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
© 2008 BioMed Central Ltd

Abstract

New technologies for rapidly assaying DNA sequences have revealed that the degree and nature of human genetic variation is far more complex than 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).
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 1](http://genomicbiology.com/2008/9/4/215)  
**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].

![Figure 2](http://genomicbiology.com/2008/9/4/215)  
**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-segregate 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.

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

| 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]      |
|                           | Multiple solid tumors            | CRAC1 (HMPS)                                        | 16 December 2007    | [46]      |
|                           | Prostate cancer                  | 8q24                                                 | 22 April 2007       | [47]      |
|                           |                                  | TCF2; 17p, 2p15, Xp11.22 and multiple others       | 1 April 2007        | [48-49]   |
|                           |                                  | 9p21                                                 | 6 January 2008      | [26]      |
| **Heart**                 | Myocardial infarction, coronary   |                                                     | 9p21, 6q25, 2q36    | [26,54]   |
|                           | artery disease, intracranial     |                                                      | 3 May 2007          | [28]      |
|                           | aneurysm                         |                                                      | 18 July 2007        | [55]      |
| **Metabolic**             | Acute liver failure              | 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                 | CDKL1 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               | IL7RAX, IL2RX                                        | 28 July 2007        | [66-68]   |
| **Immune mediated**       | Ankylosing spondylitis           | ARTS1, IL23R                                         | 21 October 2007     | [69]      |
|                           | Autoimmune thyroid disease       | TSHR, FCR5                                           | 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]      |
|                           |                                  | PKX, KIAA1542, BANK1, C8orf-BLK, ITGAM              | 20 January 2008     | [74-77]   |
| **Age-related macular**   | C3                               |                                                      | 18 July 2007        | [78]      |
| degeneration              |                                  |                                                      | 10 June 2007        | [56]      |
| **Celiac disease**        |                                  |                                                      | 10 June 2007        | [56]      |
| **Asthma (childhood)**    |                                  |                                                      | 10 June 2007        | [56]      |
| **Bipolar disorder**      |                                  |                                                      | 10 June 2007        | [56]      |
| **Crohn’s disease**       |                                  |                                                      | 10 June 2007        | [56]      |
| **Gallstone disease**     |                                  |                                                      | 10 June 2007        | [56]      |
| **Glaucoma**              |                                  |                                                      | 10 June 2007        | [56]      |
| **HIV host control**      |                                  |                                                      | 10 June 2007        | [56]      |
| **Psoriasis**             |                                  |                                                      | 10 June 2007        | [56]      |
| **Restless leg syndrome** |                                  |                                                      | 10 June 2007        | [56]      |

*References: [22, 23, 24, 26, 28, 31-38, 40-45, 48-51, 53-59, 60, 61, 62, 64-66, 68, 69, 70, 71, 72, 73, 74-77]*
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, Denovil G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, Mclntosh 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, 5:254.
2. Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. Nat Rev Genet 2006, 7:85-97.
3. Iafrete A, Feuk L, Rivera MN, Listwenik 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, Lande RS, Botstein D, Akars G, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Rogers JA, Wart DE, Kaufmann ER, Brocker A, Phipps P, Muller-Kahle H, Fulton TR, Ng S, Schumm JW, Bramcan J, Knowlton RG, Barker DF, Crooks SM, et al: A genetic linkage map of the human genome. Cell 1987, 51:1319-337.
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, 359:799-801.
10. Riordan JR, Rammens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou J, Drumml ML, Lannuzi 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, Slaughenhoupt S, Pentao L, Guzzetta V, Trask B, Saucedo-Cardenas O, Barker DF, Kilian 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, Connelly PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Young AB, Shouilou 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. Lande R: 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 WC: 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, Amanzai-Cook RM, Nusbaum C, Zody MC,rows K, 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, Hardin SM, Lasterns S, McVean G, Nejentsev S, Paggioli G, Zhang J, 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, 449:815-861.
20. Freedman ML, Reich D, Penney KL, McDonald-Gibson G, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Conlon LN, Lande RS, Sklar P, Henderson B, Hirschhorn JN, Altshuler D: Assessing the impact of population stratification on genetic association studies. Nat Genet 2004, 36:512-517.
21. Marchini J, Cardona LR, Phillips MS, Donnelly P: The effects of human population structure on large genetic association studies. Nat Genet 2004, 36:512-517.
54. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu Ding CJ, Swift AJ, Narisu N, Hu T, Prunin R, Xiao R, Li XY, Connelly K, Riebow NL, Sprau AG, Tong M, White PP, Hekit KN, Barnhart MW, Bark CW, Goldenst J, 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.

55. 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, Altmuler D, Almgren P, Florez JC, Meyer J, Ardleigh K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastenyk S, Taskinen MR, et al.: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007, 316:1331-1336.

56. Sulem P, Weidon MN, Fawcett KA, Wasson J, Debeijen SM, Daly A, Lango H, Frayling TM, Neumann R, Sherva R, Bleib I, Pharoa PD, Palmer CN, Kimmer C, Tavendale R, Morris AD, McCarthy M, Walker M, Himan G, Glasser B, Permut M, Hattersley AT, Wareham N, Balding DJ, et al.: Fine mapping at the WFS1 confers risk for type 2 diabetes. Nat Genet 2007, 39:951-953.

57. Kathiresan S, Melander O, Guiducci C, Suri A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Walshebrand B, Hegele TR, Corbetta AA, Ewens JW, Breyer JP, McMheal K, Bradford KM, Yaspan BL, Wiklund F, Stattin P, Lindström S, et al.: A common variant near IDDM1 associates with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet 2008, 40:189-197.

58. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, Gómez Pérez F, Frazer KA, Elliott P, Scott J, Milos PM, Cox DR, Thompson JF: Genome-wide scan identifies variation in MBL2 associated with plasma triglycerides. Nat Genet 2008, 40:149-151.

59. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Selzinger S, Joshipura K, Halperin RF, Stampfer C, Jensen KR, Leshter D, Hesterle SE, Peistron L, Levine T, Bertorini T, Graves MC, Mozaffar T, Jackson CE, Bosch P, McVey A, Dick A, Barohn R, Loven-Hoering C, Rosenfeld J, O’Connor DT, Zhang K, Crook R, Ryberg H, Hutton M, Katz J, Simpson W: The effect of N- and O-linked glycosylation on plasma lipoprotein-lipid levels and triglyceride levels: the FHS lipoprotein-lipoprotein. J Clin Invest 2007, 119:1674-1682.

60. Todd JA, Hutton MM, Howie RB, Pohlmann T, Tilling J, Tomazin M, Chiswell D, Drayson M, Hattersley AT, et al.: Strong associations between autoimmunity and type 1 diabetes. Nature 2009, 457:808-811.

61. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer SJ, Scerin AG, Wakefield MJ, Mowry JS, Wiesmann CM, Crain PA, Pfeiffer RS, et al.: Meta-analysis of genome-wide association studies identifies two new loci associated with multiple sclerosis. Science 2009, 326:503-509.

62. Todd JA, Hutton MM, Howie RB, Pohlmann T, Tilling J, Tomazin M, Chiswell D, Drayson M, Hattersley AT, et al.: Strong associations between autoimmunity and type 1 diabetes. Nature 2009, 457:808-811.

63. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer SJ, Scerin AG, Wakefield MJ, Mowry JS, Wiesmann CM, Crain PA, Pfeiffer RS, et al.: Meta-analysis of genome-wide association studies identifies two new loci associated with multiple sclerosis. Science 2009, 326:503-509.

64. Todd JA, Hutton MM, Howie RB, Pohlmann T, Tilling J, Tomazin M, Chiswell D, Drayson M, Hattersley AT, et al.: Strong associations between autoimmunity and type 1 diabetes. Nature 2009, 457:808-811.
