Genetics of Complex Airway Disease

William O. C. Cookson¹ and Miriam F. Moffatt¹

¹Division of Respiratory Sciences, Imperial College London, United Kingdom

The past 3 years have seen highly significant genetic effects identified for a wide variety of common complex diseases, including the airway disorders of asthma and chronic obstructive pulmonary disease. It appears that only a portion of the genetically mediated susceptibility to complex diseases has been identified, and there is much left to be discovered. This review briefly describes the results of the genome-wide association studies of asthma and gives an overview of the parallel and increasingly large-scale studies that are taking place with chronic obstructive pulmonary disease. The future impact is discussed of technological advances that allow increasingly large-scale gene expression studies, next-generation sequencing, and genome-wide testing for epigenetic effects. The use of genetic technology to examine the airway microbiota that interact with the mucosa in health and disease is described.

Keywords: asthma; COPD; genetics; gene expression; epigenetics

The past 3 years have seen many highly significant genetic effects identified for a wide variety of common complex diseases, including the airway disorders of asthma and chronic obstructive pulmonary disease (COPD). This is the result of technological advances that allow high-density genotyping of hundreds of thousands of single-nucleotide polymorphisms (SNPs) in tens of thousands of cases and control subjects. It is also the fruit of an unprecedented level of national and international cooperation. Despite these genuine successes, it appears that only a portion of the genetically mediated susceptibility to complex diseases has been identified, and there is much left to be discovered. In this review we briefly describe the results of the genome-wide association studies of asthma in which we have been involved and give an overview of the parallel and increasingly large-scale studies that are taking place with COPD. We then reflect on the future impact of technological advances that allow increasingly ambitious gene expression studies, next-generation sequencing, and genome-wide testing for epigenetic effects. Genetic findings are highlighting the importance of epithelial barriers in asthma, so we describe the use of genetic technology to examine the airway bacteria that interact with the mucosa in health and disease.

ASTHMA

Asthma is characterized by an abnormal mucosa of the airways, inflammation, and symptoms of wheeze and shortness of breath. Asthma affects more than 1 child in 10 in the developed world, and there are 300 million cases worldwide (1). Although some effective therapies for mild asthma exist, severe asthma remains difficult to treat. The societal cost of the disease is very large (2, 3).

Asthma runs strongly in families and has a heritability of up to 60% (4). Genetic studies offer a structured means of understanding the causes of asthma and of identifying targets for

treating the syndrome. As with other common complex diseases, genetic studies have led to considerable advances in the understanding of asthma. These are discussed in detail by Kathleen Barnes in an accompanying review. In general, robust genetic effects have been identified that carry substantial population-attributable risks. Several genes act in pathways that communicate the presence of epithelial damage to the adaptive immune system, providing a new focus for effective therapies. Genetic findings have also led to a reassessment of the primacy of atopy in asthma and eczema (atopic dermatitis), suggesting that atopy is secondary to epithelial or other events in both diseases.

Despite these advances, only a small component of the overall genetic contribution to asthma has been identified. This missing heritability may be due to multiple small effects (polygenes), rare highly penetrant mutations, or epigenetic modifications of gene function.

Adequately powered Genome-Wide Association Studies (GWAS) have been the method of choice for identifying complex disease genes (5). We have previously performed a first-generation GWAS of approximately 1,000 children with asthma and 1,000 control subjects that showed markers on chromosome 17q21 to be strongly associated with childhood-onset asthma (6, 7).

We extended the GWAS by measuring global gene expression in Epstein-Barr virus-transformed lymphoblastoid cell lines from children with asthma and their siblings (8). This allowed us to demonstrate that the asthma-associated SNPs were also strongly associated ($P = 10^{-23}$) with transcript abundance of *ORMDL3* (6). Subsequently, we and others have shown that the locus also regulates expression of the neighboring gene *GSDMB* (9, 10).

We have led a second-generation study of 10,000 patients with asthma and 16,000 control subjects as part of a large international initiative (the GABRIEL collaboration) to identify genetic and environmental contributions to asthma (11). We have studied cases with adult and occupational asthma as well as childhood disease.

We performed a random-effects pooled analysis of results from individual populations. We found associations for all patients with asthma with genome-wide significance on chromosome 2 within *IL1RL1/IL18R1* ($rs3771166$; $P = 3 \times 10^{-9}$); on chromosome 6 within *HLA-DQ* ($rs9273349$; $P = 7 \times 10^{-14}$), on chromosome 9 flanking *IL33* ($rs1342326$; $P = 9 \times 10^{-10}$), on chromosome 15 within *SMAD3* ($rs744910$; $P = 4 \times 10^{-9}$), and on chromosome 22 within *IL2RB* ($rs2284033$; $P = 1.1 \times 10^{-8}$). Association within the chromosome 17q21 *ORMDL3/GSDMB* locus was confined to childhood-onset disease ($rs2305480$; $P = 6 \times 10^{-23}$). Markers at two additional genes *SLC22A5* and *RORA* were identified with $P < 10^{-7}$.

Genetic studies may be used to determine whether disease-related intermediate phenotypes are causal or secondary to the disease progress (12). For example, a Mendelian randomization strategy has been used recently to show that SNPs modulating C-reactive protein do not cause coronary heart disease (13), whereas a causal role for Lp(a) has been substantiated (14).

We therefore determined the associations between asthma and the total serum IgE concentration. Only *HLA-DQ* showed association to the total serum IgE concentration ($P < 10^{-6}$), and in general loci showing strong association to IgE had mini-

(Received in original form January 10, 2011; accepted in final form February 21, 2011)

Supported by The Wellcome Trust.

Correspondence and requests for reprints should be addressed to Dr. William Cookson M.D., D.Phil., National Heart and Lung Institute, Dovehouse Street, London SW3 6LY, UK. E-mail: w.cookson@imperial.ac.uk

Proc Am Thorac Soc Vol 8, pp 149–153, 2011

DOI: 10.1513/pats.201101-003MS

Internet address: www.atsjournals.org

mal effects on the risk of asthma. This suggests that elevation of IgE is likely to be an inconstant secondary effect of asthma rather than its cause, supported by the absent relationship between atopic sensitization and asthma in many populations (15). Genetic studies in children with atopic dermatitis have shown that defects in barrier proteins such as *FLG* (16) and *SPINK5* (17) commonly predispose to the disease, indicating that increases in IgE levels may be secondary to barrier failure.

Genetic Risk

We tested our results for their ability to determine the risk of an individual developing asthma and to model how genetic variants affect the burden of disease in the populations we studied (11).

We assessed individual risk for the seven SNPs associated with childhood asthma in a classification analysis based on a logistic regression model. The sensitivity of classifying asthma or nonasthma was 35%, the specificity was 75%, the false-positive rate was 25%, and the false-negative rate was 65%. Estimation of the population-attributable risk fraction for the joint action of the GWAS-significant loci indicated that 38% of childhood-onset asthma was attributable to SNP combinations at these loci (above the lowest 10% of risk score in the control subjects; 95% confidence interval, 28–44). In an independent population-based replication sample of 517 cases and 3,486 control subjects (B58C–non-GABRIEL), the same SNPs accounted for 49% of lifetime asthma risk (from birth to middle age) (95% confidence interval, 27–65).

Although these results indicate that the SNPs would not be effective in classifying individuals as asthmatic or nonasthmatic, the genetic variants (and their presumed functional outcomes) have a significant impact on the burden of asthma in the community.

COPD

In common with other complex diseases, candidate gene studies of COPD have been disappointing (18). The use of GWAS, on the other hand, has yielded three reproducible and interesting loci that provide some insight into the disease process. Pillai and colleagues found genome-wide significant associations of the *CHRNA3/CHRNA5/IREB2* region to clinical COPD (19). *CHRNA3* and *CHRNA5* are receptors for nicotine, suggesting a potential link with smoking. Measures of lung function were mapped in subjects from the Framingham Heart Study, showing a replicated and genome-wide significant association between the *HHIP* locus and the FEV₁/FVC ratio (20). This locus was close to genome-wide significance with COPD in the study by Pillai and colleagues (19). Two substantial metaanalyses of lung function in general population samples provided further replication of the association between the *HHIP* locus and the FEV₁/FVC ratio (21, 22). One of these studies also found evidence for association of FEV₁/FVC with the *FAM13A* locus (21), which has independently been associated with COPD status (23).

The *CHRNA3/CHRNA5/IREB2* locus was associated with smoking, whereas the *HHIP* and *FAM13A* loci were not. By contrast, *HHIP* was associated with systemic components of COPD and with the frequency of COPD exacerbations, and *FAM13A* was associated with lung function (24). Examination of the *HHIP* locus in a case-control study of current or former smokers with COPD, lung cancer, or normal lung function showed a shared susceptibility between lung cancer and COPD that may be mediated through the epithelial response to smoking (25).

Unattributed Heritability

Heritability can be defined as the variance in a complex trait that is due to inherited genetic factors. Childhood asthma may

have a heritability as high as 60%, but the heritability accounted for by the loci in our large-scale GWAS was only 4%. This is consistent with other complex human traits such as height (26) and Crohn's disease (27), and the search for the missing heritability or dark heritability is a major focus for geneticists who study complex diseases.

A substantial portion of the dark heritability is expected to result from the effects of multiple loci that are too weak to detect using current sample sizes (28). This is consistent with data in yeast, where only 3% of highly heritable transcript abundances are explained by single-locus inheritance and 50% are consistent with more than five controlling loci of equal effect (29). Statistics have been developed to identify multiple causal variants in human subjects through simultaneous analysis of the entire set of SNPs from a genome-wide study (30).

Although present SNP arrays provide good coverage of common variation in the human genome (>80%), some of the unattributed heritability is due to genetic factors that reside in unmapped regions or variation that is not effectively tagged. Imputation has provided a valuable tool to deal with this, at least in subjects of European ancestry. Dominance and interaction effects may also account for some of the unattributed heritability, although they do not appear to be major factors (8).

Missing heritability may be in part due to rare mutations with high penetrance. Mutations that alter protein function can have large effects, such as *CARD15* (*NOD2*) mutations in inflammatory bowel disease (31) and *FLG* mutations in eczema (atopic dermatitis) (16). Although systematic study of complex diseases with known nonsynonymous SNPs has not in general yielded highly significant results (32), it may be valuable to search for mutations in individuals with severe spectrum disease. Full genomic sequencing of multiple diseased individuals is too expensive for general application. However, whole-exome sequencing with enrichment of exonic sequences by hybridization before ultradeep sequencing (33) has proven to be extremely effective in the identification of rare mutations predisposing to disease (34, 35) and merits application to complex disorders such as asthma.

Exome sequencing has been successful with very small subject samples in Mendelian disorders because it is safe to assume that these diseases are due to mutation rather than to common variants affecting, for example, transcription or subtle aspects of gene function. This assumption allows noncoding mutations to be ignored and common coding mutations to be filtered out of the analysis by comparison to published databases such as Hapmap and the 1,000 genomes project. This assumption is unlikely to apply to complex diseases such as asthma, but, by analogy with successful studies of obesity and metabolic disorders (36), it is possible that patients with severe or strongly familial forms of the disease may be enriched for mutations in particular genes or pathways of relevance to more common forms.

THE EPIGENOME

Epigenetic effects, mediated through mechanisms other than sequence variation, are another possible cause of familial clustering. The patterns of gene expression that determine cellular type and function become stably restricted during development, partly through methylation of CpG sequences and gene silencing (37). Islands of CpG sequences are positioned at the 5' UTRs of many human genes (38). About one fifth of islands are variably methylated, and in one third of these methylation status correlates with transcript abundance (39). Abnormalities of DNA methylation are well recognized in single-gene disorders and in cancer (40), and it is postulated that epigenetic

changes in methylation may be of importance to common human diseases (40) such as asthma and COPD.

Although it is likely that genome-wide studies of methylation status at various loci may identify new genes and pathways that mediate airway diseases, age (39, 41, 42), sex (39, 41), genetic polymorphism (43, 44), and environmental factors (42, 44) have been strongly associated with altered methylation at selected loci. The relative contribution of these factors to methylation at loci genome wide is not known, and it is not certain to what extent true epigenetic inheritance with transmission between generations takes place. These factors will have to be taken into account if methylation changes at individual loci are to help understand complex diseases.

FUNCTIONAL GENOMICS

Although robust associations between asthma and COPD loci are recognized, these signals have not been translated into a full understanding of which gene or genetic elements mediate disease susceptibility at particular loci. Variation in gene expression is an important mechanism underlying susceptibility to complex disease (9). Transcript abundances of genes are directly modified by polymorphism in regulatory elements. Consequently, transcript abundances may be considered as quantitative traits that can be mapped to genetic variation with considerable power. These have been named "expression quantitative trait loci" (45, 46).

In addition to providing information about the biological control of gene expression, such data aid in interpreting the results of GWA studies. The availability of systematically generated expression quantitative trait loci information provides immediate insight into a probable biological base for the disease associations and can help to identify networks of genes involved in disease pathogenesis, exemplified by the identification of a macrophage-enriched network affecting obesity in human subjects (47, 48). This approach allows identification of genetic variants that have minor individual effects on the trait but that can be identified as a group because of the overall perturbation of the network.

THE METAGENOME

It has been estimated that the cells of the human body are outnumbered by 10 to 1 by bacterial cells living in and on the body. Although the airways have been considered in the past to be sterile, modern understanding of the ubiquitous presence of microbes in the most severe environments suggests that airway sterility is unlikely. Only 1% of all bacteria can be cultured routinely in the laboratory (49), so culture is no longer the gold standard for the diagnosis of infections. Culture-independent molecular methods have shown that the microbiota of humans is far greater in extent than previously recognized (50–52), and investigations of inflammatory bowel disease have shown that normal bacterial flora are essential in maintaining a healthy mucosa (53).

Most culture-independent methods for bacterial identification are based on the amplification and sequencing of the bacterial 16S ribosomal RNA gene (16S rRNA). This essential gene features highly conserved sequences that allow the design of PCR primers that amplify sequences from a very wide range of bacteria while intervening highly variable sequences enable the identification of individual bacterial species. Public databases, such as the online ribosomal database, contain reference 16S rRNA sequences for more than 690,000 organisms (<http://rdp.cme.msu.edu/>). In a typical experiment, the 16S gene is amplified from a complex microbial mixture, and the resulting

sequences are used to derive a semiquantitative estimate of the bacterial species present. Most studies of bacterial communities using the 16S gene have relied on analysis of sequence traces, now accomplished most easily by direct sequencing of 16S amplicons with next-generation sequencers such as the Roche 454 (Roche Diagnostics, Branford, CT). An alternative approach uses hybridization of the amplification product to microarrays (54). In analogy to gene expression studies, significant results from such arrays may require confirmation by sequencing.

Although the causes of asthma are unknown, many studies suggest a role for microbiota in its etiology (55). Viral infections are important inducers of seasonal exacerbations of asthma (56), but there is circumstantial evidence that bacterial infections may also play a role. Asymptomatic neonates whose throats are colonized with *Streptococcus pneumoniae*, *Haemophilus influenzae*, or *Moraxella catarrhalis* are at increased risk for recurrent wheeze and asthma early in life (57). These same bacteria have consistently been associated with exacerbations of asthma (58) and COPD (59). The response of patients with asthma to antibiotics also suggests the importance of acute and chronic bacterial infections in the pathogenesis of disease (60). Epidemiological research has consistently indicated that a rich microbial environment in early life confers protection against the development of asthma (61), suggesting the need to understand the extent and nature of normal airway flora.

We have conducted a modest study that compared the airway microbiota at three levels in adult patients with asthma, in the related condition of COPD, and in control subjects (62). We also studied bronchial lavage from children with asthma and control subjects (62). We identified 5,054 16S rRNA bacterial sequences from 43 subjects, detecting greater than 70% of species present. The bronchial tree was not sterile and contained a mean of 2,000 bacterial genomes per cm² surface sampled. Pathogenic proteobacteria, particularly *Haemophilus* spp., were much more frequent in the bronchi of adult patients with asthma and in patients with COPD than in control subjects. We found similar highly significant increases in Proteobacteria in children with asthma. Conversely, Bacteroidetes, particularly *Prevotella* spp., were more frequent in control subjects than in adults or children with asthma or in patients with COPD.

A recent study used a microarray technique to test for relationships between the composition of the airway bacterial microbiota and clinical features of asthma, establishing that bacterial burden and diversity were higher among patients with asthma compared with control subjects, although this study did not identify the organisms responsible for this effect (54).

These early results suggest that the bronchial tree is likely to contain a characteristic microbiota and suggest that this microbiota is disturbed in asthmatic airways. The disturbance may be secondary to many things, including the use of antibiotics, steroids, and β -agonists. The possibility also exists that the chronic presence of known pathogens may contribute to the disease process, but rigorous testing of such hypotheses is needed.

SUMMARY

Genome-wide association studies have made great advances in identifying the most important genes predisposing to airway diseases. However, it appears that most genetic effects on asthma and COPD have yet to be identified. Even though the loci yet to be discovered may have minor individual effects on disease susceptibility, they may identify important pathogenic pathways and lead to novel therapeutic approaches. Further work is necessary to understand the biological consequences of susceptibility variants and the genes they affect. Next-generation

sequencing and measurements of gene expression will be a key tool in future studies. Genome-wide mapping of epigenetic effects are anticipated to be valuable but are likely to be very complex in their interpretation. The genetics of airway microbiome may be just as helpful as human genetics to our understanding of airway inflammation.

Author Disclosure: Neither author has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

References

- ISAAC. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. *Lancet* 1998;351:1225–1232.
- Smith DH, Malone DC, Lawson KA, Okamoto LJ, Battista C, Saunders WB. A national estimate of the economic costs of asthma. *Am J Respir Crit Care Med* 1997;156:787–793.
- Stanford R, McLaughlin T, Okamoto LJ. The cost of asthma in the emergency department and hospital. *Am J Respir Crit Care Med* 1999;160:211–215.
- Duffy DL, Martin NG, Battistutta D, Hopper JL, Mathews JD. Genetics of asthma and hay fever in Australian twins. *Am Rev Respir Dis* 1990;142:1351–1358.
- Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166–176.
- Moffatt MF, Kabisch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007;448:470–473.
- Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008;359:1985–1994.
- Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Depner M, von Berg A, Bufe A, Rietschel E, et al. A genome-wide association study of global gene expression. *Nat Genet* 2007;39:1202–1207.
- Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. *Nat Rev Genet* 2009;10:184–194.
- Verlaan DJ, Berlivet S, Hunninghake GM, Madore AM, Lariviere M, Moussette S, Grundberg E, Kwan T, Ouimet M, Ge B, et al. Allele-specific chromatin remodeling in the ZBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. *Am J Hum Genet* 2009;85:377–393.
- Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WO. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010;363:1211–1221.
- Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med* 2008;5:e177.
- Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, Erdmann J, Braund P, Engert JC, Bennett D, et al. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* 2009;302:37–48.
- Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med* 2009;361:2518–2528.
- Weinmayr G, Weiland SK, Bjorksten B, Brunekreef B, Buchele G, Cookson WO, Garcia-Marcos L, Gotua M, Gratzou C, van Hage M, et al. Atopic sensitization and the international variation of asthma symptom prevalence in children. *Am J Respir Crit Care Med* 2007;176:565–574.
- Palmer CN, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, Goudie DR, Sandilands A, Campbell LE, Smith FJ, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 2006;38:441–446.
- Walley AJ, Chavanas S, Moffatt MF, Esnouf RM, Ubhi B, Lawrence R, Wong K, Abecasis GR, Jones EY, Harper JJ, et al. Gene polymorphism in Netherton and common atopic disease. *Nat Genet* 2001;29:175–178.
- Smolonska J, Wijmenga C, Postma DS, Boezen HM. Meta-analyses on suspected chronic obstructive pulmonary disease genes: a summary of 20 years' research. *Am J Respir Crit Care Med* 2009;180:618–631.
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.
- Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009;5:e1000429.
- Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loefer LR, Marcianti KD, Franceschini N, van Durme YM, Chen TH, Barr RG, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45–52.
- Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, et al. Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2010;42:36–44.
- Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua A, Sparrow D, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42:200–202.
- Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, Coxson HO, Lomas DA, Silverman EK. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2010;182:1498–1505.
- Young RP, Whittington CF, Hopkins RJ, Hay BA, Epton MJ, Black PN, Gamble GD. Chromosome 4q31 locus in COPD is also associated with lung cancer. *Eur Respir J* 2010;36:1375–1382.
- Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;40:575–583.
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;40:955–962.
- Petretto E, Mangion J, Dickens NJ, Cook SA, Kumaran MK, Lu H, Fischer J, Maatz H, Kren V, Pravenec M, et al. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet* 2006;2:e172.
- Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci USA* 2005;102:1572–1577.
- Hoggart CJ, Whittaker JC, De Iorio M, Balding DJ. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genet* 2008;4:e1000130.
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599–603.
- Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, Kiatkowski DP, McCarthy MI, Ouwehand WH, Samani NJ, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007;39:1329–1337.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–276.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30–35.
- Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet* 2010;42:13–14.
- O'Rahilly S. Human genetics illuminates the paths to metabolic disease. *Nature* 2009;462:307–314.
- Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16:6–21.
- Bird AP. CpG islands as gene markers in vertebrate nucleus. *Trends Genet* 1987;3:342–347.
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;38:1378–1385.
- Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007;447:433–440.
- Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. *PLoS ONE* 2010;5:e10028.

42. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 2009;5:e1000602.
43. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, Kahn RS, Ophoff RA, *et al.* The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS ONE* 2009;4:e6767.
44. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum Mol Genet* 2007;16:547–554.
45. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* 2003;422:297–302.
46. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004;430:743–747.
47. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil Zhang C, Lamb J, Edwards S, Sieberts SK, *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* 2008;452:429–435.
48. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, *et al.* Genetics of gene expression and its effect on disease. *Nature* 2008;452:423–428.
49. Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 1985;39:321–346.
50. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007;449:804–810.
51. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, Macfarlane S. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol* 2007;73:7435–7442.
52. Grice EA, Kong HH, Renaud G, Young AC, Bouffard GG, Blakesley RW, Wolfsberg TG, Turner ML, Segre JA. A diversity profile of the human skin microbiota. *Genome Res* 2008;18:1043–1050.
53. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, Yu D, Schilter HC, Rolph MS, Mackay F, Artis D, *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* 2009;461:1282–1286.
54. Huang YJ, Nelson CE, Brodie EL, Desantis TZ, Baek MS, Liu J, Woyke T, Allgaier M, Bristow J, Wiener-Kronish JP, *et al.* Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J Allergy Clin Immunol* 2011;127:372–381.
55. Cookson W. The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat Rev Immunol* 2004;4:978–988.
56. Johnston S, Pattemore P, Sanderson G, Smith S, Lampe F, Josephs L, Symington P, O'Toole S, Myint S, Tyrrell D, *et al.* Community study of role of viral infections in exacerbations of asthma in 9–11 year old children. *BMJ* 1995;310:1225–1229.
57. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bonnelykke K, Brasholt M, Heltberg A, Vissing NH, Thorsen SV, *et al.* Childhood asthma after bacterial colonization of the airway in neonates. *N Engl J Med* 2007;357:1487–1495.
58. Kraft M. The role of bacterial infections in asthma. *Clin Chest Med* 2000;21:301–313.
59. Sethi S, Evans N, Grant BJ, Murphy TF. New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. *N Engl J Med* 2002;347:465–471.
60. Blasi F, Johnston SL. The role of antibiotics in asthma. *Int J Antimicrob Agents* 2007;29:485–493.
61. Eder W, Ege MJ, von Mutius E. The asthma epidemic. *N Engl J Med* 2006;355:2226–2235.
62. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, *et al.* Disordered microbial communities in asthmatic airways. *PLoS ONE* 2010;5:e8578.