Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Genetics & Development

Prediction of individual genetic risk of complex disease

Naomi R Wray[1], Michael E Goddard[2] and Peter M Visscher[1]

Most common diseases are caused by multiple genetic and environmental factors. In the last 2 years, genome-wide association studies (GWAS) have identified polymorphisms that are associated with risk to common disease, but the effect of any one risk allele is typically small. By combining information from many risk variants, will it be possible to predict accurately each individual person's genetic risk for a disease? In this review we consider the lessons from GWAS and the implications for genetic risk prediction to common disease. We conclude that with larger GWAS sample sizes or by combining studies, accurate prediction of genetic risk will be possible, even if the causal mutations or the mechanisms by which they affect susceptibility are unknown.

Addresses
[1] Genetic Epidemiology and Queensland Statistical Genetics, Queensland Institute of Medical Research, Brisbane, Australia
[2] Faculty of Land and Food Resources, University of Melbourne and Department of Primary Industries, Victoria, Australia

Corresponding author: Wray, Naomi R (Naomi.Wray@qimr.edu.au)

Current Opinion in Genetics & Development 2008, **18**:257–263

This review comes from a themed issue on
Genetics of disease
Edited by Nick Hastie and Aravinda Chakravarti

Available online 28th August 2008

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2008.07.006

Introduction

Common complex diseases, such as psychiatric disorders, cancer, diabetes, heart disease and asthma, are caused by multiple genetic and environmental factors. Significant heritabilities and increased risk to relatives quantify the importance of the genetic factors. To predict the risk of a complex disease for a healthy individual we need to know and be able to measure risk factors, their effect sizes and how they interact. Although prediction of total risk is an ultimate goal, prediction of *genetic* risk, the risk that can be attributed to inherited genetic variants, is an important component and is the focus of this review.

Predictive genetic tests are already available for a huge range of Mendelian disorders, those for which a single genetic mutation is known to cause the disease [1], but for common complex diseases, very few causal genetic risk factors have been identified. Consequently genetic prediction has been mostly limited to family history information. However, the value of family history information in clinical diagnosis is limited: for example, all children in a nuclear family are predicted to have the same genetic risk based on the history of disease in their parents and more distant relatives, yet they are genetically different. Great progress has been made in the last 2 years in the identification of common polymorphisms that are associated with risk of disease in the population. Now that these risk variants have been identified, can they be used to predict an individual's genetic risk for a particular disease more accurately than can be done using family history information?

The possible impact of prediction of genetic risk on individual and population health has been recognized for sometime [2,3,4•], but it is only the new developments in high-density genotyping technology [5] that make genetic risk prediction within reach. In this review, we discuss the recent advances made by genome-wide association studies in identifying genetic variants associated with disease and the way in which these results can be used for 'genomic profiling' [6], the prediction of genetic risk to disease.

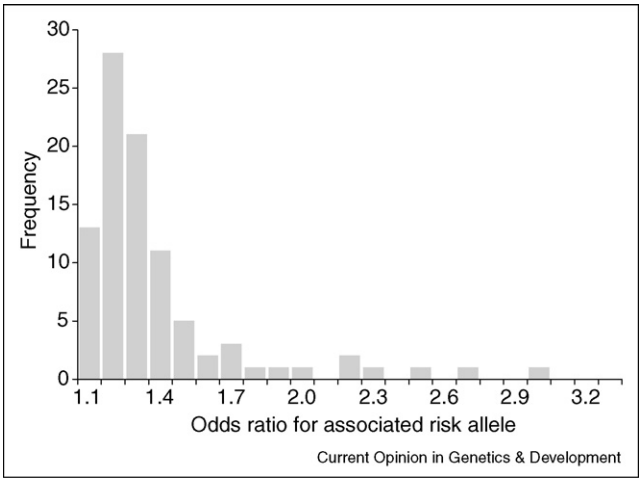
Genetic architecture of complex diseases

Genetic architecture of a disease refers to the number of genetic polymorphisms that affect risk of disease, the distribution of their allelic frequencies, the distribution of their effect sizes and their genetic mode of action (additive, dominant and/or epistatic). Prediction of genetic risk is dependent on the underlying genetic architecture because as the number of causal variants increases, the proportion of variance explained by each decreases. As a consequence, it becomes harder to detect them experimentally and to estimate their allelic effects accurately.

Genome-wide association studies (GWAS) – what can they tell us about genetic architecture of complex diseases?

Following the sequencing of the human genome and the creation of a map of common haplotypes [7], commercial panels of ~300 to ~500 thousand single nucleotide polymorphisms (SNPs) were created that covered most common variation. These SNP chips have been used extensively in the last two years in large GWAS for common diseases and quantitative traits. To date, more than 25 (e.g. [8••]) of these studies have been published and many more are under way. Validation studies have shown that at least some of the associations found in GWAS are replicated in independent samples. What

Figure 1



Frequency distribution of effect sizes expressed as Odds ratio for the risk allele of 92 validated associated SNPs identified from GWAS. These SNPs represent associations with one of 16 disorders (listed in Appendix A). The power of the GWAS to detect variants with effect size of 1.1 or smaller was low.

general conclusions can we draw from these recent studies about the genetic architecture of complex diseases?

(i) Most associated variants that have been detected are common (minor allele frequency, MAF >0.05), although this is mostly a reflection of study design in terms of SNPs selected for genotyping and power of the sample size. Risk variants are almost equally likely to be the minor or major alleles.

(ii) Although a few variants of large effects (allelic Odds ratio, OR > 2) have been detected, the vast majority of the effect sizes of risk alleles are small, typically OR <1.5 but many around 1.1 and 1.2 (Figure 1), which are the limits of detection given the experimental sample sizes employed to date.

(iii) For most diseases, the detected variants explain little of the total genetic variance that we know exist (from twin and family studies). For breast cancer, prior to GWAS, about 20% of the familial risk could be accounted for by variants in six genes (16% from BRACA1 and BRAC2 alone) [9•,10]. Results from GWAS have identified variants that account for only an additional 2.3% of the familial risk [9•].

(iv) The genetic mode of action is additive on the logarithm of risk (log-risk) scale (multiplicative on the risk scale) for the majority of variants identified, as expected from the significant heritabilities [11], but the power to detect interactions between variants has been limited by the sample sizes used to date [8••].

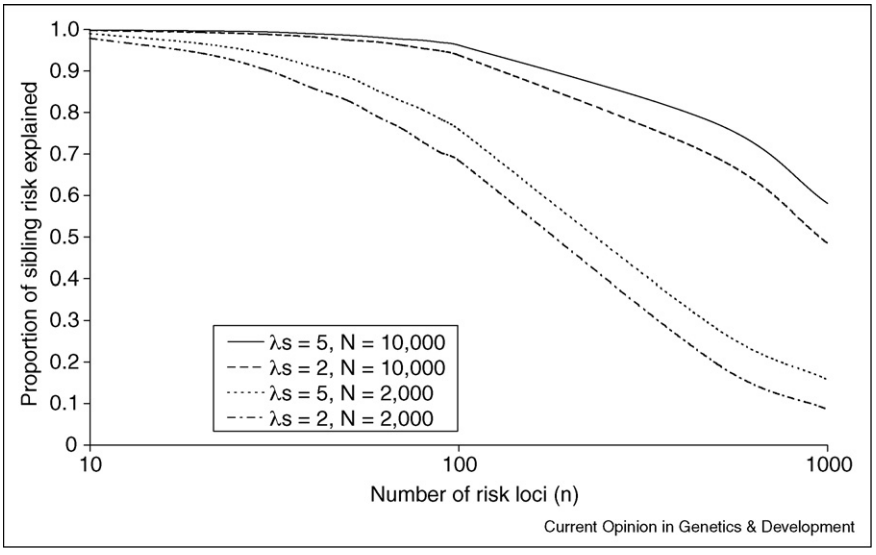
Together these studies point to a genetic architecture of a few risk variants of large or moderate size, but a very substantial number of variants which generate small increases in disease risk [8••]. But why has so little of the genetic variance been explained? First, the effect sizes may be underestimated as the associated variants may only be in linkage disequilibrium (LD) with the causal variants. Second, the genotyped markers may not be representative of all the inherited genetic variance, especially structural variation [12•,13•,14•]. Thirdly, the GWAS SNP panels are limited to polymorphisms with MAF >0.01. Rare causal variants are likely to exist but will not be represented ('tagged') by the genotyped SNPs, as high r^2 LD between SNPs requires a close matching of allele frequencies [15]. However, such variants cannot individually (by virtue of being rare) explain much of the genetic variance [16,17•].

Since most of the identified associated SNPs have effect size close to the limit dictated by the power of the studies, a likely explanation, at least in part, is that there are many common polymorphisms with effects that were too small to pass the stringent significance thresholds. Given the effect size of variants detected so far, study samples of the order of 10 000 cases and controls would be needed to detect variants which explain the majority of genetic variance [8••,18•].

Prediction of genetic risk from genome-wide markers

Many genetic variants each conferring only a small increased risk to disease are individually not useful in predicting a person's genetic risk to disease. However, a risk equation combining presence/absence of each risk variant and its effect size can generate a personalised prediction of genetic risk. Given the emerging evidence for the genetic architecture underlying complex diseases, how accurate would a prediction equation be in predicting individual genetic risk? The SNPs (or other markers) do not have to be the causative mutations: they just need to be in high LD with the causative mutations so that there is a consistent association between the SNP and disease risk. We investigated this problem using simulation of GWAS based on realistic assumptions of underlying genetic architecture [18•]. For simplicity we assumed perfect LD between causal and genotyped SNPs and that the genotyped SNPs could explain all the genetic variance by combining additively on the log-risk scale. Our results confirmed that GWAS of only a couple of thousand of cases and controls (the current norm) are still too small to detect most of the causal variants and those detected will not explain enough of the genetic variance to be useful in risk prediction (Figure 2). However, for a case-control study with 10 000 cases and controls our results show that even for diseases controlled by 1000 loci with mean relative risk of only 1.04 (and maximum ~1.3), it will be possible to identify ~75 of the larger loci that explain >50% of the genetic variance and yield a correlation between predicted and true genetic risk >0.7 [18•]. If 500 loci affect disease risk,

Figure 2



Proportion of sibling risk (λ_s) explained from GWAS of either $N = 2000$ or 10 000 cases and controls for a common complex disease with $\lambda_s = 2$ or 5. Simulation results using method presented in Ref. [18•] assuming a GWAS of 300 000 SNPs of which n explain all the genetic variance; these n have allele frequencies drawn from a uniform distribution and have effect size drawn from an exponential distribution and act additively on the log-risk scale. As n increases the mean effect size for associated alleles decreases. The mean effect size has OR 1.08 when 1000 loci and 1.30 when 100 loci explain the genetic variance, when $\lambda_s = 5$. Results are robust to the prevalence rate of the disease in the population.

the correlation is ~0.9. Therefore, as results from multiple GWAS are combined, a larger fraction of the genetic variance is likely to be explained and accurate prediction of genetic risk to disease will become possible even though the risks conveyed by individual variants are small.

A practical issue is identification of optimal statistical methods for selecting and combining SNPs into a prediction equation. The central problem of selecting a set of predictors when the number of measured variables is very large and greater than the number of observed phenotypic values is a common statistical challenge – the large p small n paradigm – arising in a wide range of diverse fields (including gene expression data, for example [19]) and a number of these methods have been applied to SNP selection [20–25,26•]. Just as the individual SNP associations discovered in GWAS need to be validated in independent datasets, prediction equations estimated from one dataset need to be validated in other independent data.

Assessing the utility of prediction of genetic risk

The precision of a prediction of genetic risk can be assessed by the correlation (ρ) between true and predicted genetic risk or by ρ^2 which is the proportion of the genetic variance explained by the associated variants in data independent of that used to identify them. The magnitude of ρ depends on the proportion of the genetic variance tagged by the genotyped markers and the accuracy with which their effects on risk are estimated. Whereas ρ is a tangible measure for quantitative traits, for disease traits it is perhaps best expressed at the proportion of the sibling risk ratio (λ_s) explained by the prediction (Figure 2).

The utility of the multi-marker equation for predicting *absolute* risk of disease depends on the heritability of the trait. If the heritability is low, the value of predicting the genetic component of risk in predicting disease will also be low. The ability to predict disease status given perfect knowledge of genetic risk ($\rho = 1$) was the subject of [27]. The receiver operator characteristic (ROC) curve can be used to compare the sensitivity and specificity of a genomic profile test in correctly classifying diseased and non-diseased individuals. However, the maximum area under the curve explained by a genetic prediction is limited by the heritability of the disease (and also disease prevalence if applied to a population sample). Therefore, we have argued that use of ROC statistics can be confusing for assessing the accuracy of the genetic component of risk prediction [18•].

The epidemiological measure population attributable fraction (PAF) can be used to quantify the decrease in disease prevalence if the identified set of risk factors is eliminated. PAF measures the change in the mean of disease prevalence and can be high even when the proportion of genetic variance explained by the markers is low. For example, the 11 validated variants for type 2 diabetes identified by GWAS [28], have risk allele

frequencies ranging from 0.31 to 0.87 and OR ranging from 1.10 to 1.37; together they explain <5% of the genetic variance or λ_s but have PAF >80%. PAF describes the effect of eliminating risk alleles on the population mean whereas ρ^2 quantifies the precision of genetic prediction to individuals. These two measures represent utility to the so-called 'population' and 'high-risk' based approaches to prevention of common disease, discussed in detail in the context of genomic profiling by [6].

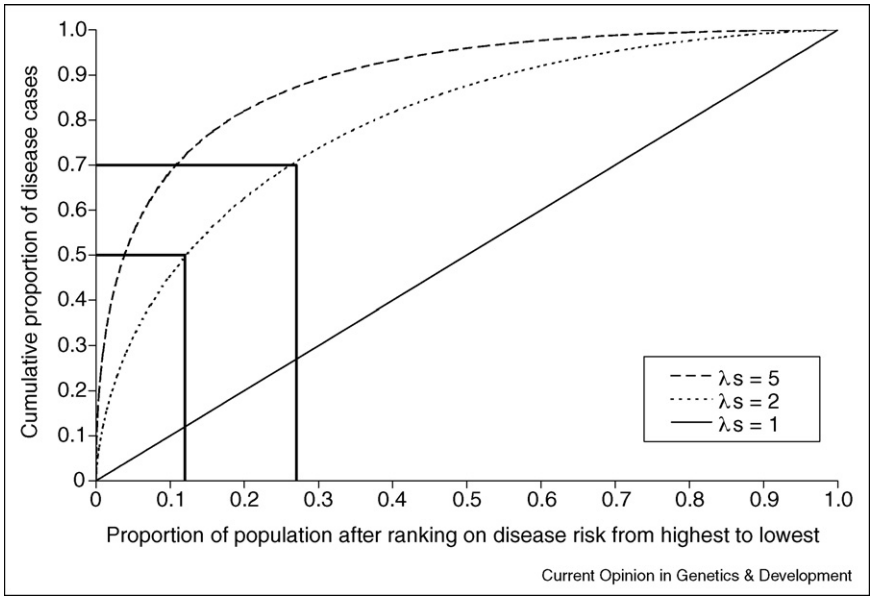
The use of genetic risk prediction

Prediction of genetic risk will be useful in diagnosis, treatment, prognosis and prevention strategies, as family history is currently, but with greater power. For diseases with very low population prevalence, genetic risk prediction might be limited to those with family history of the disease, providing differential risk prediction for family members. For higher prevalence disease, genetic risk prediction could be applied at a population level to identify a sub-group genetically most at risk. Figure 3 shows that, even for diseases with λ_s of only 2, the majority of future disease cases occur in the 15% of people with highest genetic risk [29]. If we assume that environmental (E) risk factors, like genetic (G) risk factors, act additively on the log- risk scale, then on the risk scale they will act multiplicatively, implying substantial G × E interaction on this scale. Therefore, the effects of environmental risk factors or intervention strategies will be larger for those individuals with high genetic risk to disease. For example, if an intervention strategy reduces relative risk by 50% then people with a 40% genetic risk of disease can cut their absolute risk to 20% whereas people at a low genetic risk of 1% will cut their risk to 0.5%. Therefore, it would be efficient use of resources to target the intervention to the small proportion of the population at high risk (Figure 3). Such strategies would have even more importance if G × E is present on the log-risk scale.

A number of studies have already used associated genetic variants to predict genetic risk of disease, often in combination with environmental factors [30–35], but their success has been limited by the small proportion of genetic variance explained by the variants identified to date. Time will tell if larger study samples in GWAS are sufficient for identification of variants that explain the majority of the genetic variance. For some diseases (e.g. psychiatric disorders [36,37]), increasing phenotypic homogeneity of clinical diagnosis in samples used in GWAS may be needed before accurate predictions can be made. Moreover, biologically driven, genotypically defined diagnoses, so-called 'reverse phenotyping' [38], may result [37].

Implementation of risk prediction in a clinical context has the potential to be of major economic benefit to population health [4•]. It also raises serious ethical and social concerns [4•,39,40], but these may have been by-passed as November 2007 saw the launch of three new companies (Decodeme, 23andme, Navigenics) offering genomic profile services for less that US $1000.

Figure 3



The cumulative proportion of disease cases present in the population when ranked from lowest to highest genetic risk for diseases with different λ_s (after [29]). Even for diseases with λ_s of only 2, 70% of cases occur within the top 27% and 50% of cases occur in the top 12% of those ranked on genetic risk. These results are independent of disease prevalence, although λ_s is constrained to be low for high prevalence diseases.

Conclusion

The results of GWAS provide empirical evidence that the genetic architecture of complex disease is one of many common causal variants each of small, additive effect. GWAS have not yet identified variants that explain enough of the genetic variance to make accurate predictions of genetic risk. However, simulation studies and experiments suggest that with larger sample sizes or by combining studies, accurate prediction of genetic risk will be possible. In the long term, complete sequencing of each person's genome [41] means that risk of a disease could be predicted from all sequence variants. In the short term, genome variation can be represented by the genotyping of a representative set of SNPs. The value from predicting individual disease risk from multiple associated variants could be reaped long before the causal mechanism of each is determined. Cost-effective intervention strategies could be targeted to the subset of the population most at genetic risk to disease. In the long-term prediction of this additive genetic risk can be combined with predictions from environmental risk factors, non-additive genetic factors and interactions between them, along with other molecular (such as gene or protein expression) profiles [42•]. However, an individual's genomic profile could be available from birth, long before exposure to most environmental risk factors takes place.

Acknowledgements

This work was supported by the Australian National Health and Medical Research Council grants 389892, 442915, 443011 and 496688.

Appendix A

References for the 16 disorders used in Figure 2:

1. Abdominal aortic aneurysm [43].
2. Age-related macular degeneration [44].
3. Amyotrophic lateral sclerosis [45].
4. Ankylosing spondylitis [46].
5. Asthma [47].
6. Autoimmune thyroid disease [46].
7. Breast cancer [48].
8. Childhood obesity [49].
9. Colorectal cancer [50].
10. Coronary heart disease [51,52].
11. Crohn's disease [53–56].
12. Prostate cancer [57,58].
13. Rheumatoid arthritis [59,60].
14. Systemic lupus erythematosus [61–63].
15. Type 1 diabetes [64,65].
16. Type 2 diabetes [28].

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. McKusick V: **Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders**. edn 12. Baltimore: Johns Hopkins University Press; 1998.

2. Collins FS, Green ED, Guttmacher AE, Guyer MS: **A vision for the future of genomics research**. *Nature* 2003, **422**:835-847.

3. Bell J: **Predicting disease using genomics**. *Nature* 2004, **429**:453-456.

4. • Khoury MJ, Jones K, Grosse SD: **Quantifying the health benefits of genetic tests: the importance of a population perspective**. *Genet Med* 2006, **8**:191-195.
An epidemiologic framework is used to investigate the value of genomic profiling.

5. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J *et al.*: **Large-scale genotyping of complex DNA**. *Nat Biotechnol* 2003, **21**:1233-1237.

6. Khoury MJ, Yang Q, Gwinn M, Little J, Dana Flanders W: **An epidemiologic assessment of genomic profiling for measuring susceptibility to common diseases and targeting interventions**. *Genet Med* 2004, **6**:38-47.

7. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM *et al.*: **A second generation human haplotype map of over 3.1 million SNPs**. *Nature* 2007, **449**:851-861.

8. •• WTCCC: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls**. *Nature* 2007, **447**:661-678.
The pioneering GWAS from the Wellcome Trust Case Control Consortium of Bipolar Disorder, Coronary Artery Disease, Crohn's Disease, Hypertension, Rheumatoid Arthritis, Type 1 Diabetes and Type 2 Diabetes.

9. • Stratton MR, Rahman N: **The emerging landscape of breast cancer susceptibility**. *Nat Genet* 2008, **40**:17-22.
Genetic architecture of a complex human disease is probably most explored, both empirically and theoretically, for breast cancer.

10. Easton DF: **How many more breast cancer predisposition genes are there?** *Breast Cancer Res* 1999, **1**:14-17.

11. Risch N: **Linkage strategies for genetically complex traits. I. Multilocus models**. *Am J Hum Genet* 1990, **46**:222-228.

12. • 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.
Same as Ref. [14•].

13. • Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen WW *et al.*: **Global variation in copy number in the human genome**. *Nature* 2006, **444**:444-454.
Same as Ref. [14•]

14. • Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L *et al.*: **Paired-end mapping reveals extensive structural variation in the human genome**. *Science* 2007, **318**:420-426.
Three papers which illustrate the emerging evidence of the extent of structural variation. The latest SNP chips for use in GWAS provide good coverage of copy number variant regions and so it will become clearer if structural variants include important risk variants for common complex diseases.

15. Wray NR: **Allele frequencies and the r^2 measure of linkage disequilibrium: impact on design and interpretation of association studies**. *Twin Res Hum Genet* 2005, **8**:87-94.

16. Reich DE, Lander ES: **On the allelic spectrum of human disease**. *Trends Genet* 2001, **17**:502-510.

17. • Craddock N, O'Donovan MC, Owen MJ: **Phenotypic and genetic complexity of psychosis – Invited commentary on ...** Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry* 2007, **190**:200-203.
The authors argue against the importance of rare variants in explaining a substantial amount of the genetic variance for schizophrenia.

18. • Wray NR, Goddard ME, Visscher PM: **Prediction of individual risk to disease from genome-wide association studies**. *Genome Res* 2007, **17**:1520-1528.

We use simulation and a quantitative genetic framework to investigate the value of genomic profiling.

19. Dudoit S, Fridlyand J, Speed T: **Comparison of discrimination methods for the classification of tumors using gene expression data**. *J Am Stat Assoc* 2002, **97**:77-87.

20. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps**. *Genetics* 2001, **157**:1819-1829.

21. Long N, Gianola D, Rosa GJM, Weigel KA, Avendano S: **Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers**. *J Anim Breed Genet* 2007, **124**:377-389.

22. Perez-Enciso M: **Multiple association analysis via simulated annealing (MASSA)**. *Bioinformatics* 2006, **22**:573-580.

23. Bureau A, Dupuis J, Falls K, Lunetta KL, Hayward B, Keith TP, Van Eerdewegh P: **Identifying SNPs predictive of phenotype using random forests**. *Genet Epidemiol* 2005, **28**:171-182.

24. Auro K, Alanne M, Kristiansson K, Silander K, Kuulasmaa K, Salomaa V, Peltonen L, Perola M: **Combined effects of thrombosis pathway gene variants predict cardiovascular events**. *PLoS Genet* 2007, **3**:e120.

25. Heidema AG, Feskens EJ, Doevendans PA, Ruven HJ, van Houwelingen HC, Mariman EC, Boer JM: **Analysis of multiple SNPs in genetic association studies: comparison of three multi-locus methods to prioritize and select SNPs**. *Genet Epidemiol* 2007, **31**:910-921.

26. • Goddard ME, Hayes BJ: **Genomic selection**. *J Anim Breed Genet* 2007, **124**:323-330.
Use of genome-wide markers for assessing genetic value of individuals in breeding programmes, so-called 'genomic selection', is an important area of research in livestock genetics. Many of the methods developed are relevant to genomic profiling for disease risk in humans.

27. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM: **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med* 2006, **8**:395-400.

28. Frayling TM: **Genome-wide association studies provide new insights into type 2 diabetes aetiology**. *Nat Rev Genet* 2007, **8**:657-662.

29. Pharoah PDP, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BAJ: **Polygenic susceptibility to breast cancer and implications for prevention**. *Nat Genet* 2002, **31**:33-36.

30. Morrison AC, Bare LA, Chambless LE, Ellis SG, Malloy M, Kane JP, Pankow JS, Devlin JJ, Willerson JT, Boerwinkle E: **Prediction of coronary heart disease risk using a genetic risk score: the atherosclerosis risk in communities study**. *Am J Epidemiol* 2007, **166**:28-35.

31. Bare LA, Morrison AC, Rowland CM, Shiffman D, Luke MM, Iakoubova OA, Kane JP, Malloy MJ, Ellis SG, Pankow JS *et al.*: **Five common gene variants identify elevated genetic risk for coronary heart disease**. *Genet Med* 2007, **9**:682-689.

32. Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, Seddon JM: **Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration**. *Nat Genet* 2006, **38**:1055-1059.

33. van Hylckama Vlieg A, Baglin CA, Bare LA, Rosendaal FR, Baglin TP: **Proof of principle of potential clinical utility of multiple SNP analysis for prediction of recurrent venous thrombosis**. *J Thromb Haemost* 2008, **6**:161-169.

34. Pharoah PDP, Tyrer J, Dunning AM, Easton DF, Ponder BAJ: **Association between common variation in 120 candidate genes and breast cancer risk**. *Plos Genet* 2007, **3**:401-406.

35. Lyssenko V, Almgren P, Anevski D, Orho-Melander M, Sjogren M, Saloranta C, Tuomi T, Groop L: **Genetic Prediction of Future Type 2 Diabetes**. *PLoS Med* 2005, **2**:e345.

36. Schulze TG, Hedeker D, Zandi P, Rietschel M, McMahon FJ: **What is familial about familial bipolar disorder? Resemblance among relatives across a broad spectrum of phenotypic characteristics**. *Arch Gen Psychiatry* 2006, **63**:1368-1376.

37. Craddock N, Owen MJ: **Rethinking psychosis: the disadvantages of a dichotomous classification now outweigh the advantages**. *World Psychiatry* 2007, **6**:20-27.

38. Schulze TG, McMahon FJ: **Defining the phenotype in human genetic studies: Forward genetics and reverse phenotyping**. *Hum Hered* 2004, **58**:131-138.

39. Grosse SD, Khoury MJ: **What is the clinical utility of genetic testing?** *Genet Med* 2006, **8**:448-450.

40. Hodge JG Jr: **Ethical issues concerning genetic testing and screening in public health**. *Am J Med Genet C Semin Med Genet* 2004, **125**:66-70.

41. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G *et al.*: **The diploid genome sequence of an individual human**. *Plos Biol* 2007, **5**:2113-2144.

42. • Ioannidis JPA: **Is molecular profiling ready for use in clinical decision making?** *Oncologist* 2007, **12**:301-311.
Discussions about the use of molecular profiling (combining of any molecular measures for use in disease diagnosis, prevention or treatment) in clinical decision making are relevant also to genomic profiling.

43. Helgadottir A, Thorleifsson G, Magnusson KP, Grétarsdottir S, Steinthorsdottir V, Manolescu A, Jones GT, Rinkel GJ, Blankensteijn JD, Ronkainen A *et al.*: **The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm**. *Nat Genet* 2008, **40**:217-224.

44. Swaroop A, Branham KEH, Chen W, Abecasis G: **Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits**. *Hum Mol Genet* 2007, **16**:R174-R182.

45. 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 *et al.*: **Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis**. *Nat Genet* 2008, **40**:29-31.

46. Newport M, Sirugo G, Lyons E, Vannberg F, Hill AVS, Bradbury LA, Farrar C, Pointon JJ, Wordsworth P, Brown MA *et al.*: **Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants**. *Nat Genet* 2007, **39**:1329-1337.

47. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, Depner M, von Berg A, Bufe A, Rietschel E *et al.*: **Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma**. *Nature* 2007, **448**:470-473.

48. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R *et al.*: **Genome-wide association study identifies novel breast cancer susceptibility loci**. *Nature* 2007, **447**:1087-1093.

49. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB *et al.*: **A common genetic variant is associated with adult and childhood obesity**. *Science* 2006, **312**:279-283.

50. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z *et al.*: **Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk**. *Nat Genet* 2008, **40**:26-28.

51. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR *et al.*: **A common allele on chromosome 9 associated with coronary heart disease**. *Science* 2007, **316**:1488-1491.

52. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM *et al.*: **Newly identified loci that influence lipid concentrations and risk of coronary artery disease**. *Nat Genet* 2008, **40**:161-169.

53. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D *et al.*: **Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility**. *Nat Genet* 2007, **39**:830-832.

54. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A *et al.*: **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.

55. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A *et al.*: **A genome-wide association study identifies IL23R as an inflammatory bowel disease gene**. *Science* 2006, **314**:1461-1463.

56. 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.

57. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A *et al.*: **Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24**. *Nat Genet* 2007, **39**:631-637.

58. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A *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.

59. Plenge RM, Cotsapas C, Davies L, Price AL, Bakker PIW, Maller J, Pe'er I, Burtt NP, Blumenstiel B, DeFelice M *et al.*: **Two independent alleles at 6q23 associated with risk of rheumatoid arthritis**. *Nat Genet* 2007, **39**:1477-1482.

60. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, Liew A, Khalili H, Chandrasekaran A, Davies LRL *et al.*: **TRAF1-C5 as a risk locus for rheumatoid arthritis – A genomewide study**. *New Engl J Med* 2007, **357**:1199-1209.

61. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD: **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.

62. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, Chen W, Zhu C, McEver RP, Kimberly RP *et al.*: **A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus**. *Nat Genet* 2008, **40**:152-154.

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

64. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M *et al.*: **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.

65. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C, Ionescu-Tirgoviste C, Widmer B, Dunger DB *et al.*: **A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region**. *Nat Genet* 2006, **38**:617-619.