and dAkt phosphorylation. Previous RNAi screens have used reporter gene assays[7,8], and these too could be easily adapted to microarray format. However, a cell is only a cell, and for systematically studying gene function by RNAi in the context of a whole organism, *Caenorhabditis elegans* will remain the system of choice at present, because of the ease of systemic RNAi delivery[9].

So why would a researcher choose to perform a genome-wide RNAi screen in Drosophila cells, when the same screen could be performed in mammalian cells using arrayed siRNAs or shRNAs? First, triggering RNAi using long dsRNAs in Drosophila cells is a very robust procedure. Second, genome-scale RNAi libraries are already available for Drosophila [7,8,10], whereas only smaller libraries are currently available for mammalian cells[11–13]. Third, Drosophila has a less redundant genome than mammals, which should allow functions to be identified for a greater proportion of genes by RNAi screens. Finally, and perhaps most excitingly, two distinct dsRNAs can be used in combination to effectively silence two genes at once in Drosophila cells. It is likely that reproducibility targeting any two mammalian genes simultaneously by RNAi will prove a more difficult challenge.

Targeting two genes at once by RNAi opens up the ability to perform large-scale screens for synthetic or epistatic genetic relationships. Sabatini and co-workers demonstrate the power of this approach by identifying genes that affect the phosphorylation levels of the antiapoptotic protein dAkt when the major dAkt phosphatase dPTEN is simultaneously inactivated by RNAi. Understanding the phenotypic consequences of interactions between two or more genes is one of the great challenges for many biologists, whether they work on model organisms or complex genetic diseases in humans. Indeed, recent work using the gene-deletion collection in yeast has suggested that the extent of synthetic interactions between genes is probably much larger than previously imagined[14]. It seems likely that the RNAi microarrays described by Sabatini and co-workers may provide a powerful platform with which to start systematically testing the vast number of gene combinations required to dissect synthetic genetic interactions in metazoans.

1. Wheeler, D.B. *et al. Nat. Methods* **1**, 127–132 (2004).
2. Ziauddin, J. & Sabatini, D.M. *Nature* **411**, 107–110 (2001).
3. Mousses, S. *et al.* RNAi microarray analysis in cultured mammalian cells. *Genome Res* **13**, 2341–2347 (2003).
4. Kumar, R., Conklin, D. S. & Mittal, V. *Genome Res.* **13**, 2333–2340 (2003).
5. Silva, J.M., Mizuno, H., Brady, A., Lucito, R. & Hannon, G.J. *Proc. Natl. Acad. Sci. USA* **101**, 6548–6552 (2004).
6. Baghdoyan, S. *et al. Nucleic Acids Res.* **32**, e77 (2004).
7. Lum, L. *et al. Science* **299**, 2039–2045 (2003).
8. Foley, E. & O'Farrell, P.H. *PLoS Biol* **2**, E203 (2004).
9. Kamath, R.S. *et al. Nature* **421**, 231–237 (2003).
10. Boutros, M. *et al. Science* **303**, 832–835 (2004).
11. Paddison, P. J. *et al. Nature* **428**, 427–431 (2004).
12. Berns, K. *et al. Nature* **428**, 431–437 (2004).
13. Zheng, L. *et al. Proc. Natl. Acad. Sci. USA* **101**, 135–140 (2004).
14. Tong, A. H. *et al. Science* **303**, 808–813 (2004).

SNPs made routine

Karl W Broman & Eleanor Feingold

With the sequencing of the human genome, millions of single-nucleotide polymorphisms, or SNPs, have been discovered and can be used as markers to identify genes contributing to common human diseases. Two large sets of SNPs have now been organized in panels for high-throughput genotyping.

Two new platforms for high-throughput genotyping of single-nucleotide polymorphisms (SNPs) are presented in this issue of *Nature Methods*. These new technologies promise very fast, high-quality data, and have the potential to improve our ability to map genes for common human diseases by either linkage or association. Murray *et al.*[1] report on a panel of approximately 4,700 SNPs for use in human genetic linkage studies, in which researchers seek to identify an association *within families* between a disease trait and the pattern of inheritance of DNA at marker loci. Matsuzaki *et al.*[2] report on a complementary pair of chips for the simultaneous genotyping of over 100,000 SNPs that could potentially be useful for genome-wide association studies, in which researchers seek to establish an association *across a population* between individuals' disease status and their marker genotypes. Genetic linkage provides lower-resolution mapping than does association analysis, as the extent of association along a chromosome is broader within a family than across a population, but for the same reason, several orders of magnitude fewer markers are needed for linkage than for association mapping.

Short tandem repeat polymorphisms (STRPs, also known as microsatellites) have been the most commonly used genetic marker for human genetic linkage studies over the last decade. STRPs are differences in the lengths of repeats of motifs such as CA or GATA, and typically have many alleles; SNPs are single base differences, and generally have just two alleles. As a result, an individual STRP marker is generally more informative than a SNP, and so linkage analysis with STRPs may be done with fewer markers; a typical genome scan with STRPs includes just 400 markers.

Murray *et al.* genotyped nearly 500 individuals in 28 large families with their 4,700-SNP assay, and used these data to estimate a genetic map for the markers. Although the map itself provides little new biological information, because it is of lower resolution than the earlier map produced by deCODE Genetics Inc.[3], these data allow the identification of genotyping errors that can not be identified by other means, and provide relatively precise estimates of genotyping error rates. Using this information, Murray *et al.* report extremely low rates of genotyping errors, on the order of 1/10,000, and low rates of missing data. In comparison, the genotyping error and missing data rates for STRPs are 50–100 times higher[4].

Matsuzaki *et al.* report similarly low error

Karl W. Broman is in the Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. Eleanor Feingold is in the Departments of Human Genetics and Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.
e-mail: kbroman@jhsph.edu

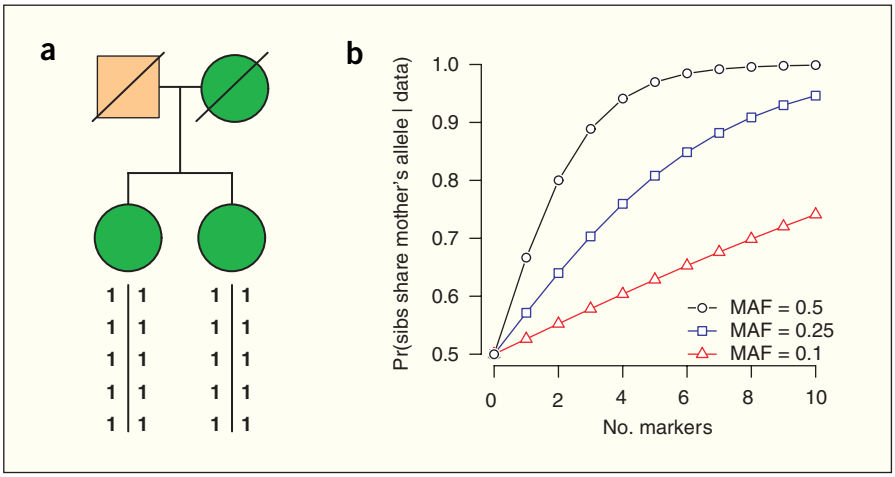

Figure 1 | The essential calculation in human genetic linkage analysis is that of the probability that two relatives share an ancestral allele, given the available marker genotype data. Shown here is the probability (Pr) that two siblings received the same allele from their mother, given that they are both homozygous for the more frequent allele at a set of tightly linked SNPs, as a function of the number of SNPs and the frequency of the less frequent allele (the minor allele frequency, MAF) at each SNP. This assumes linkage equilibrium between markers. In the case of complete linkage disequilibrium, multiple markers provide no more information than does a single marker, and incorrectly assuming linkage equilibrium will cause one to overestimate the information in the data. For example, in the case of four SNPs in complete disequilibrium and having equally frequent alleles, the true probability is 67%, but with the assumption of linkage equilibrium, one would calculate a probability of 94%.

rates for their SNP panel. However, they genotyped a small number of individuals in replicate and 60 parents-offspring trios, and not all genotyping errors may be identified with such data. Although their estimate of the average error rate is small, the true error rate is likely to be larger, and the estimated error rates for individual SNPs are imprecise.

SNPs for linkage

Use of the 4,700-SNP panel of Murray *et al.* could dramatically increase the efficiency of human genetic linkage studies. The time for genotyping is reduced from weeks or months to just days, and, though the analysis of STRP genotype data must be preceded by time-consuming and tedious work to identify and resolve genotyping errors, the tiny error rate associated with this new technology suggests that data cleaning will require little effort: there will be few errors to find, and if an error is detected for a particular SNP in a particular family, one can simply omit the marker completely, with little loss of information.

This is not the first high-throughput SNP genotyping assay for human linkage studies. A 10,000-SNP assay[5,6], based on technology similar to that of the 100,000 SNP-assay of Matsuzaki *et al.*, is available and has been applied to linkage studies of rheumatoid arthritis[7] and bipolar disorder[8]. But the panel produced by Murray *et al.* seems to give more complete and higher-quality data.

Murray *et al.* report that their 4,700-SNP panel provides considerably more information than the commonly used STRP panels of around 400 markers, and that it is more informative than a scan with 800 STRPs. This is exciting, but should be considered with caution; in many gene mapping studies it is not marker information but rather study size that limits mapping power and precision. In addition, users should be aware that the information content of a panel depends on the pedigree structure and even on the underlying etiology of the disease.

For the full promise of this new panel to be realized, several important issues of statistical methodology will need to be resolved. Single-point linkage analysis (sometimes called two-point linkage analysis), in which each marker is considered one at a time, will no longer be usable, as, unlike an STRP, a single SNP provides little information in isolation. Multipoint methods, which combine the information from multiple markers, are in routine use, but there are serious computational limitations on how many markers can be considered simultaneously as the pedigree size increases. Moreover, all current multipoint linkage software assumes linkage equilibrium between markers (that is, that the frequency of multilocus genotypes may be obtained by multiplying the frequencies of individual marker genotypes). In the context of a dense set of SNPs showing linkage disequilibrium, this assumption can lead to overestimation of the information in the marker data (**Fig. 1**), which could cause either a loss of power or a spurious linkage.

SNPs for genome-wide association

Eight years ago, Risch and Merikangas[9] proposed the use of whole-genome association studies for common human diseases; the 100,000-SNP panel of Matsuzaki *et al.* could allow that concept to be applied in practice. Although 100,000 SNPs is probably not enough to guarantee a SNP in association with every causal variant, it is probably sufficient to find some variants affecting a typical complex disease (with adequate study design and sample size, of course). We will probably not know how many SNPs are needed for a fully powered whole-genome association study until it has been attempted multiple times.

Key statistical issues in whole-genome association studies, including how to adjust for multiple testing and whether to test individual markers or haplotypes, are well known, and a variety of strategies have been proposed. With whole-genome, high-density SNP genotypes now feasible, there will be an opportunity to study the appropriateness of these strategies with actual data.

In summary, these fast, accurate SNP genotyping assays are exciting and provide good cause for celebration by gene mappers; these new technologies, once accompanied by similar advances in tools for statistical genetic analyses, will dramatically increase the rate of discovery of the complex etiology underlying common human diseases. However, study design, subject recruitment, disease phenotype definition, and the identification, measurement and proper account of environmental risk factors will remain critical.

1. Murray, S.S. *et al. Nat. Methods* **1**, 113–117 (2004).
2. Matsuzaki, H. *et al. Nat. Methods* **1**, 109–111 (2004).
3. Kong, A. *et al. Nat. Genet.* **31**, 241–247 (2002).
4. Weber, J.L. & Broman, K.W. in *Genetic Dissection of Complex Traits* (eds. Rao, D.C. & Province, M.A.) 77–96 (Academic Press, New York, 2001).
5. Kennedy, G.C. *et al. Nat. Biotechnol.* **21**, 1233–1237 (2003).
6. Matsuzaki, H. *et al. Genome Res.* **14**, 414–425 (2004).
7. John, S. *et al. Am. J. Hum. Genet.* **75**, 54–64 (2004).
8. Middleton, F.A. *et al. Am. J. Hum. Genet.* **74**, 886–897 (2004).
9. Risch, N. & Merikangas, K. *Science* **273**, 1516–1517 (1996).