Review

Genetic determinants of phenotypic diversity in humans

Nazli G Rahim, Olivier Harismendy, Eric J Topol and Kelly A Frazer

Address: Scripps Genomic Medicine, The Scripps Research Institute, North Torrey Pines Road MEM 275, La Jolla, CA 92037, USA.

Correspondence: Kelly A Frazer. Email: kfrazer@scripps.edu

Published: 24 April 2008

*Genome **Biology*** 2008, **9:**215 (doi:10.1186/gb-2008-9-4-215)

The electronic version of this article is the complete one and can be
found online at http://genomebiology.com/2008/9/4/215

Abstract

New technologies for rapidly assaying DNA sequences have revealed that the degree and nature of human genetic variation is far more complex then previously realized. These same technologies have also resulted in the identification of common genetic variants associated with more than 30 human diseases and traits.

Human genetic variation was named “breakthrough of the year” by *Science* in 2007, reflecting the marked advances in understanding the genetic basis of normal human phenotypic diversity and susceptibility to a wide range of diseases. The human genome is composed of 3 billion nucleotides with approximately 0.5% of these nucleotides differing among individuals [1]. This genetic variation, the nucleotides that differ from person to person, affects the majority of human phenotypic differences, from eye color and height to disease susceptibility and responses to drugs.

Classification of genetic variants

Phenotypic variation in humans is a direct consequence of genetic variation, which acts in conjunction with environmental and behavioral factors to produce phenotypic diversity. Genetic variants are classified by two basic criteria: their genetic composition and their frequency in the population. In terms of composition, polymorphisms can be classified as sequence variants or structural variants. Sequence variants range from single nucleotide differences between individuals to 1 kilobase (kb)-sized insertions or deletions (indels) of a segment of DNA (Figure 1) [2]. Larger insertions and deletions, as well as duplications, inversions and translocations, are collectively called structural variants. These variants can range in size from 1 kb to those spanning more than 5 megabases (Mb) of DNA [3].

Genetic variants are also classified in terms of their frequency within the population, with common variants defined as those in which the minor allele is present at a frequency of greater than 5% in the population, while for rare variants it is present at a frequency of less than 5%. The fundamental source of genetic variation is mutation, and the majority of common genetic variants arose once in human history and are shared by many individuals today through descent from common ancient ancestors. A polymorphism is, by convention, defined as a genetic variant that is present in at least 1% of the population and thereby excludes rare variants that may have arisen in relatively recent human history. Much of the study of genetic variation to date has focused on characterizing the 10 million estimated single nucleotide polymorphisms (SNPs), as they comprise approximately 78% of human variants, thus accounting for most genetic diversity. SNPs are located, on average, every 100 to 300 bases in the genome. Structural variants account for only an estimated 22% of all variants in the genome, but they comprise an estimated 74% of the nucleotides that differ between individuals [1]. As a result of technological advances that enable their detection, there has been a flurry of recent efforts to catalogue structural polymorphisms on a genomic scale [4-6].

The study of inheritance of genetic variation depends on two key concepts: genetic linkage and linkage disequilibrium (Figure 2). Two loci are in genetic linkage if they are physically close enough to one another such that recombination occurs between them with a less than 50% probability in a single generation, resulting in their co-segregation more often than if they were independently inherited (Figure 2a,b).

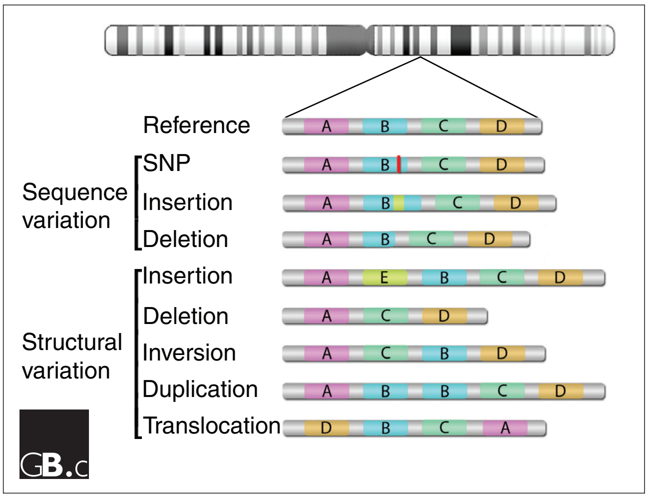

Figure 1
Classification of genetic variants by composition. Schematic of sequence and structural variants compared to reference sequence. Sequence variation (indicated by red line) refers to single-nucleotide variants and small (less than 1 kb) indels. Structural variation includes inversions, translocations and copy-number variants, which result in the presence of a segment of DNA in variable numbers compared to the reference sequence, as in duplications, deletions or insertions. Adapted from [4].

Recombination frequency is measured in units of centimorgans, with 1 centimorgan equal to a 1% chance that two loci will segregate independently due to recombination in a single generation. One centimorgan is, on average, equivalent to 1 million base pairs (bp) in the human genome.

Linkage disequilibrium is a measure of the co-occurrence in a population of a particular allele at one locus with a particular allele at a second locus at a higher frequency than would be predicted by random chance. Linkage disequilibrium is created when a new mutation occurs in a genomic interval that already contains a particular variant allele, and is eroded over the course of many generations by recombination. Various statistics have been used to measure the amount of linkage disequilibrium between two variant alleles, one of the most useful being the coefficient of correlation r^2. When $r^2 = 1$ the two variant alleles are in complete linkage disequilibrium, whereas values of $r^2 < 1$ indicate that the ancestral complete linkage disequilibrium has been eroded. Thus, while genetic linkage results from recombination in the last two to three generations and measures co-segregation in a pedigree, linkage disequilibrium depends on the association of variant alleles within a population of

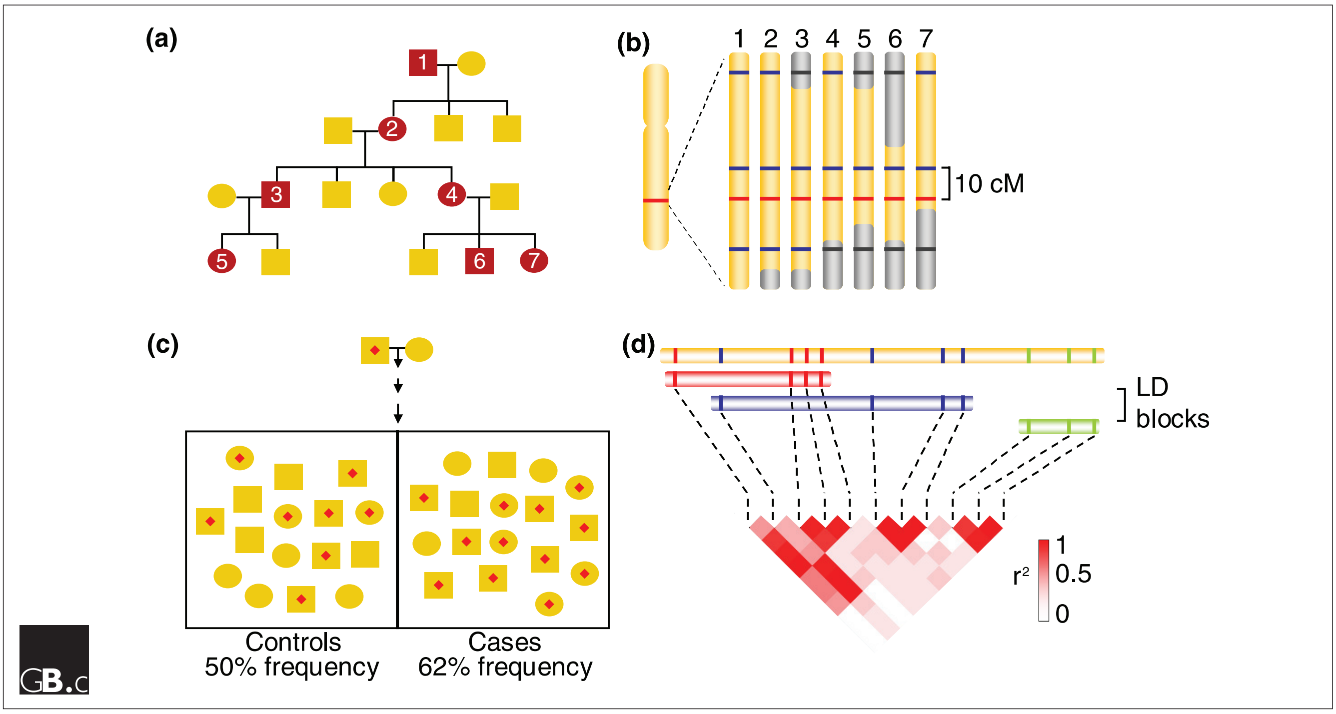

Figure 2
Identification of genetic variation underlying human disease using linkage analysis and genome-wide association studies. **(a)** Rare Mendelian traits, such as a monogenic disease with autosomal dominance inheritance, can be studied using linkage analysis in a family. The disease status is followed within a pedigree (seven affected individuals depicted in red). **(b)** The disease loci (red bar) co-segregates with the genetic marker (blue bar), located 10 centimorgans (cM) apart. Each of the seven individuals with the disease carries the blue genetic marker, both inherited from the affected 'parent' chromosome (yellow). **(c)** Genetic variants underlying common diseases can be statistically identified by using SNP-based linkage disequilibrium (LD) maps. The frequency of a causative variant (red diamond) will be higher (62%) among those with the disease when compared with a control population (50%). **(d)** LD map of 11 variants cluster into three blocks of correlation $r^2 > 0.8$ (red scale correlation matrix). The LD between polymorphisms needs to be empirically determined by genotyping a population and calculating the correlation.

unrelated individuals and reflects evolutionary history (Figure 2c,d).

Advances in identification of genetic variants underlying human traits

The first disease traits to be ascribed to particular genes were Mendelian traits, which are controlled by a single gene and follow well defined models of inheritance, such as autosomal dominant, autosomal recessive, and X-linked (Figure 2a). Genetic variants underlying Mendelian diseases are highly penetrant by definition (that is, the variant is associated with a very high relative risk of having the disease) and, as a result of negative selection, they tend to be rare (Figure 3).

In the 1980s and 1990s, the creation of genetic-linkage maps was based on sequence-dependent data such as restriction-fragment length polymorphisms [7,8] and microsatellite markers [9]. These techniques established genetic-linkage analysis as the traditional method for identifying genetic variation underlying monogenic genetic disorders. Linkage studies consisted of mapping broad genetic regions that segregate with a disease in families and then using positional cloning to narrow down the candidate region in order to isolate disease-causing genes or variants. Linkage analyses were successful in identifying genetic variants in genes responsible for many notable Mendelian diseases, including cystic fibrosis [10], for which the major disease variant has a deletion of a single amino acid, Charcot-Marie-Tooth Disease Type 1A [11], for which the underlying genetic variant is a DNA duplication, and Huntington's disease [12], which is a trinucleotide repeat disorder. By 1995, genetic linkage mapping had been used to uncover variants underlying hundreds of human Mendelian traits and diseases. Thus, almost a decade before the elucidation of the human genome sequence, it was fully appreciated that DNA variants of all classes, both common and rare as well as sequence and structural, play important roles in single-gene traits and rare Mendelian diseases.

The next, and more difficult, stage was to determine genes associated with the far more common complex (multigene) diseases such as diabetes, heart disease and cancer. The conceptual framework for statistical association studies to identify common genetic variants underlying common diseases was established by Risch and Merikangas in 1996 [13], and is now referred to as the common disease/common variant (CD/CV) hypothesis. This hypothesis states that common diseases are caused by multiple genetic variants that are present at a high frequency in the population and confer cumulative incremental effects on disease risk (Figure 3) [14,15]. It is thought that due to the low penetrance and modest risk associated with these common variant alleles, they do not undergo the same strong negative selection as highly penetrant rare variants underlying Mendelian diseases. In addition, environment and behavior are believed to contribute over 70% of the susceptibility to diseases such as cancer, coronary heart disease and type 2 diabetes [16]. On the basis of these assumptions in the CD/CV model, it was posited that to identify variant that occur at a high frequency in the population yet confer a small risk for disease, it would be feasible to use SNP-based linkage disequilibrium maps to survey the common genetic variation present in the entire genomes of a large number of individuals.

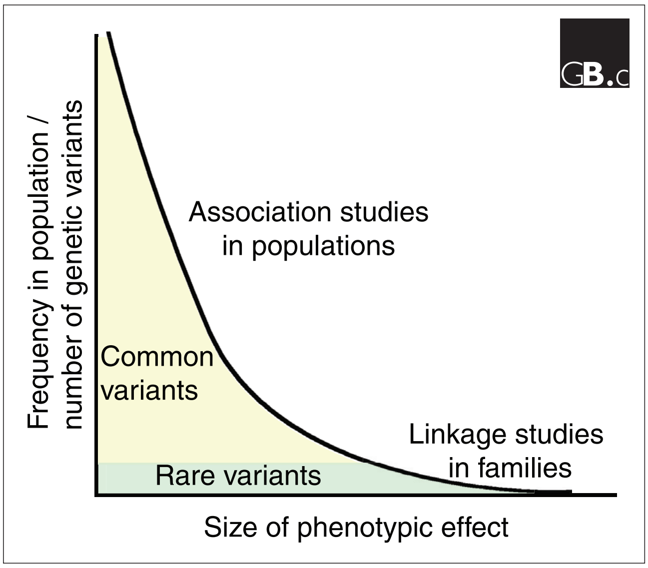

Figure 3
The allelic spectrum of disease is dependent on the number of genetic variants, their frequency in a population and on the size of their phenotypic effect. Family-based linkage studies have proved successful in identifying causative genetic variants in rare Mendelian disorders, which are, by definition, caused by highly penetrant variants that have a low frequency in the population. Complex diseases are caused by multiple genetic variants that confer incremental risk of disease. Genome-wide association studies have sufficient power to detect genetic variants with modest phenotypic effects, provided that they occur at a high frequency in the population. Adapted from [92].

Several key technological advances laid the foundation for the eventual successful implementation of genome-wide association studies in identifying common genetic variants underlying complex traits. The first was the completion of the 3 billion bp human genome sequence in 2001, which served as a reference sequence to which genotype or sequence information from individuals could be compared [17,18]. Then, large-scale efforts led to the discovery of a substantial fraction of the 10 million estimated SNPs in the human population. By genotyping millions of these SNPs in hundreds of individuals, the International HapMap Project created SNP linkage disequilibrium maps, reducing the vast majority of common genetic variation in the 3 billion bp human genome to around 500,000 tag SNPs that are proxies for other SNPs in high linkage disequilibrium [19]. This resource has driven a wave of critical technological advances in the design of genome-wide SNP arrays that allow the rapid and cost-effective genotyping of hundreds of thousands to millions of tag SNPs in each individual, thus allowing the examination of common genetic variation across the genome.

Genome-wide association studies using SNP-based arrays compare the frequency of SNP alleles in the genomes of a group of individuals with a complex trait (the cases) to a control group (Figure 2c). This approach allows the identification of common genetic variants that are either causative or in linkage disequilibrium with a causative allele. In reviewing the design of successful genome-wide association studies, three key features become clear. First, because of the moderate risk conferred by many common genetic variants, it is imperative to design an adequately powered study with large sample sizes that are carefully controlled to minimize bias [20-22]. Second, SNP selection and detection is critical, and there is an ongoing effort to catalog more SNPs across the genome and to create methods to assay SNP genotypes more densely. Finally, even statistically convincing associations require validation by replication in an independent cohort.

Identifying genetic variants underlying complex (multigene) traits

During 2007, the first wave of genome-wide association studies using tag SNPs resulted in the identification of common genetic variants associated with a broad range of common diseases and traits, including cancer, metabolic diseases, immune-mediated diseases and neurodegenerative diseases (Table 1). The findings of these genome-wide scans can best be reviewed by discussing the results of studies investigating specific complex diseases and traits. Gout and its associated serum uric acid concentration has been studied in two genome-wide association studies [23,24], resulting in the identification of variants in the gene *SLC2A9* (solute carrier family 2 member 9). *SLC2A9* variants were associated with high concentration of uric acid in the serum (between 1.7% and 5.3% increase) and the expression level of the isoform 2 of *SLC2A9* was correlated with serum uric acid concentration [24]. This isoform encodes the protein Glut9ΔN, a putative fructose transporter expressed in kidney. As fructose is upstream in the pathway generating uric acid, an impaired expression of this protein possibly leads to the increased level of serum uric acid observed in gout [23,24].

Multiple genome-wide association studies investigating coronary artery disease have independently identified a strong association with SNPs in a chromosomal region at 9p21. Individuals homozygous for the 9p21 risk allele have a 1.9 higher relative risk of suffering from coronary artery disease than individuals homozygous for the non-risk alleles [22,25-28]. Interestingly, this region does not harbor any known genes, and the underlying biological reason for the association is unknown. Beyond diseases, genome-wide scans have identified variants associated with human height: *HMG2A* (a transcription factor) and *GDF5-UQCC* (a locus associated with osteoarthritis) [29,30]. In addition, variants in *FTO* (fat mass and obesity associated gene) have been associated with obesity: adults homozygous for the risk allele have an increased relative risk of 1.67 for being obese compared with the non-risk allele carriers [31].

In spite of the exciting successes of recent SNP-based genome scans, the results of studies investigating specific complex diseases indicate that the approach frequently identifies common variants that account for only a small fraction (less than 10%) of the heritable component of the disease [32]. Most of the associated SNPs typically result in an increased relative risk of around 1.2 for heterozygotes and for many diseases only a few SNPs have been identified. Thus, we are left asking where is the remaining genetic variance underlying these heritable diseases? It is likely that some of this missing variation is accounted for by common variants with very small effects, which the current studies, despite the rather large cohorts used, are not powerful enough to capture. The additive or even multiplicative integrated effect of common SNPs may be important, as recently shown with five SNPs that increase susceptibility to prostate cancer [33]. Such gene-gene interactions are typically not accounted for in the analysis of genome scans. It is well established that SNP-based genome scans have limited power to capture the association of rare variants, which are likely to be important contributors to complex diseases. Structural variants have been demonstrated to underlie phenotypic diversity of complex traits [34,35] but have not generally been captured with current SNP-centric platforms for ultra-high throughput genotyping. Recent studies have shown that this class of variants is enriched in segmentally duplicated regions of the genome, in which there is a paucity of tag SNPs because of technical difficulties [36]. Thus, the missing variation in SNP-based genome scans indicates that systematically examining these other types of variants for their contribution to complex diseases is important.

Functional annotation of genetic variants

Although the discoveries of SNP-based genome-wide association studies are exciting, it is important to note that they are limited to the statistical association of DNA variants with common diseases and that the biological mechanisms underlying most of these findings are not yet known. For example, multiple studies have shown that three SNPs on chromosome 16p13 in the vicinity of *KIAA0350* are unequivocally associated with type 1 diabetes, but it is unclear how the risk and non-risk alleles differ; is it in expression, alternative splicing patterns, or the function of the protein encoded by *KIAA0350*? [37] This uncertainty in

Table 1

Genetic loci associated with disease and phenotypic variation

Disease type	Disease	Associated loci	Date of publication	Reference
Cancer	Acute lymphoblastic leukemia	PAX5 and others	12 April 2007	[40]
	Breast cancer	FGFR2, TNCR9, MAP3K1, LSP and others	27 May 2007	[41]
	Colon, prostate cancer	8q24	8 July 2007	[42-44]
	Colorectal cancer	SMAD7	14 October 2007	[45]
		CRAC1 (HMPS)	16 December 2007	[46]
	Multiple solid tumors	CASP8	22 April 2007	[47]
	Prostate cancer	8q24	1 April 2007	[48,49]
		TCF2; 17p	1 July 2007	[50]
		2p15, Xp11.22 and multiple others	10 February 2008	[51-53]
Heart	Myocardial infarction, coronary artery disease, intracranial aneurysm	9p21	6 January 2008	[26]
		9p21	3 May 2007	[26,54]
		6q25, 2q36	18 July 2007	[28]
	Atrial fibrillation	4q25	1 July 2007	[55]
Metabolic	Celiac disease	IL-2, IL-21	10 June 2007	[56]
	Diabetes, type 1	12q24 and others	6 June 2007	[57]
		KIAA0350	15 July 2007	[37]
		IL2RA	5 August 2007	[58]
	Diabetes, type 2	CDKAL1 and six others	26 April 2007	[59,60]
		WFS1	1 July 2007	[61]
	Gout	SLCA9	9 March 2008	[23,24]
	Hypercholesterolemia	CELSR2	9 February 2008	[62]
	Lipoprotein disorders	MLX1PL and multiple others	13 January 2008	[32,62,63]
	Obesity	FTO	12 April 2007	[31]
Neurodegenerative	Amyotrophic lateral sclerosis	FLJ10986	1 August 2007	[64]
		DPP6	16 December 2007	[65]
	Multiple sclerosis	IL7Rα, IL2Rα	28 July 2007	[66-68]
Immune mediated	Ankylosing spondylitis	ARTS1, IL23R	21 October 2007	[69]
	Autoimmune thyroid disease	TSHR, FCRL3	21 October 2007	[69]
	Rheumatoid arthritis	6p21, 1p13	7 June 2007	[22]
		TRAF1-C5	31 August 2007	[70]
		6q23	4 Nov 2007	[71,72]
	Systemic lupus erythematosus	TNFSF4	2 December 2007	[73]
		PXK, KIAA1542, BANK1, C8orf-BLK, ITGAM	20 January 2008	[74-77]
	Age-related macular degeneration	C3	18 July 2007	[78]
	Celiac disease	IL-2, IL-21	10 June 2007	[56]
Other	Asthma (childhood)	ORMDL3	4 July 2007	[79]
	Bipolar disorder	16p12	7 June 2007	[22]
	Crohn's disease	IRGM	6 June 2007	[80]
		ILR23	26 October 2006	[81]
		IBD5	October 2001	[82]
		ATG16L1	15 April 2007	[83,84]
		5p13.1	5 March 2007	[85]
		NOD2	16 June 2001	[86]
	Gallstone disease	ABCG8	15 July 2007	[87]
	Glaucoma	LOXL1	9 August 2007	[88]
	HIV host control	HLA-B*5701	19 July 2007	[89]
	Psoriasis	β-Defensin, CNV	2 December 2007	[90]
	Restless leg syndrome	MEIS1, BTBD9, MAP2K5	18 July 2007	[91]

the underlying biological cause of an association is especially pronounced when the variant lies in a chromosomal interval that does not contain a gene, such as the association of the 9p21 interval with coronary artery disease. Therefore, the findings of most association studies currently can only be used for crude predictions of the likelihood that an individual will develop a certain disease.

To translate the findings of SNP-based genome scans into clinical practice to improve human health, it is necessary to establish new, highly innovative approaches for assaying intervals containing associated variants for functional differences between the risk and non-risk alleles. This will require access to diverse and large patient populations to obtain biological samples. Each genomic interval has a different landscape of functional sequences, and this, together with the fact that each disease affects different biological processes, makes it impossible to develop a 'one-size-fits-all' strategy to annotate associated sequences for functional differences between risk and non-risk alleles. Thus, it is also essential to make use of diverse experimental methods and technologies in all the various biological 'omics': genomics, proteomics, epigenomics, metabolomics, structural genomics and glycomics.

Several public and private initiatives are developing 'next generation' sequencing technologies based on pyrosequencing (Roche-454) [38], sequencing by synthesis (Illumina-Solexa) [39] or sequencing by ligation (ABI-SOLiD). These technologies, capable of the cost-effective generation of massive amounts of DNA sequence, are already being used to sequence targeted regions, and in the near future will be capable of sequencing whole genomes of individuals to simultaneously examine SNPs and other genetic variants for associations with specific diseases. The statistical analysis methods for assessing the relationship between rare genetic variants identified in sequence data and complex traits are beginning to be developed. Results of sequence-based studies conducted so far suggest that associated intervals will be identified on the basis that the frequency of rare genetic variants with functional consequences will be greater in individuals with the complex disease versus controls. Thus, next-generation sequencing technologies, by detecting a myriad more SNPs and other types of variation associated with complex disease, will increase the difficulty and at the same time, the importance of functional annotation of genetic variants. At this point, it appears that we are just beginning to appreciate the extent of human genomic variation. Projects like the '1000 Genomes' and large-scale efforts to perform deep-coverage sequencing in both healthy patients and those with complex traits will help propel this exciting field further.

References

1. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, *et al.*: **The diploid genome sequence of an individual human.** *PLoS Biol* 2007, **5:**e254.
2. Feuk L, Carson AR, Scherer SW: **Structural variation in the human genome.** *Nat Rev Genet* 2006, **7:**85-97.
3. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36:**949-951.
4. Estivill X, Armengol L: **Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies.** *PLoS Genet* 2007, **3:**1787-1799.
5. Abecasis G, Tam PK-H, Bustamante CD, Ostrander EA, Scherer SW, Chanock SJ, Kwok P-Y, Brookes AJ: **Human genome variation 2006: emerging views on structural variation and large-scale SNP analysis.** *Nat Genet* 2007, **39:**153-155.
6. Sharp AJ, Cheng Z, Eichler EE: **Structural variation of the human genome.** *Annu Rev Genomics Hum Genet* 2006, **7:**407-442.
7. Botstein D, White RL, Skolnick M, Davis RW: **Construction of a genetic linkage map in man using restriction fragment length polymorphisms.** *Am J Hum Genet* 1980, **32:**314-331.
8. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, Botstein D, Akots G, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Powers JA, Watt DE, Kauffman ER, Bricker A, Phipps P, Muller-Kahle H, Fulton TR, Ng S, Schumm JW, Braman JC, Knowlton RG, Barker DF, Crooks SM, *et al.*: **A genetic linkage map of the human genome.** *Cell* 1987, **51:**319-337.
9. Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M: **A second-generation linkage map of the human genome.** *Nature* 1992, **359:**794-801.
10. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, Drumm ML, Lannuzzi MC, Collins FS, Tsui LC: **Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.** *Science* 1989, **245:**1066-1073.
11. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, Saucedo-Cardenas O, Barker DF, Killian JM, Garcia CA, Chakravarti A, Patel PI: **DNA duplication associated with Charcot-Marie-Tooth disease type 1A.** *Cell* 1991, **66:**219-232.
12. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Young AB, Shoulson I, Bonilla E, Martin JB: **A polymorphic DNA marker genetically linked to Huntington's disease.** *Nature* 1983, **306:** 234-238.
13. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273:**1516-1517.
14. Lander ES: **The new genomics: global views of biology.** *Science* 1996, **274:**536-539.
15. Reich DE, Lander ES: **On the allelic spectrum of human disease.** *Trends Genet* 2001, **17:**502-510.
16. Willett W: **Balancing life-style and genomics research for disease prevention**. *Science* 2002, **296:**695-698.
17. International Human Genome Sequencing Consortium: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
18. 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, *et al*: **The sequence of the human genome.** *Science* 2001, **291:**1304-1351.
19. International HapMap Consortium, 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, *et al.*: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449:**851-861.
20. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D: **Assessing the impact of population stratification on genetic association studies.** *Nat Genet* 2004, **36:**388-393.
21. Marchini J, Cardon LR, Phillips MS, Donnelly P: **The effects of human population structure on large genetic association studies.** *Nat Genet* 2004, **36:**512-517.

22. Wellcome Trust Case Control Consortium: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447:**661-678.
23. Döring A, Gieger C, Mehta D, Gohlke H, Prokisch H, Coassin S, Fischer G, Henke K, Klopp N, Kronenberg F, Paulweber B, Pfeufer A, Rosskopf D, Völzke H, Illig T, Meitinger T, Wichmann HE, Meisinger C: **SLC2A9 influences uric acid concentrations with pronounced sex-specific effects.** *Nat Genet* 2008, **40:**430-436.
24. Vitart V, Rudan I, Hayward C, Gray NK, Floyd J, Palmer CN, Knott SA, Kolcic I, Polasek O, Graessler J, Wilson JF, Marinaki A, Riches PL, Shu X, Janicijevic B, Smolej-Narancic N, Gorgoni B, Morgan J, Campbell S, Biloglav Z, Barac-Lauc L, Pericic M, Klaric IM, Zgaga L, Skaric-Juric T, Wild SH, Richardson WA, Hohenstein P, Kimber CH, Tenesa A, *et al.*: **SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout.** *Nat Genet* 2008, **40:**437-442.
25. Helgadottir A, Thorleifsson G, Magnusson KP, Grétarsdottir S, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankensteijn JD, Ronkainen A, Jääskeläinen JE, Kyo Y, Lenk GM, Sakalihasan N, Kostulas K, Gottsäter A, Flex A, Stefansson H, Hansen T, Andersen G, Weinsheimer S, Borch-Johnsen K, Jorgensen T, Shah SH, Quyyumi AA, Granger CB, Reilly MP, Austin H, Levey AI, Vaccarino V, *et al.*: **The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm.** *Nat Genet* 2008, **40:**217-224.
26. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, *et al.*: **A common variant on chromosome 9p21 affects the risk of myocardial infarction.** *Science* 2007, **316:**1491-1493.
27. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC: **A common allele on chromosome 9 associated with coronary heart disease.** *Science* 2007, **316:**1488-1491.
28. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, *et al.*; WTCCC and the Cardiogenics Consortium: **Genomewide association analysis of coronary artery disease.** *N Engl J Med* 2007, **357:**443-453.
29. Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, Shen H, Timpson N, Lettre G, Usala G, Chines PS, Stringham HM, Scott LJ, Dei M, Lai S, Albai G, Crisponi L, Naitza S, Doheny KF, Pugh EW, Ben-Shlomo Y, Ebrahim S, Lawlor DA, Bergman RN, Watanabe RM, Uda M, Tuomilehto J, Coresh J, Hirschhorn JN, *et al.*: **Common variants in the GDF5-UQCC region are associated with variation in human height.** *Nat Genet* 2008, **40:**198-203.
30. Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JR, Elliott KS, Hackett R, Guiducci C, Shields B, Zeggini E, Lango H, Lyssenko V, Timpson NJ, Burtt NP, Rayner NW, Saxena R, Ardlie K, Tobias JH, Ness AR, Ring SM, Palmer CN, Morris AD, Peltonen L, Salomaa V; Diabetes Genetics Initiative; Wellcome Trust Case Control Consortium, Davey Smith G, Groop LC, Hattersley AT, McCarthy MI, Hirschhorn JN, Frayling TM: **A common variant of HMGA2 is associated with adult and childhood height in the general population.** *Nat Genet* 2007, **39:**1245-1250.
31. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, *et al.*: **A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.** *Science* 2007, **316:**889-894.
32. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, *et al.*: **Newly identified loci that influence lipid concentrations and risk of coronary artery disease.** *Nat Genet* 2008, **40:**161-169.
33. Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, Bälter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang BL, Isaacs WB, Xu J, Grönberg H: **Cumulative association of five genetic variants with prostate cancer.** *New Engl J Med* 2008, **358:**910-919.
34. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC: **Diet and the evolution of human amylase gene copy number variation.** *Nat Genet* 2007, **39:**1256-1260.
35. Schaeffeler E, Schwab M, Eichelbaum M, Zanger UM: **CYP2D6 genotyping strategy based on gene copy number determination by TaqMan real-time PCR.** *Hum Mutat* 2003, **22:**476-485.
36. McCarroll SA, Altshulter D: **Copy-number variation and association studies of human disease.** *Nat Genet* 2007, **39:**S37-S42.
37. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C: **A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene.** *Nature* 2007, **448:**591-594.
38. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, *et al.*: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437:**376-380.
39. Bennett S: **Solexa Ltd.** *Pharmacogenomics* 2004, **5:**433-438.
40. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, Relling MV, Evans WE, Shurtleff SA, Downing JR: **Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia.** *Nature* 2007, **446:**758-764.
41. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R; SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, *et al*: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **447:**1087-1093.
42. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K; CORGI Consortium, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, *et al.*: **A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21.** *Nat Genet* 2007, **39:**984-988.
43. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, *et al.*: **Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24.** *Nat Genet* 2007, **39:**989-994.
44. Haiman CA, Le Marchand L, Yamamato J, Stram DO, Sheng X, Kolonel LN, Wu AH, Reich D, Henderson BE: **A common genetic risk factor for colorectal and prostate cancer.** *Nat Genet* 2007, **39:** 954-956.
45. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O; CORGI Consortium, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I, Houlston RS: **A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk.** *Nat Genet* 2007, **39:**1315-1317.
46. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z, Sullivan K, Heinimann K, Lubbe S, Domingo E, Barclay E, Martin L, Gorman M, Chandler I, Vijayakrishnan J, Wood W, Papaemmanuil E, Penegar S, Qureshi M; CORGI Consortium, Farrington S, Tenesa A, Cazier JB, Kerr D, Gray R, Peto J, Dunlop M, *et al.*: **Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk.** *Nat Genet* 2008, **40:**26-28.

47. Sun T, Gao Y, Tan W, Ma S, Shi Y, Yao J, Guo Y, Yang M, Zhang X, Zhang Q, Zeng C, Lin D: **A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers.** *Nat Genet* 2007, **39:**605-613.
48. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, *et al.*: **Genome-wide association study of prostate cancer identifies a second risk locus at 8q24.** *Nat Genet* 2007, **39:**645-649.
49. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D: **Multiple regions within 8q24 independently affect risk for prostate cancer.** *Nat Genet* 2007, **39:**638-644.
50. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, *et al.*: **Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes.** *Nat Genet* 2007, **39:**977-983.
51. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, Kristinsson KT, Birgisdottir B, Ghosh S, Magnusdottir DN, Thorlacius S, Thorleifsson G, Zheng SL, Sun J, Chang BL, Elmore JB, Breyer JP, McReynolds KM, Bradley KM, Yaspan BL, Wiklund F, Stattin P, Lindström S, *et al.*: **Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer.** *Nat Genet* 2008, **40:**281-283.
52. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, *et al.*: **Multiple newly identified loci associated with prostate cancer susceptibility.** *Nat Genet* 2008, **40:**316-321.
53. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cussenot O, Valeri A, *et al.*: **Multiple loci identified in a genome-wide association study of prostate cancer.** *Nat Genet* 2008, **40:**310-315.
54. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC: **A common allele on chromosome 9 associated with coronary heart disease.** *Science* 2007, **316:**1488-1491.
55. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, *et al.*: **Variants conferring risk of atrial fibrillation on chromosome 4q25.** *Nature* 2007, **448:** 353-357.
56. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, Feighery C, Jewell D, Kelleher D, Kumar P, Travis S, Walters JR, Sanders DS, Howdle P, Swift J, Playford RJ, McLaren WM, Mearin ML, Mulder CJ, McManus R, McGinnis R, Cardon LR, Deloukas P, Wijmenga C: **A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21.** *Nat Genet* 2007, **39:**827-829.
57. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, *et al.*: **Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.** *Nat Genet* 2007, **39:**857-864.
58. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, Clayton DG, Wicker LS, Todd JA: **Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes.** *Nat Genet* 2007, **39:**1074-1082.
59. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, *et al.*: **A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.** *Science* 2007, **316:**1341-1345.
60. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Boström K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Råstam L, Speliotes EK, Taskinen MR, *et al.*: **Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.** *Science* 2007, **316:**1331-1336.
61. Sandhu MS, Weedon MN, Fawcett KA, Wasson J, Debenham SL, Daly A, Lango H, Frayling TM, Neumann RJ, Sherva R, Blech I, Pharoah PD, Palmer CN, Kimber C, Tavendale R, Morris AD, McCarthy MI, Walker M, Hitman G, Glaser B, Permutt MA, Hattersley AT, Wareham NJ, Barroso I: **Common variants in WFS1 confer risk of type 2 diabetes.** *Nat Genet* 2007, **39:**951-953.
62. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M: **Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans.** *Nat Genet* 2008, **40:**189-197.
63. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, Gómez Pérez FJ, Frazer KA, Elliott P, Scott J, Milos PM, Cox DR, Thompson JF: **Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides.** *Nat Genet* 2008, **40:**149-151.
64. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, Halperin RF, Stamper C, Jensen KR, Letizia D, Hesterlee SE, Pestronk A, Levine T, Bertorini T, Graves MC, Mozaffar T, Jackson CE, Bosch P, McVey A, Dick A, Barohn R, Lomen-Hoerth C, Rosenfeld J, O'connor DT, Zhang K, Crook R, Ryberg H, Hutton M, Katz J, Simpson EP, *et al.*: **Whole-genome analysis of sporadic amyotrophic lateral sclerosis.** *N Engl J Med* 2007, **357:**775-788.
65. van Es MA, van Vught PW, Blauw HM, Franke L, Saris CG, Van den Bosch L, de Jong SW, de Jong V, Baas F, van't Slot R, Lemmens R, Schelhaas HJ, Birve A, Sleegers K, Van Broeckhoven C, Schymick JC, Traynor BJ, Wokke JH, Wijmenga C, Robberecht W, Andersen PM, Veldink JH, Ophoff RA, van den Berg LH: **Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis.** *Nat Genet* 2008, **40:**29-31.
66. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL: **Risk alleles for multiple sclerosis identified by a genomewide study.** *N Engl J Med* 2007, **357:**851-862.
67. Gregory SG, Schmidt S, Seth P, Oksenberg JR, Hart J, Prokop A, Caillier SJ, Ban M, Goris A, Barcellos LF, Lincoln R, McCauley JL, Sawcer SJ, Compston DA, Dubois B, Hauser SL, Garcia-Blanco MA, Pericak-Vance MA, Haines JL, Multiple Sclerosis Genetics Group: **Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis.** *Nat Genet* 2007, **39:**1083-1091.
68. Lundmark F, Duvefelt K, Iacobaeus E, Kockum I, Wallström E, Khademi M, Oturai A, Ryder LP, Saarela J, Harbo HF, Celius EG, Salter H, Olsson T, Hillert J: **Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis.** *Nat Genet* 2007, **39:**1108-1113.
69. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Barrett JC, Davison D, Easton D,

Evans DM, Leung HT, Marchini JL, Morris AP, Spencer CC, Tobin MD, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, *et al.*: **Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants.** *Nat Genet* 2007, **39:**1329-1337.

70. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LR, Li W, Tan AK, Bonnard C, Ong RT, Thalamuthu A, Pettersson S, Liu C, Tian C, Chen WV, Carulli JP, Beckman EM, Altshuler D, Alfredsson L, Criswell LA, Amos CI, Seldin MF, Kastner DL, Klareskog L, Gregersen PK: **TRAF1-C5 as a risk locus for rheumatoid arthritis - a genomewide study**. *New Engl Med* 2007, **357:**1199-1209.

71. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PL, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M, Parkin M, Barry R, Winslow W, Healy C, Graham RR, Neale BM, Izmailova E, Roubenoff R, Parker AN, Glass R, Karlson EW, Maher N, Hafler DA, Lee DM, Seldin MF, Remmers EF, Lee AT, Padyukov L, Alfredsson L, Coblyn J, *et al.*: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis.** *Nat Genet* 2007, **39:**1477-1482.

72. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, Donn R, Symmons D, Hider S, Bruce IN, Wellcome Trust Case Control Consortium, Wilson AG, Marinou I, Morgan A, Emery P, YEAR Consortium, Carter A, Steer S, Hocking L, Reid DM, Wordsworth P, Harrison P, Strachan D, Worthington J: **Rheumatoid arthritis association at 6q23.** *Nat Genet* 2007, **39:**1431-1433.

73. Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, Gaffney PM, Moser KL, Rioux JD, Altshuler D, Behrens TW, Vyse TJ: **Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus.** *Nat Genet* 2008, **40:**83-89.

74. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jönsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrup H, González-Escribano MF, Martin J, Abderrahim H, Alarcón-Riquelme ME: **Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40:**211-216.

75. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, Chen W, Zhu C, McEver RP, Kimberly RP, Alarcón-Riquelme ME, Vyse TJ, Li QZ, Wakeland EK, Merrill JT, James JA, Kaufman KM, Guthridge JM, Harley JB: **A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40:**152-154.

76. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönnblom L, Syvänen AC, Criswell LA, Gregersen PK, Behrens TW: **Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX.** *N Engl J Med* 2008, **358:**900-909.

77. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, *et al.*: **Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci.** *Nat Genet* 2008, **40:**204-210.

78. Yates JR, Sepp T, Matharu BK, Khan JC, Thurlby DA, Shahid H, Clayton DG, Hayward C, Morgan J, Wright AF, Armbrecht AM, Dhillon B, Deary IJ, Redmond E, Bird AC, Moore AT, Genetic Factors in AMD Study Group: **Complement C3 variant and the risk of age-related macular degeneration.** *N Engl J Med* 2007, **357:**553-561.

79. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E, Heinzmann A, Simma B, Frischer T, Willis-Owen SA, Wong KC, Illig T, Vogelberg C, Weiland SK, von Mutius E, Abecasis GR, Farrall M, Gut IG, Lathrop GM, Cookson WO: **Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma.** *Nature* 2007, **448:**470-473.

80. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, Burke DA, Todhunter CE, Ahmad T, Onnie CM, McArdle W, Strachan D, Bethel G, Bryan C, Lewis CM, Deloukas P, Forbes A, Sanderson J, Jewell DP, Satsangi J, Mansfield JC, Wellcome Trust Case Control Consortium, Cardon L, Mathew CG: **Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility.** *Nat Genet* 2007, **39:**830-832.

81. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AJ, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH: **A genome-wide association study identifies IL23R as an inflammatory bowel disease gene.** *Science* 2006, **314:**1461-1463.

82. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, O'Leary S, Winchester E, Dewar K, Green T, Stone V, Chow C, Cohen A, Langelier D, Lapointe G, Gaudet D, Faith J, Branco N, Bull SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, Hudson TJ: **Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease.** *Nat Genet* 2001, **29:**223-228.

83. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR: **Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis.** *Nat Genet* 2007, **39:**596-604.

84. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega FM, Briggs J, Günther S, Prescott NJ, Onnie CM, Häsler R, Sipos B, Fölsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S: **A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1.** *Nat Genet* 2007, **39:**207-211.

85. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, Demarche B, Gut I, Heath S, Foglio M, Liang L, Laukens D, Mni M, Zelenika D, Van Gossum A, Rutgeerts P, Belaiche J, Lathrop M, Georges M: **Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4.** *PLoS Genet* 2007, **3:**e58.

86. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPherson AJ, Bridger S, van Deventer S, Forbes A, Nikolaus S, Lennard-Jones JE, Foelsch UR, Krawczak M, Lewis C, Schreiber S, Mathew CG: **Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations.** *Lancet* 2001, **357:**1925-1928.

87. Buch S, Schafmayer C, Völzke H, Becker C, Franke A, von Eller-Eberstein H, Kluck C, Bässmann I, Brosch M, Lammert F, Miquel JF, Nervi F, Wittig M, Rosskopf D, Timm B, Höll C, Seeger M, ElSharawy A, Lu T, Egberts J, Fändrich F, Fölsch UR, Krawczak M, Schreiber S, Nürnberg P, Tepel J, Hampe J: **A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease.** *Nat Genet* 2007, **39:**995-999.

88. Thorleifsson G, Magnusson KP, Sulem P, Walters GB, Gudbjartsson DF, Stefansson H, Jonsson T, Jonasdottir A, Jonasdottir A, Stefansdottir G, Masson G, Hardarson GA, Petursson H, Arnarsson A, Motallebipour M, Wallerman O, Wadelius C, Gulcher JR, Thorsteinsdottir U, Kong A, Jonasson F, Stefansson K: **Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma.** *Science* 2007, **317:**1397-1400.

89. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB: **A whole-genome association study of major determinants for host control of HIV-1.** *Science* 2007, **317:**944-947.

90. Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JA, Schalkwijk J: **Psoriasis is associated with increased beta-defensin genomic copy number.** *Nat Genet* 2008, **40:**23-25.

91. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, Fulda S, Pütz B, Eckstein G, Hauk S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Peglau I, Eisensehr I, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann HE, Holsboer F, Müller-Myhsok B, Meitinger T: **Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions.** *Nat Genet* 2007, **39:**1000-1006.

92. Petretto E, Liu ET, Aitman TJ: **A gene harvest revealing the archeology and complexity of human disease.** *Nat Genet* 2007, **39:**1299-1301.