are causal within GWAS loci. Second, early findings in rare variant genetics suggested that this class of variation might have large effects on phenotypes.[37] This is intuitive: the frequency of strong detrimental alleles should be controlled by purifying selection and is also consistent with the observation that most common SNPs identified by GWAS have weak effects. The *poster child* example of this rationale is the identification of low-frequency missense variants in *PCSK9* that are associated with low low-density lipoprotein (LDL)-cholesterol levels and reduced coronary heart disease risk.[38] This finding led to the development of a new class of therapeutics to treat patients with hypercholesterolaemia, paving the way for similar approaches following genetic discoveries.[39] As we will discuss below, it seems that the large phenotypical effect observed for *PCSK9* coding variants is more an exception than the rule. In fact, the weak phenotypical effect observed for many rare variants is consistent with early population genetic work. By considering mutations that cause Mendelian diseases, human–chimpanzee

Table 1 Challenges in the analysis of rare and low-frequency variants in human genetics

Challenge	Description
Technology	Choice between next-generation DNA sequencing and genotyping arrays recently developed to capture rare/low-frequency coding variation. Arrays are less expensive and easier to analyse, but are limited to known genetic variants—this might be more of a concern for experiments in non-European populations. Sequencing is becoming more affordable, but still expensive and computationally intense. Sequencing candidate genes, the whole exome or the whole genome will impact the class of genetic variation discovered and the multiple hypothesis burdens.
Study design	Most published rare-variant association analyses have used unrelated individuals given the relative ease to assemble such experimental design. For the same number of participants, a cohort of related individuals has less power to discover new genetic variants (given that fewer independent chromosomes are tested) than a cohort of unrelated individuals. However, the allele frequency might be higher and the phenotypical effect stronger, thus increasing power. Additional methodological work is needed to compare statistical power to find genetic associations with rare/low-frequency variants in pedigrees vs unrelated individuals, in particular, in the context of gene-based tests.
Statistical analysis	Minor allele frequency (MAF) impacts statistical power. For instance, under some assumptions (OR=1.5, α=5×10^{-8}, population prevalence=5%), we would need >400 000 individuals to have 80% power to find an association with a rare variant (MAF=0.1%). For a common variant (MAF=10%), ∼4600 individuals would be sufficient. Furthermore, because the number of rare variants is higher than the number of common variants in the human genome, the multiple hypothesis burdens for rare-variant association studies is higher, again decreasing statistical power.Statistical tests that combine variants, for instance by gene, have been developed (recently reviewed in ref. 7), although the optimal tests will likely depend on the specific genetic architecture of each phenotype.
Variant annotation	Coding variants are more likely to have phenotypical effects, although a large fraction will be neutral. Bioinformatic tools have been developed to prioritise functional variants, and thus decrease the signal-to-noise ratio, but they are imperfect.[77 78] These tools often also ignore non-coding variants. Private rare non-coding variants can cause Mendelian diseases.[79] Although there are only few (if any) examples of rare non-coding variants associated with complex human traits, they probably exist but we have not carefully looked for them yet. Ideally, experimental validation should guide the selection of likely functional variants before association testing, although this is difficult to implement using high-throughput methods.
Population stratification	Following the original observation that current statistical methods (eg, principal component adjustment) cannot properly account for population stratification of rare variants,[65] a large number of reports have been published, although the optimal method is unclear. Inflation due to population stratification of rare variants might also depend on the type of gene-based tests used.[80] Ideally, having a large number of genotyped or sequenced controls would allow ancestry-based matching with cases.[81]
Phenotypical variance explained	The phenotypical variance explained by a variant depends on the effect size and the allele frequency. For rare variants to explain a large fraction of the missing heritability, phenotypical effects would need to be high. Although this is the case for *PCSK9* and a handful of other genes that harbour penetrant rare alleles, most rare variants will likely have weak-to-modest effects. Using calculations based on empirical data, a recent report suggests that the heritability explained by rare variants could be substantial (18%–84%) but that we would need a very large sample size (>1 000 000 individuals) to find all the associated variants.

divergence and DNA sequence data in a large number of individuals, investigators showed that most rare missense mutations are deleterious in humans and may therefore influence complex human phenotypes. However, the estimated selection coefficients that best fit the data are small, suggesting that most rare deleterious missense variants have small effects on fitness.[40] And finally, from a more practical point of view, rare variant experiments in large DNA collections are only now becoming possible with NGS technologies. It does remain expensive and analytically complicated, but NGS is mature. Several large-scale sequencing projects are now ongoing or completed, such as the Exome Sequence Project that surveyed genetic variation in the exome of 6515 cohort participants.[41]

Initially, DNA resequencing efforts to find rare variants were targeted to candidate genes. These genes were selected based on previous molecular, cellular or genetic (Mendelian diseases, GWAS results) knowledge. Such approach was proven to be successful for blood lipid traits,[38 42–44] but also for other phenotypes such as type 1 and 2 diabetes,[45 46] fetal haemoglobin levels[47] and age-related macular degeneration (AMD).[48 49] A main challenge when sequencing excellent candidate genes pertains to distinguishing pathological from neutral mutations. Two recent studies sequenced genes implicated in diabetes and cardiomyopathies and identified a large number of non-synonymous variants in healthy individuals, highlighting the difficulty in using this genetic information to develop prognostic tests.[50 51] Validating functionally the impact of DNA sequence variants identified remains a priority, and a series of guidelines to demonstrate causality in genotype–phenotype analyses was recently proposed.[52]

Except for neurocognitive disorders, for which NGS has implicated de novo variants,[6] there are currently few examples of WES or whole-genome sequencing (WGS) experiments that have identified rare or low-frequency variants associated with complex human diseases or traits. WES of 91 patients with cystic fibrosis (a monogenic disease) identified missense variants in *DCTN4* that are associated with resistance to *Pseudomonas aeruginosa* infections (a complex trait).[53] WGS in 962 participants did not identify new genetic association with high-density lipoprotein-cholesterol,[54] whereas WES in 2005 individuals found rare variants in one gene, *PNPLA5*, that are associated with LDL-cholesterol.[55] In the cystic fibrosis and LDL-cholesterol studies, 91 and 554 individuals were selected from the extremes of bacterial resistance and LDL-cholesterol levels, respectively. Under an additive genetic effect model, this 'extreme' study design increases statistical power to find variants while limiting the number of samples to sequence.[56]

There is one example where WGS has been successful for common human diseases. The Iceland-based deCODE genetics company has reported several associations between strong effect rare/low-frequency variants identified by WGS and diseases. These include variants in *TREM2* and *APP* associated with Alzheimer's disease,[57 58] a nonsense variant in *LGR4* with osteoporosis,[59] a variant in *C3* with AMD[60] and several variants with type 2 diabetes.[61] Importantly, other investigators have replicated some of the associations with Alzheimer's disease, AMD and type 2 diabetes.[48 49 62–64] For all these findings, deCODE's approach was similar: they identified genetic variation in the Icelandic population by WGS of ∼2000 participants. Then, they imputed the identified genetic variants using long-phase haplotyping methodology in ∼90 000 participants genotyped on GWAS-type arrays. Finally, they used the extensive genealogy of this population to infer genotypes in >250 000 individuals. Although the sample size of these studies

is very large, the number of cases remains in the 'normal' range for association studies: for instance, there were 1143 and 11 114 cases in the recent AMD and type 2 diabetes studies, respectively.[60 61] The high control-to-case ratio (45:1 for AMD, 24:1 for type 2 diabetes) improves power, although it stabilises as the number of controls increases. deCODE's successes are also explained by the phenotypical, genetic and environmental homogeneity of its participants, which minimises potential confounders. This might be particularly important for association studies of rare and low-frequency variants.[65 66] Further supporting the importance to work with homogenous populations, a WES experiment in large families identified a rare missense variant in *PLD3* that is associated with late-onset Alzheimer's disease.[67] The deCODE studies highlight that population isolates and large pedigrees might be particular useful for rare and low-frequency variant studies. Furthermore, imputing variants into already genotyped samples might be a powerful approach to minimise sequence costs while maximising power. Recently, we used a similar strategy—WES in 761 African-Americans and imputation in ~13 000 genotyped African-Americans—to find new associations with blood cell phenotypes.[68]

SEQUENCING BY DIRECT GENOTYPING

One of the conclusions from the early large-scale NGS experiments is that we need large sample size to find new genetic associations with rare or low-frequency variants. The variants identified so far have large effect size—often OR >2—but we have found only a handful despite having sequenced large cohorts with different complex phenotypes available. And retrospectively, we probably have not performed to date well-powered NGS experiments: we found large effect variants because we only had power to find such variants. Based on our few findings, it seems likely that most rare or low-frequency variants will have modest-to-weak effect on phenotypes. But how to test rare/low-frequency variants in tens of thousands of samples?

Exome arrays were designed precisely to answer this need, that is, to develop a tool that would allow large-scale testing of coding variation in very large sample sizes at moderate costs (<10% of what WES costs if we include analysis time). To design the Illumina Infinium HumanExome Beadchip, investigators combined genetic variation identified by WES or WGS of ~12 000 individuals and selected ~250 000 variants for the exome array (http://genome.sph.umich.edu/wiki/Exome_Chip_Design). These variants have been seen at least three times in two different studies and are highly enriched for protein altering functions (missense, nonsense, splice site). Affymetrix has also generated a similar exome array. Exome chips are convenient because of their simplicity, but also have certain limitations. First, many coding and all non-coding rare variants are not tested by exome arrays. For an exhaustive analysis of this class of genetic variation, direct DNA sequencing remains necessary. Second, exome chips might not capture as well coding variation in different populations. Most of the sequence data used to generate the genetic variation catalogue for the exome chip was from individuals of European ancestry. Thus, exome chip experiments in other populations might miss a large fraction of the coding variation that is ancestry-specific or population-specific. As a dramatic example, we recently sequenced the exome of 164 African-Americans that were also genotyped on the Illumina exome chip: 67% of the coding variation—mostly very rare, however—was not surveyed by the exome array (Ken Sin Lo and GL, unpublished). This is an important flag to remember in deciding between NGS and exome chip genotyping for experiments in non-European ancestry populations, especially because LD will not be helpful to tag variants at such low MAF.

Genetic discovery experiments based on the exome array approach already have some successes (table 2). The first report focused on insulin processing and secretion in individuals from Finland.[64] The authors identified four missense and one nonsense variants strongly associated with these insulin traits. Two of these variants fell within, but were independent from, GWAS signals for the same phenotypes; these low-frequency variants implicate *SGSM2* and *MADD* as causal genes for insulin secretion (table 2). The three remaining variants did not overlap with GWAS loci for insulin indexes. This study identified the same variant in *PAM* (p.Asp563Gly) that was found to be associated with type 2 diabetes risk by the deCODE group.[61] Blood lipid traits were also analysed in large populations genotyped on exome arrays, leading to the identification of coding variation at five loci (table 2).[69 70] A low-frequency variant in *TM6SF2* (p.Glu167Lys) is associated with total cholesterol levels and alanine transaminase (a marker of liver injury), as well as two related clinical endpoints: myocardial infarction and non-alcoholic fatty liver disease.[70 71] This *TM6SF2* variant explains the GWAS signal for these phenotypes at the locus. Finally, we used the exome chip to identify coding variants associated with blood cell phenotypes in ~30 000 Europeans or individuals of European descent.[72] We reported the first erythropoietin variant associated with haemoglobin and haematocrit levels, a rare missense variant in the thrombocytopenia gene *TUBB1* associated with platelet count, and a collection of eight missense variants in the chemokine receptor gene *CXCR2* associated with white blood cell counts (table 2). We further demonstrated that a *CXCR2* frameshift mutation segregating in a family is responsible for congenital neutropenia.[72] Several large consortia, with access to exome chip genotype data for hundreds of thousands of individuals, are in progress and should yield many additional rare and low-frequency coding variants associated with human phenotypes.

AND THERE IS THE PART OF OUR GENOME THAT WE DON'T UNDERSTAND: REPETITIVE SEQUENCES

We often present NGS methods as a solution to all our genetic problems given their unprecedented capacity to generate DNA sequences. But we forget that a non-negligible fraction of our genome—repetitive DNA sequences that cover over half of the human genome—is largely refractory to this technology. Repeats correspond to segments of DNA, almost identical, that can be found at several locations and on different chromosomes. They can be short (1–2 bps motif) or long (several kilobases). The transposon element *Alu* is our most abundant repetitive sequence, representing ~11% of the human genome.[1 2] Variation in the number of repeats at specific loci has been linked to many human pathologies, most notably the expansion of triplet nucleotides in Huntington's disease, fragile X syndrome, myotonic dystrophy and other disorders.[73] From a NGS perspective, repeats are problematic because the corresponding sequence reads are usually too short and cannot be mapped unambiguously. This introduces sequence errors and difficulties in interpreting results.[74]

Medullary cystic kidney disease type 1 (MCKD1) is a Mendelian disease that was mapped to a two megabases interval on chromosome 1 by linkage studies more than a decade ago. More recently, investigators used WES and WGS but did not find mutations that segregated perfectly with disease status in affected pedigrees. They eventually used 'old-fashioned'

Table 2 New genetic associations between rare or low-frequency variants and human complex traits identified using the ExomeChip

Trait	Population	Sample size	Gene	Variant	Minor allele frequency	Effect size (in SD units)	GWAS locus	Reference
Insulin processing and secretion	Europeans (Finland)	8229	*SGSM2*	rs61741902 (p.Val996Ile)	1.4%	0.41	Yes, but independent from SNP	[64]
			MADD	rs35233100 (p.Arg766X)	3.7%	−0.32	Yes, but independent from SNP	
			TBC1D30	rs150781447 (p.Arg279Cys)	2.0%	0.50	No	
			KANK1	rs3824420 (p.Arg667His)	2.9%	0.28	No	
			PAM	rs35658696 (p.Asp563Gly)	5.3%	−0.21	No	
Alanine transaminase (a marker of liver injury)	EA, AA and HA	882 (EA), 1324 (AA), 467 (HA)	*TM6SF2*	rs58542926 (p.Glu167Lys)	7.2% (EA), 3.4% (AA), 4.7% (HA)	2.0 alanine transaminase unit	Yes, explain the GWAS signal	[71]
Blood lipids	EA and AA	42 208 (EA), 14 330 (AA)	*ANGPTL8*	rs145464906 (p.Gln121Stop)	0.1% (EA), 0.01% (AA)	0.77	Yes, but independent from SNP	[69]
			PAFAH1B2	rs186808413 (p.Ser161Leu)	1.1% (EA), 0.2% (AA)	0.23 (HDL), −1.46 (TG)	Yes, but independent from SNPs	
			COL18A1	rs114139997 (p.Gly111Arg)	0.003% (EA), 1.9% (AA)	0.15 (HDL), −2.34 (TG)	No	
			PCSK7	rs142953140 (p.Arg504His)	0% (EA), 0.2% (AA)	1.31 (HDL), −4.39 (TG)	Yes, but independent from SNPs	
	Europeans (Norway)	10 309	*TM6SF2*	rs58542926 (p.Glu167Lys)	8.9%	−0.19 (TC)	Yes, explain the GWAS signal	[70]
Blood cell traits	EA, French Canadians and Europeans (Germany)	31 340	*EPO*	rs62483572 (p.Asp70Asn)	0.4%	−0.22 (HCT), −0.21 (HGB)	Yes, but independent from SNP	[72]
			TUBB1	rs41303899 (p.Gly109Glu)	0.2%	−0.49 (PLT)	Yes, but independent from SNP	
			CXCR2	8 missense variants	0.005%–0.5%	−0.23 (WBC)	No	

Otherwise noted, effect sizes are in SD units.
AA, African-Americans; EA, European Americans; HA, Hispanic Americans; HCT, haematocrit; HDL, high-density lipoprotein cholesterol; HGB, haemoglobin; PLT, platelet; SNP, single nucleotide variation; TC, total cholesterol; TG, triglycerides; WBC, white blood cell; GWAS, genome-wide association studies.

positional cloning, capillary sequencing and de novo assembly methods to discover that MCKD1 is caused by a cytosine insertion in one repeat of a variable number tandem repeat (VNTR) in the *MUC1* gene.[75] The *MUC1* VNTR, very guanine–cytosine-rich, could not be sequenced by WES and was underrepresented in the WGS data. The identification of the causal mutation for MCKD1 serves as an illustrative example in considering the challenges to analyse repetitive DNA sequences by NGS. Whether such repeat sequence variation (common or rare) could also impact complex trait genetics remains to be tested.

CONCLUSION

Driven by the sequencing of the human genome and technological advancements, human geneticists have made great progress in the identification of genetic variation that cause simple and complex human diseases or that influence other human phenotypes. The new excitement in the field is in the characterisation of rare and low-frequency variants, in part because such variants might have larger phenotypical effects and might therefore be more clinically actionable than GWAS SNPs in the context of personalised medicine and drug development. Although there are clearly rare/low-frequency large-effect variants, their number is likely going to be small given insights from the completed studies. Large sample sizes are needed for comprehensive studies of rare and low-frequency variants. Other challenges include the development of new statistical methods to test association between functionally related groups of variants (gene-based, but could also be pathway-based, promoter-based or enhancer-based) as well as to explore the contribution of rare non-coding genetic variation on human phenotypical variation. Finally, because rare variation is mostly population-specific, it will be important to improve methods to correct for confounders such as population stratification because existing approaches are not appropriate.[65 66] This is particularly important to avoid some of the early pitfalls of the common variant association testing the literature.[76] The coming years will mark another chapter in the history on the exploration of our genome. It will be interesting to see how this rare/low-frequency variant adventure contrasts with the previous chapters on positional cloning, capillary sequencing and GWAS. And how it may provide ideas and tools to study in the future repetitive DNA sequences as it relates to human phenotypical variation.

Acknowledgements I would like to thank Chris Cotsapas and Ekat Kritikou, as well as all the members of my laboratory for suggestions and comments on an early version of this manuscript. I apologise to all my colleagues whose work could not be cited because of space constraints. Work in my laboratory was funded by the Canadian Institute of Health Research (#243400), the Canada Research Chair programme, Genome Canada/Genome Quebec, the Doris Duke Charitable Foundation (#2012126) and the Montreal Heart Institute Foundation.

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

1 Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing C. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.

2 Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science* 2001;291:1304–51.

3 Lander ES. Initial impact of the sequencing of the human genome. *Nature* 2011;470:187–97.

4 Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet* 2010;11:773–85.

5 Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet* 2014;15:335–46.

6 Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet* 2012;13:565–75.

7 Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* 2014;95:5–23.

8 Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;14:681–91.

9 Wang Z, Liu X, Yang BZ, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. *Front Genet* 2013;4:160.
10 Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb* 1918;52:399–433.
11 Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001;29:229–32.
12 Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature* 2005;437:1299–320.
13 Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Sun W, Wang H, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Yakub I, Birren BW, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–61.
14 Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010;467:52–8.
15 Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
16 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.
17 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–53.
18 McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 2008;17:R156–165.
19 Chen K, Gao C. TALENs: customizable molecular DNA scissors for genome engineering of plants. *J Genet Genomics* 2013;40:271–9.
20 Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 2014;32:347–55.
21 Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shoresh N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kelllis M, Khatun J, Kheradpour P, Kundaje A, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elinitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigo R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Starnatoyannopoulos JA, Tennebaum SA, Weng Z, White KP, Wold B, Khatun J, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Giddings MC, Bernstein BE, Epstein CB, Shoresh N, Ernst J, Kheradpour P, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Ward LD, Altshuler RC, Eaton ML, Kellis M, Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena HP, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Risk BA, Robyr D, Ruan X, Sammeth M, Sandu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Hayashizaki Y, Harrow J, Gerstein M, Hubbard TJ, Reymond A, Antonarakis SE, Hannon GJ, Giddings MC, Ruan Y, Wold B, Carninci P, Guigo R, Gingeras TR, Rosenbloom KR, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kent WJ, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Furey TS, Song L, Grasfeder LL, Giresi PG, Lee BK, Battenhouse A, Sheffield NC, Simon JM, Showers KA, Safi A, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Birney E, Iyer VR, Lieb JD, Crawford GE, Li G, Sandhu KS, Zheng M, Wang P, Luo OJ, Shahab A, Fullwood MJ, Ruan X, Ruan Y, Myers RM, Pauli F, Williams BA, Gertz J, Marinov GK, Reddy TE, Vielmetter J, Partridge EC, Trout D, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, King B, Muratet MA, Antoshechkin I, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Gunter C, Newberry JS, Levy SE, Absher DM, Mortazavi A, Wong WH, Wold B, Blow MJ, Visel A, Pennachio LA, Elnitski L, Margulies EH, Parker SC, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Chrast J, Davidson C, Derrien T, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Howald C, Hunt T, Jungreis I, Kay M, Khurana E, Kokocinski F, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tanzer A, Tapanan E, Tress ML, van Baren MJ, Walters N, Washieti S, Wilming L, Zadissa A, Zhengdong Z, Brent M, Haussler D, Kellis M, Valencia A, Gerstein M, Raymond A, Guigo R, Harrow J, Hubbard TJ, Landt SG, Frietze S, Abyzov A, Addleman N, Alexander RP, Auerbach RK, Balasubramanian S, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Cheng C, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyenger S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Larnarre-Vincent N, Leng J, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Pei B, Raha D, Ramirez L, Reed B, Rozowsky J, Sboner A, Shi M, Sisu C, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Yip KY, Zhang Z, Struhl K, Weissman SM, Gerstein M, Farnham PJ, Snyder M, Tenebaum SA, Penalva LO, Doyle F, Karmakar S, Landt SG, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Patacsil D, Slifer T, Victorsen A, Yang X, Snyder M, White KP, Auer T, Centarin L, Eichenlaub M, Gruhl F, Heerman S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Weng Z, Whitfield TW, Wang J, Collins PJ, Aldred SF, Trinklein ND, Partridge EC, Myers RM, Dekker J, Jain G, Lajoie BR, Sanyal A, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Jain G, Johnson AK, Johnson EM, Kutyavin TM, Lajoie BR, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sabo P, Sanchez ME, Sandstrom RS, Sanyal A, Shafer AO, Stergachis AB, Thomas S, Thurman RE, Vernot B, Vierstra J, Vong S, Wang H, Weaver MA, Yan Y, Zhang M, Akey JA, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Kaul R, Dekker J, Stamatoyannopoulos JA, Dunham I, Beal K, Brazma A, Flicek P, Herrero J, Johnson N, Keefe D, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Wilder SP, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Kundaje A, Hardison RC, Miller W, Giardine B, Harris RS, Wu W, Bickel PJ, Banfai B, Boley NP, Brown JB, Huang H, Li Q, Li JJ, Noble WS, Bilmes JA, Buske OJ, Hoffman MM, Sahu AO, Kharchenko PV, Park PJ, Baker D, Taylor J, Weng Z, Iyer S, Dong X, Greven M, Lin X, Wang J, Xi HS, Zhuang J, Gerstein M, Alexander RP, Balasubramanian S, Cheng C, Harmanci A, Lochovsky L, Min R, Mu XJ, Rozowsky J, Yan KK, Yip KY, Birney E. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
22 Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S,

Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012;337:1190–5.

23 Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jorgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Muller F, Consortium F, Forrest AR, Carninci P, Rehli M, Sandelin A. An atlas of active enhancers across human cell types and tissues. *Nature* 2014;507:455–61.

24 Consortium F; the RP, Clst. A promoter-level mammalian expression atlas. *Nature* 2014;507:462–70.

25 Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Wellcome Trust Case Control CTyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Origins and functional impact of copy number variation in the human genome. *Nature* 2010;464:704–12.

26 Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stutz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurles ME, Lee C, McCarroll SA, Korbel JO, Genomes P. Mapping copy number variation by population-scale genome sequencing. *Nature* 2011;470:59–65.

27 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bolte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010;466:368–72.

28 Stefansson H, Rujescu D, Cichon S, Pietilainen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD, Muglia P, Francks C, Matthews PM, Gylfason A, Halldorsson BV, Gudbjartsson D, Thorgeirsson TE, Sigurdsson A, Jonasdottir A, Jonasdottir A, Bjornsson A, Mattiasdottir S, Blondal T, Haraldsson M, Magnusdottir BB, Giegling I, Moller HJ, Hartmann A, Shianna KV, Ge D, Need AC, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Paunio T, Toulopoulou T, Bramon E, Di Forti M, Murray R, Ruggeri M, Vassos E, Tosato S, Walshe M, Li T, Vasilescu C, Muhleisen TW, Wang AG, Ullum H, Djurovic S, Melle I, Olesen J, Kiemeney LA, Franke B, GroupSabatti C, Freimer NB, Gulcher JR, Thorsteinsdottir U, Kong A, Andreassen OA, Ophoff RA, Georgi A, Rietschel M, Werge T, Petursson H, Goldstein DB, Nothen MM, Peltonen L, Collier DA, St Clair D, Stefansson K. Large recurrent microdeletions associated with schizophrenia. *Nature* 2008;455:232–6.

29 McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O, Krause V, Kumar RA, Grozeva D, Malhotra D, Walsh T, Zackai EH, Kaplan P, Ganesh J, Krantz ID, Spinner NB, Roccanova P, Bhandari A, Pavon K, Lakshmi B, Leotta A, Kendall J, Lee YH, Vacic V, Gary S, Iakoucheva LM, Crow TJ, Christian SL, Lieberman JA, Stroup TS, Lehtimaki T, Puura K, Haldeman-Englert C, Pearl J, Goodell M, Willour VL, Derosse P, Steele J, Kassem L, Wolff J, Chitkara N, McMahon FJ, Malhotra AK, Potash JB, Schulze TG, Nothen MM, Cichon S, Rietschel M, Leibenluft E, Kustanovich V, Lajonchere CM, Sutcliffe JS, Skuse D, Gill M, Gallagher L, Mendell NR, Wellcome Trust Case Control CCraddock N, Owen MJ, O'Donovan MC, Shaikh TH, Susser E, Delisi LE, Sullivan PF, Deutsch CK, Rapoport J, Levy DL, King MC, Sebat J. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 2009;41:1223–7.

30 McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, Duerr RH, Silverberg MS, Taylor KD, Rioux JD, Altshuler D, Daly MJ, Xavier RJ. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* 2008;40:1107–12.

31 Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C, Lettre G, Lim N, Lyon HN, McCarroll SA, Papadakis K, Qi L, Randall JC, Roccasecca RM, Sanna S, Scheet P, Weedon MN, Wheeler E, Zhao JH, Jacobs LC, Prokopenko I, Soranzo N, Tanaka T, Timpson NJ, Almgren P, Bennett A, Bergman RN, Bingham SA, Bonnycastle LL, Brown M, Burtt NP, Chines P, Coin L, Collins FS, Connell JM, Cooper C, Smith GD, Dennison EM, Deodhar P, Elliott P, Erdos MR, Estrada K, Evans DM, Gianniny L, Gieger C, Gillson CJ, Guiducci C, Hackett R, Hadley D, Hall AS, Havulinna AS, Hebebrand J, Hofman A, Isomaa B, Jacobs KB, Johnson T, Jousilahti P, Jovanovic Z, Khaw KT, Kraft P, Kuokkanen M, Kuusisto J, Laitinen J, Lakatta EG, Luan J, Luben RN, Mangino M, McArdle WL, Meitinger T, Mulas A, Munroe PB, Narisu N, Ness AR, Northstone K, O'Rahilly S, Purmann C, Rees MG, Ridderstrale M, Ring SM, Rivadeneira F, Ruokonen A, Sandhu MS, Saramies J, Scott LJ, Scuteri A, Silander K, Sims MA, Song K, Stephens J, Stevens S, Stringham HM, Tung YC, Valle TT, Van Duijn CM, Vimaleswaran KS, Vollenweider P, Waeber G, Wallace C, Watanabe RM, Waterworth DM, Watkins N, Witteman JC, Zeggini E, Zhai G, Zillikens MC, Altshuler D, Caulfield MJ, Chanock SJ, Farooqi IS, Ferrucci L, Guralnik JM, Hattersley AT, Hu FB, Jarvelin MR, Laakso M, Mooser V, Ong KK, Ouwehand WH, Salomaa V, Samani NJ, Spector TD, Tuomi T, Tuomilehto J, Uda M, Uitterlinden AG, Wareham NJ, Deloukas P, Frayling TM, Groop LC, Hayes RB, Hunter DJ, Mohlke KL, Peltonen L, Schlessinger D, Strachan DP, Wichmann HE, McCarthy MI, Boehnke M, Barroso I, Abecasis GR, Hirschhorn JN. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009;41:25–34.

32 Falchi M, El-Sayed Moustafa JS, Takousis P, Pesce F, Bonnefond A, Andersson-Assarsson JC, Sudmant PH, Dorajoo R, Al-Shafai MN, Bottolo L, Ozdemir E, So HC, Davies RW, Patrice A, Dent R, Mangino M, Hysi PG, Dechaume A, Huyvaert M, Skinner J, Pigeyre M, Caiazzo R, Raverdy V, Vaillant E, Field S, Balkau B, Marre M, Visvikis-Siest S, Weill J, Poulain-Godefroy O, Jacobson P, Sjostrom L, Hammond CJ, Deloukas P, Sham PC, McPherson R, Lee J, Tai ES, Sladek R, Carlsson LM, Walley A, Eichler EE, Pattou F, Spector TD, Froguel P. Low copy number of the salivary amylase gene predisposes to obesity. *Nat Genet* 2014;46:492–7.

33 Wellcome Trust Case Control CCraddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S, Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G, Langford CF, Lango-Allen H, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL, McGuffin P, McLay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C, Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR, Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ, Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, RenwickRing SM, Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ, Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z, Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker M, Walker NM, Wallace C, Warren-Perry M, Watkins NA, Webster J, Weedon MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP, Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010;464:713–20.

34 Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;11:446–50.
35 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;42:565–9.
36 Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 2012;109:1193–8.
37 Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008;40:695–701.
38 Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* 2006;354:1264–72.
39 Roth EM, McKenney JM, Hanotin C, Asset G, Stein EA. Atorvastatin with or without an antibody to PCSK9 in primary hypercholesterolemia. *N Engl J Med* 2012;367:1891–900.
40 Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 2007;80:727–39.
41 Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, Akey JM. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013;493:216–20.
42 Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004;305:869–72.
43 Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, Schwartz SM, Voight BF, Elosua R, Salomaa V, O'Donnell CJ, Dallinga-Thie GM, Anand SS, Yusuf S, Huff MW, Kathiresan S, Hegele RA. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 2010;42:684–7.
44 Sanna S, Li B, Mulas A, Sidore C, Kang HM, Jackson AU, Piras MG, Usala G, Maninchedda G, Sassu A, Serra F, Palmas MA, Wood WH III, Njolstad I, Laakso M, Hveem K, Tuomilehto J, Lakka TA, Rauramaa R, Boehnke M, Cucca F, Uda M, Schlessinger D, Nagaraja R, Abecasis GR. Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet* 2011;7:e1002198.
45 Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009;324:387–9.
46 Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, Mahajan A, Fuchsberger C, Atzmon G, Benediktsson R, Blangero J, Bowden DW, Brandslund I, Brosnan J, Burslem F, Chambers J, Cho YS, Christensen C, Douglas DA, Duggirala R, Dymek Z, Farjoun Y, Fennell T, Fontanillas P, Forsen T, Gabriel S, Glaser B, Gudbjartsson DF, Hanis C, Hansen T, Hreidarsson AB, Hveem K, Ingelsson E, Isomaa B, Johansson S, Jorgensen T, Jorgensen ME, Kathiresan S, Kong A, Kooner J, Kravic J, Laakso M, Lee JY, Lind L, Lindgren CM, Linneberg A, Masson G, Meitinger T, Mohlke KL, Molven A, Morris AP, Potluri S, Rauramaa R, Ribel-Madsen R, Richard AM, Rolph T, Salomaa V, Segre AV, Skarstrand H, Steinthorsdottir V, Stringham HM, Sulem P, Tai ES, Teo YY, Teslovich T, Thorsteinsdottir U, Trimmer JK, Tuomi T, Tuomilehto J, Vaziri-Sani F, Voight BF, Wilson JG, Boehnke M, McCarthy MI, Njolstad PR, Pedersen O, Go TDC, Consortium TDG, Groop L, Cox DR, Stefansson K, Altshuler D. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* 2014;46:357–63.
47 Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat Genet* 2010;42:1049–51.
48 Zhan X, Larson DE, Wang C, Koboldt DC, Sergeev YV, Fulton RS, Fulton LL, Fronick CC, Branham KE, Bragg-Gresham J, Jun G, Hu Y, Kang HM, Liu D, Othman M, Brooks M, Ratnapriya R, Boleda A, Grassmann F, von Strachwitz C, Olson LM, Buitendijk GH, Hofman A, van Duijn CM, Cipriani V, Moore AT, Shahid H, Jiang Y, Conley YP, Morgan DJ, Kim IK, Johnson MP, Cantsilieris S, Richardson AJ, Guymer RH, Luo H, Ouyang H, Licht C, Pluthero FG, Zhang MM, Zhang K, Baird PN, Blangero J, Klein ML, Farrer LA, DeAngelis MM, Weeks DE, Gorin MB, Yates JR, Klaver CC, Pericak-Vance MA, Haines JL, Weber BH, Wilson RK, Heckenlively JR, Chew EY, Stambolian D, Mardis ER, Swaroop A, Abecasis GR. Identification of a rare coding variant in complement 3 associated with age-related macular degeneration. *Nat Genet* 2013;45:1375–9.
49 Seddon JM, Yu Y, Miller EC, Reynolds R, Tan PL, Gowrisankar S, Goldstein JI, Triebwasser M, Anderson HE, Zerbib J, Kavanagh D, Souied E, Katsanis N, Daly MJ, Atkinson JP, Raychaudhuri S. Rare variants in CFI, C3 and C9 are associated with high risk of advanced age-related macular degeneration. *Nat Genet* 2013;45:1366–70.
50 Flannick J, Beer NL, Bick AG, Agarwala V, Molnes J, Gupta N, Burtt NP, Florez JC, Meigs JB, Taylor H, Lyssenko V, Irgens H, Fox E, Burslem F, Johansson S, Brosnan MJ, Trimmer JK, Newton-Cheh C, Tuomi T, Molven A, Wilson JG, O'Donnell CJ, Kathiresan S, Hirschhorn JN, Njolstad PR, Rolph T, Seidman JG, Gabriel S, Cox DR, Seidman CE, Groop L, Altshuler D. Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nat Genet* 2013;45:1380–5.
51 Bick AG, Flannick J, Ito K, Cheng S, Vasan RS, Parfenov MG, Herman DS, DePalma SR, Gupta N, Gabriel SB, Funke BH, Rehm HL, Benjamin EJ, Aragam J, Taylor HA Jr, Fox ER, Newton-Cheh C, Kathiresan S, O'Donnell CJ, Wilson JG, Altshuler DM, Hirschhorn JN, Seidman JG, Seidman C. Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. *Am J Hum Genet* 2012;91:513–19.
52 MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, Conrad DF, Cooper GM, Cox NJ, Daly MJ, Gerstein MB, Goldstein DB, Hirschhorn JN, Leal SM, Pennacchio LA, Stamatoyannopoulos JA, Sunyaev SR, Valle D, Voight BF, Winckler W, Gunter C. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014;508:469–76.
53 Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, Wright FA, Rieder MJ, Tabor HK, Nickerson DA, Barnes KC, National Heart L, Blood Institute GOESPLung GO, Gibson RL, Bamshad MJ. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat Genet* 2012;44:886–9.
54 Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A, Muzny D, Yu F, Rice K, Zhu C, Bis J, Heiss G, O'Donnell CJ, Psaty BM, Cupples LA, Gibbs R, Boerwinkle E, Cohorts for H, Aging Research in Genetic Epidemiology C. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet* 2013;45:899–901.
55 Lange LA, Hu Y, Zhang H, Xue C, Schmidt EM, Tang ZZ, Bizon C, Lange EM, Smith JD, Turner EH, Jun G, Kang HM, Peloso G, Auer P, Li KP, Flannick J, Zhang J, Fuchsberger C, Gaulton K, Lindgren C, Locke A, Manning A, Sim X, Rivas MA, Holmen OL, Gottesman O, Lu Y, Ruderfer D, Stahl EA, Duan Q, Li Y, Durda P, Jiao S, Isaacs A, Hofman A, Bis JC, Correa A, Griswold ME, Jakobsdottir J, Smith AV, Schreiner PJ, Feitosa MF, Zhang Q, Huffman JE, Crosby J, Wassel CL, Do R, Franceschini N, Martin LW, Robinson JG, Assimes TL, Crosslin DR, Rosenthal EA, Tsai M, Rieder MJ, Farlow DN, Folsom AR, Lumley T, Fox ER, Carlson CS, Peters U, Jackson RD, van Duijn CM, Uitterlinden AG, Levy D, Rotter JI, Taylor HA, Gudnason V Jr, Siscovick DS, Fornage M, Borecki IB, Hayward C, Rudan I, Chen YE, Bottinger EP, Loos RJ, Saetrom P, Hveem K, Boehnke M, Groop L, McCarthy M, Meitinger T, Ballantyne CM, Gabriel SB, O'Donnell CJ, Post WS, North KE, Reiner AP, Boerwinkle E, Psaty BM, Altshuler D, Kathiresan S, Lin DY, Jarvik GP, Cupples LA, Kooperberg C, Wilson JG, Nickerson DA, Abecasis GR, Rich SS, Tracy RP, Willer CJ, Project NGOES. Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am J Hum Genet* 2014;94:233–45.
56 Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol* 2011;35:790–9.
57 Jonsson T, Stefansson H, Steinberg S, Jonsdottir I, Jonsson PV, Snaedal J, Bjornsson S, Huttenlocher J, Levey AI, Lah JJ, Rujescu D, Hampel H, Giegling I, Andreassen OA, Engedal K, Ulstein I, Djurovic S, Ibrahim-Verbaas C, Hofman A, Ikram MA, van Duijn CM, Thorsteinsdottir U, Kong A, Stefansson K. Variant of TREM2 associated with the risk of Alzheimer's disease. *N Engl J Med* 2013;368:107–16.
58 Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, Hoyte K, Gustafson A, Liu Y, Lu Y, Bhangale T, Graham RR, Huttenlocher J, Bjornsdottir G, Andreassen OA, Jonsson EG, Palotie A, Behrens TW, Magnusson OT, Kong A, Thorsteinsdottir U, Watts RJ, Stefansson K. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 2012;488:96–9.
59 Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, Jonasdottir A, Oddsson A, Helgason A, Magnusson OT, Walters GB, Frigge ML, Helgadottir HT, Johannsdottir H, Bergsteinsdottir K, Ogmundsdottir MH, Center JR, Nguyen TV, Eisman JA, Christiansen C, Steingrimsson E, Jonasson JG, Tryggvadottir L, Eyjolfsson GI, Theodors A, Jonsson T, Ingvarsson T, Olafsson I, Rafnar T, Kong A, Sigurdsson G, Masson G, Thorsteinsdottir U, Stefansson K. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* 2013;497:517–20.
60 Helgason H, Sulem P, Duvvari MR, Luo H, Thorleifsson G, Stefansson H, Jonsdottir I, Masson G, Gudbjartsson DF, Walters GB, Magnusson OT, Kong A, Rafnar T, Kiemeney LA, Schoenmaker-Koller FE, Zhao L, Boon CJ, Song Y, Fauser S, Pei M, Ristau T, Patel S, Liakopoulos S, van de Ven JP, Hoyng CB, Ferreyra H, Duan Y, Bernstein PS, Geirsdottir A, Helgadottir G, Stefansson E, den Hollander AI, Zhang K, Jonasson F, Sigurdsson H, Thorsteinsdottir U, Stefansson K. A rare nonsynonymous sequence variant in C3 is associated with high risk of age-related macular degeneration. *Nat Genet* 2013;45:1371–4.
61 Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, Helgadottir HT, Johannsdottir H, Magnusson OT, Gudjonsson SA, Justesen JM, Harder MN, Jorgensen ME, Christensen C, Brandslund I, Sandbaek A, Lauritzen T,

Vestergaard H, Linneberg A, Jorgensen T, Hansen T, Daneshpour MS, Fallah MS, Hreidarsson AB, Sigurdsson G, Azizi F, Benediktsson R, Masson G, Helgason A, Kong A, Gudbjartsson DF, Pedersen O, Thorsteinsdottir U, Stefansson K. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* 2014;46:294–8.

62 Jin SC, Benitez BA, Karch CM, Cooper B, Skorupa T, Carrell D, Norton JB, Hsu S, Harari O, Cai Y, Bertelsen S, Goate AM, Cruchaga C. Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet* 2014. doi:10.1093/hmg/ddu277

63 Guerreiro R, Wojtas A, Bras J, Carrasquillo M, Rogaeva E, Majounie E, Cruchaga C, Sassi C, Kauwe JS, Younkin S, Hazrati L, Collinge J, Pocock J, Lashley T, Williams J, Lambert JC, Amouyel P, Goate A, Rademakers R, Morgan K, Powell J, St George-Hyslop P, Singleton A, Hardy J, Alzheimer Genetic Analysis G. TREM2 variants in Alzheimer's disease. *N Engl J Med* 2013;368:117–27.

64 Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stancakova A, Stringham HM, Sim X, Yang L, Fuchsberger C, Cederberg H, Chines PS, Teslovich TM, Romm JM, Ling H, McMullen I, Ingersoll R, Pugh EW, Doheny KF, Neale BM, Daly MJ, Kuusisto J, Scott LJ, Kang HM, Collins FS, Abecasis GR, Watanabe RM, Boehnke M, Laakso M, Mohlke KL. Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* 2013;45:197–201.

65 Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet* 2012;44:243–6.

66 O'Connor TD, Kiezun A, Bamshad M, Rich SS, Smith JD, Turner E, Project NES, Esp Population Genetics SAWGLeal SM, Akey JM. Fine-scale patterns of population stratification confound rare variant association tests. *PLoS ONE* 2013;8:e65834.

67 Cruchaga C, Karch CM, Jin SC, Benitez BA, Cai Y, Guerreiro R, Harari O, Norton J, Budde J, Bertelsen S, Jeng AT, Cooper B, Skorupa T, Carrell D, Levitch D, Hsu S, Choi J, Ryten M, Consortium UKBE, Hardy J, Ryten M, Trabzuni D, Weale ME, Ramasamy A, Smith C, Sassi C, Bras J, Gibbs JR, Hernandez DG, Lupton MK, Powell J, Forabosco P, Ridge PG, Corcoran CD, Tschanz JT, Norton MC, Munger RG, Schmutz C, Leary M, Demirci FY, Bamne MN, Wang X, Lopez OL, Ganguli M, Medway C, Turton J, Lord J, Braae A, Barber I, Brown K, Alzheimer's Research UKCPassmore P, Craig D, Johnston J, McGuinness B, Todd S, Heun R, Kolsch H, Kehoe PG, Hooper NM, Vardy ER, Mann DM, Pickering-Brown S, Brown K, Kalsheker N, Lowe J, Morgan K, David Smith A, Wilcock G, Warden D, Holmes C, Pastor P, Lorenzo-Betancor O, Brkanac Z, Scott E, Topol E, Morgan K, Rogaeva E, Singleton AB, Hardy J, Kamboh MI, St George-Hyslop P, Cairns N, Morris JC, Kauwe JS, Goate AM. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature* 2014;505:550–4.

68 Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, Zhang G, Franceschini N, Fox K, Lange EM, Rich SS, O'Donnell CJ, Jackson RD, Wallace RB, Chen Z, Graubert TA, Wilson JG, Tang H, Lettre G, Reiner AP, Ganesh SK, Li Y. Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in African Americans. NHLBI GO Exome Sequencing Project. *Am J Hum Genet* 2012;91:794–808.

69 Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO, Brody JA, Khetarpal SA, Crosby JR, Fornage M, Isaacs A, Jakobsdottir J, Feitosa MF, Davies G, Huffman JE, Manichaikul A, Davis B, Lohman K, Joon AY, Smith AV, Grove ML, Zanoni P, Redon V, Demissie S, Lawson K, Peters U, Carlson C, Jackson RD, Ryckman KK, Mackey RH, Robinson JG, Siscovick DS, Schreiner PJ, Mychaleckyj JC, Pankow JS, Hofman A, Uitterlinden AG, Harris TB, Taylor KD, Stafford JM, Reynolds LM, Marioni RE, Dehghan A, Franco OH, Patel AP, Lu Y, Hindy G, Gottesman O, Bottinger EP, Melander O, Orho-Melander M, Loos RJ, Duga S, Merlini PA, Farrall M, Goel A, Asselta R, Girelli D, Martinelli N, Shah SH, Kraus WE, Li M, Rader DJ, Reilly MP, McPherson R, Watkins H, Ardissino D, Project NGES, Zhang Q, Wang J, Tsai MY, Taylor HA, Correa A, Griswold ME, Lange LA, Starr JM, Rudan I, Eiriksdottir G, Launer LJ, Ordovas JM, Levy D, Chen YD, Reiner AP, Hayward C, Polasek O, Deary IJ, Borecki IB, Liu Y, Gudnason V, Wilson JG, van Duijn CM, Kooperberg C, Rich SS, Psaty BM, Rotter JI, O'Donnell CJ, Rice K, Boerwinkle E, Kathiresan S, Cupples LA. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet* 2014;94:223–32.

70 Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Lochen ML, Ganesh SK, Vatten L, Skorpen F, Dalen H, Zhang J, Pennathur S, Chen J, Platou C, Mathiesen EB, Wilsgaard T, Njolstad I, Boehnke M, Chen YE, Abecasis GR, Hveem K, Willer CJ. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* 2014;46:345–51.

71 Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjaerg-Hansen A, Vogt TF, Hobbs HH, Cohen JC. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2014;46:352–6.

72 Auer PL, Teumer A, Schick U, O'Shaughnessy A, Lo KS, Chami N, Carlson C, de Denus S, Dube MP, Haessler J, Jackson RD, Kooperberg C, Perreault LP, Nauck M, Peters U, Rioux JD, Schmidt F, Turcot V, Volker U, Volzke H, Greinacher A, Hsu L, Tardif JC, Diaz GA, Reiner AP, Lettre G. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet* 2014;46:629–34.

73 Budworth H, McMurray CT. A brief history of triplet repeat diseases. *Methods Mol Biol* 2013;1010:3–17.

74 Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012;13:36–46.

75 Kirby A, Gnirke A, Jaffe DB, Baresova V, Pochet N, Blumenstiel B, Ye C, Aird D, Stevens C, Robinson JT, Cabili MN, Gat-Viks I, Kelliher E, Daza R, DeFelice M, Hulkova H, Sovova J, Vylet'al P, Antignac C, Guttman M, Handsaker RE, Perrin D, Steelman S, Sigurdsson S, Scheinman SJ, Sougnez C, Cibulskis K, Parkin M, Green T, Rossin E, Zody MC, Xavier RJ, Pollak MR, Alper SL, Lindblad-Toh K, Gabriel S, Hart PS, Regev A, Nusbaum C, Kmoch S, Bleyer AJ, Lander ES, Daly MJ. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet* 2013;45:299–303.

76 Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33:177–82.

77 Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat* 2013;34:E2393–2402.

78 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–15.

79 Weedon MN, Cebola I, Patch AM, Flanagan SE, De Franco E, Caswell R, Rodriguez-Segui SA, Shaw-Smith C, Cho CH, Lango Allen H, Houghton JA, Roth CL, Chen R, Hussain K, Marsh P, Vallier L, Murray A, International Pancreatic Agenesis CEllard S, Ferrer J, Hattersley AT. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* 2014;46:61–4.

80 Zawistowski M, Reppell M, Wegmann D, St Jean PL, Ehm MG, Nelson MR, Novembre J, Zollner S. Analysis of rare variant population structure in Europeans explains differential stratification of gene-based tests. *Eur J Hum Genet* 2014;22:1137–44.

81 Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, Branham KE, Heckenlively J, Study F, Fulton R, Wilson RK, Mardis ER, Lin X, Swaroop A, Zollner S, Abecasis GR. Ancestry estimation and control of population stratification for sequence-based association studies. *Nat Genet* 2014;46:409–15.

82 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;9:356–69.

Rare and low-frequency variants in human common diseases and other complex traits

Guillaume Lettre

J Med Genet published online September 3, 2014
doi: 10.1136/jmedgenet-2014-102437