ELSEVIER

European Journal of Cancer 40 (2004) 2537–2543

European Journal of Cancer

www.ejconline.com

The human genome project and the discovery of genetic determinants of cancer susceptibility

R. Taramelli *, F. Acquati

Dipartimento di Biologia Strutturale e Funzionale, Universita Insubria, Via JH Dunant 3, 21100 Varese, Italy

Received 30 January 2004; received in revised form 4 June 2004; accepted 7 July 2004
Available online 17 September 2004

Abstract

The Human Genome Project has recently provided a great deal of information on the sequence that comprises the human genome. We are now in the process of structuring and deciphering the 3×10^9 base sequence in order to gain insights into its functional role. Several efforts are focusing on the search for DNA sequence variations underlying common/complex diseases that constitute a real burden in terms of public health measures. As expected, the genetic architecture of these complex traits, shows tremendous complexity, and the discovery and characterisation of susceptibility alleles constitute a real challenge for the geneticist. Conceptual and experimental genetic approaches aimed at dissecting the molecular features of susceptibility genes, in particular those predisposing to cancer, are outlined and discussed in this reveiw.

Keywords: Gene; Gene interaction; Complex diseases; Linkage disequilibrium; Association; SNPs; Cancer; Susceptibility; Mouse; QTL

1. Introduction

The complete sequence determination of the human genome has given rise to several challenging questions, in particular with regard to the genetic basis underlying common or complex diseases. As is widely recognised, these disorders are, from a genetic point of view, the most difficult to approach in that a complex architecture of genes, gene interactions, environment and gene-environment relationships are associated with the full manifestation of the disease phenotype [1–5].

One of the most valuable outcomes of completing the human genome sequence was that the demand for larger data-sets concerning the range of genetic differences among individuals could be met.

We have in common 99% of our genome, but the 1% difference is crucial because it harbours the base changes that are associated with disease risk or with a predisposition to pathological states [6]. It is therefore logical that much of the present effort should be focused on the identification of the many naturally occurring polymorphic genetic variants that determine an individual's susceptibility to a range of common diseases, including cancer [5–7]. Indeed, cancer can be considered a complex genetic trait in which genes and the environment interact, conferring a variable degree of risk on people who inherit predisposing alleles [8]. A great deal of genetic information has already been obtained through the study of high-penetrance genes that are responsible for the striking familial aggregation of cancer (i.e., the *BRCA* mutations and their association with breast cancer), but, unfortunately, these genes can only explain a fraction of the genetic component of risk [8]. A more challenging problem is to isolate and characterise the genes that contribute only modestly to cancer risk, as is the case for the vast majority of the genes involved in common malignancies [8–10]. The latter statement highlights the real and crucial problem in the study of complex diseases, namely the formidable task of assessing and capturing

* Corresponding author. Tel.: +39 332 421514; fax: +39 332 421500.
E-mail address: roberto.taramelli@uninsubria.it (R. Taramelli).

0959-8049/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2004.07.030

the genetic contributions (likely to be modest) underlying conditions in which 'environmental factors', in broader terms, are playing a crucial role.

This Review will focus on what the Human Genome Project has brought to the study of common cancers, which theoretically and practically can be interpreted as a complex disease, with particular attention given to the identification of predisposing genes.

We envisage that purpose two main issues are relevant and worth considering: (1) the systematic search for the many naturally occurring polymorphic genetic variants, coupled with the assessment of their direct (causally related to the phenotype) or indirect (in linkage or linkage disequilibrium with other causative variants influencing the phenotype) role in determining cancer risk; and (2) the exploitation of the mouse as a model organism for the study of cancer.

2. Assessing genetic variation

The most abundant sequence differences between individuals are the single-nucleotide polymorphisms (SNPs), which are responsible for most of the observed variability in typical sequence surveys [6,7]. It has been estimated that there are approximately 10 million SNPs with allelic frequencies above 1% [11], and that these 10 million common SNPs represent 90% of the variation in the populations surveyed; the remaining 10% is due to variants that are rare [10,11]. SNPs are classified according to the nature of the nucleotide that is affected. Non-coding SNPs are located in a 5′ or 3′ non-transcribed region, in a 5′ or 3′ untranslated region, in an intron, or they may be intragenic. By contrast, coding SNPs (or cSNPs) may be replacement polymorphisms (they change the amino acid that is encoded) or synonymous polymorphisms (they change the codon, but not the amino acid). Non-replacement polymorphisms include synonymous and non-coding polymorphisms, but many of these may affect gene function through an effect on transcriptional or translational regulation, splicing or RNA stability [3].

What is relevant for genetic analysis is the combination of SNPs that is inherited together on the same DNA strand to form haplotypes [12]. The latter represent a sort of structured architecture of the human genome, with the result that several haplotypes are similar and of the same order in many people from different populations around the world [13].

3. Genetic methodologies: candidate gene- and linkage disequilibrium-based approaches

Currently, many research studies in human genetics are focused on finding associations between SNPs or haplotypes and cancer risk [14]. Two methodological approaches are commonly used:

(1) The 'candidate gene' approach analyses uses association studies involving cases and properly matched controls to identify individual cSNPs within candidate genes that could contribute to cancer susceptibility [3]. This approach is driven by a biological model in which candidate genes, pathways and interactions between genes and environment are fairly well known. The recent implementation of three highly integrated cancer projects fuelled by the Human Genome Project (Cancer Genome Anatomy Project; Human Cancer Genome Project; Cancer Genome Project) [15] is the main source for the discovery and characterisation of such cancer candidate genes.

Despite intensive efforts in the last decade, the identification of common, but 'low penetrance', susceptibility genes predicting risks for cancer and other complex diseases has been rather disappointing [16]. Indeed, results from several studies are mostly inconclusive and, more importantly, difficult to replicate [17]. Lack of optimal designs, quality and size are often responsible for the inconsistency of the results [18–21]. Moreover, and importantly, as the environment has an obligate role, a careful assessment of exposure to minimise any gene-environment interplay has only rarely been performed.

(2) In the other approach, proposed by authors in Refs. [22–24] and known as the indirect approach, a set of SNPs is used as genetic markers to detect any association between a genomic region and a disease, irrespective of whether the markers themselves have any functional effect. This has the advantage of narrowing down to a limited region the search for a causative variant (SNP) of the disease. This approach makes use of whole-genome scans as a crucial tool for mapping variants associated with disease and relies heavily on linkage disequilibrium among various SNPs comprising different haplotypes. Another assumption, implicit in this indirect approach, is that approximately 90% of sequence variation among individuals is due to common variants that have arisen as a result of historical mutational events. These are therefore associated with close variants that were present on the ancestral chromosome on which the mutation occurred [24,25]. Such associations should be highly efficient in identifying variants in candidate genes, chromosome regions or even in the whole genome. These approaches have provided the rationale underlying the International HapMap Project, which can be considered a logical extension of the Human Genome Project. Indeed, the aim of this Project is to determine common patterns of sequence variations in the human genome and establish correlations between sequence variants from genomes of people from African, Asian and European ancestry. By determining the genotype of more than a million sequence variants, the hope

is to discover which of these variants affects common diseases [25].

4. Alternative approaches

This 'titanic' endeavour has attracted several critics due to the contrasting views that exist among geneticists concerning how the genetic architecture underlies common diseases. According to the proponents of the HapMap Project, it is assumed (but not yet examined) that most disease-susceptibility variants are common in the population (with a frequency higher than 0.01). Thus, this hypothesis, also known as the common disease/common variant hypothesis (or CD/CV), proposes that individuals with disease have an excess of predictable and common susceptibility alleles, which are also neutral with respect to selection. Hence, these disease-associated alleles are potentially detectable in large-scale, patient-control association analyses [24,25].

However, according to the opponents of this theory, the CD/CV hypothesis misrepresents the very peculiar nature of common/complex diseases. These traits have what Kenneth Weiss and Joseph Terwilliger call a 'low detectance' – the fairly low probability of possessing any particular predisposition genotype, given a particular disease or trait phenotype [26]. The essence of this proposition is that common diseases originate from the interaction of many genetic and environment determinants, making correlations with any one of them very weak [26–33]. Additionally, it is largely established that there is an inverse relationship between the magnitude of the genetic effect and allele frequency at a certain locus [34], implying that pathologically relevant variants would be expected to be rare in human populations. Furthermore, a recent modelling-based paper has predicted that neutral susceptibility alleles contribute little to the overall genetic variance for a certain disease [27]. It seems likely that the bulk of genetic variance underlying complex diseases is due to alleles under weak selection at loci characterised by an elevated mutation rate [27]. The logical conclusion from these predictions is that allelic heterogeneity is going to be extensive, making association studies rather difficult to carry out [27–35]. By contrast, if we consider the situation shown by the so-called late-onset Mendelian disorders, where the main feature is a broad spectrum of allelic diversity, we realise that the challenge posed by complex diseases is going to be formidable. For instance, of the more than 400 alleles at the *BRCA2* locus, only six are common and among them just one is associated with a modest risk of breast cancer. In another Mendelian condition, premature coronary artery disease due to familial hypercholesterolaemia, more than 700 different alleles have been described at the low-density lipoprotein receptor, although none of them are commonly associated with relevant clinical features. Table 1 shows some examples of allelic heterogeneity in a few common diseases.

With regard to complex and late-onset diseases, such as cancers, we are therefore dealing with high allele diversity, suggesting negative selection and low allele frequencies in the founder populations [31,35]. A final, but important point to be considered has to do with the crucial concept of locus heterogeneity, which plays a major role in these diseases [31]. Indeed, the presence of more than one locus contributing to the disease risk is perhaps the most difficult obstacle to overcome in association studies. Further, the apparently 'simple' Mendelian diseases are infact 'complex', in terms of locus heterogeneity [36–38]. It is therefore not unreasonable to postulate for common diseases an even more subtle complexity, where several disease loci interact in various combinations in different people, making each of us a distinct genetic entity. This poses a real challenge for linkage disequilibrium-based approaches.

These considerations described above form the conceptual basis for the so-called common diseases/rare variant (CD/RV) or multilocus/multiallele hypothesis [26–33,35], which clearly contrasts with the CD/CV hypothesis. At present, the genetic community is split over these two contrasting views [39,40], which can be summarised as follows: are the susceptibility alleles common or rare, neutral or deleterious, few or many? According to which of the different philosophies is embraced, the methodological approaches followed will clearly differ: the proponents of the CD/CV hypothesis think the HapMap Project is a priority and they expect that the Project, together with improvements in statistical tools for analysing genome-wide scans, will allow the isolation and characterisation of the genetic determinants of common diseases. The proponents of the CD/RV hypothesis are taking a different route to address the problem, which relies more on genetic and biological insights than on technology [26–35]. They recommend, for instance, the study of traits that have a clear genetic basis and families with one or more members presenting the trait are considered to be the appropriate starting point [26,31]. The greater the familial correlation, the

Table 1
Allelic heterogeneity for genes involved in common human diseases

Disease	Locus	Number of disease-associated alleles
Cardiovascular	*LDLR*	>700
	APoB100	>20
Cancer	*BRCA1*	>500
	BRCA2	>400
	MLH1	>150
	MSH2	>100
	TP53	>150
Obesity/metabolic	*MC4R*	>30
	CFTR	>950

more plausible is the genetic effect. Age of onset is another important clue and early-onset adult cancers should be identified and studied, as they are likely to have a genetic basis [26,31]. Subclinical phenotypes, which are simpler than the disease itself, should be sought, and disease severity noted [26,31]. It is important to realise that there is not a simple one-to-one correspondence with a specific gene and 'cancer' (with a few notable exceptions such as some of the childhood cancers); rather, genes might be associated with certain features of carcinogenesis, such as DNA repair/metabolism, progression, metastasis, particular histological variants or exposures causally linked to cancer, such as tobacco [8–10]. In other words, it is more fruitful, to seek associations with conditions related to the disease rather than to the disease itself.

Furthermore, to study isolated subpopulations in which linkage disequilibrium is high and aetiology homogeneous [26,31,41–43], or ethnic groups with different disease prevalences, despite similar environmental exposures, could be more rewarding [26,31,43,44]. A reduction on the environmental heterogeneity of the study group, or adequate exposure measurements are needed for association studies aimed at finding these cancer susceptibility genes.

5. Will the mouse help?

The Human Genome Project has also fuelled the study of genomes from other species, using animal models, phylogenetic sequence comparisons and other approaches, to characterise functionally human coding and non-coding sequences, should increase our knowledge in the genetic determinants of cancer susceptibility.

In the mouse the processes of cancer development, both biologically and genetically, are remarkably similar to those in humans [45]. Indeed the mouse is considered a good model organism for the study of the influence on cancer risk of environment and genetic background [46,47], both of which strongly affect tumour susceptibility in mice and men [46–49].

Moreover, the search for low-penetrance susceptibility genes in the mouse (also known as QTLs) [50,51] is facilitated by the presence of resistance/susceptibility strains whose phenotype is genetically controlled by multiple loci, some of which have already been mapped in the mouse genome [49,52]. Through positional cloning, these genes can be cloned and human homologues sought, thereby facilitating careful candidate studies in humans.

The exploitation of the mouse as a tool for identifying low-penetrance susceptibility genes is certainly crucial given the difficulties and obstacles encountered in studies focused on human populations. The 'armoury' available to the mouse geneticist is remarkable: dozens of inbred strains, outbred mice with different evolutionary histories, recombinant inbred, recombinant congenics, genome-tagged mice [46–49], transgenic, knock-out and knock-in mice [53–56]. Furthermore, the power of mouse genetics lies in the possibility of generating crosses between members of strains that are highly susceptible or highly resistant to cancer and eventually following the genes involved in the corresponding phenotypes.

Although many mouse tumour-susceptibility QTLs have been mapped (more than a hundred) [49,52], only a few of them have been cloned and even fewer with mutations conferring altered risk have been identified [48]. To date, four cancer-susceptibility genes have been identified, either by positional cloning and/or candidate-gene analysis [49,52]. The first was *Pla2g2a* coding for a member of the phospholipase A_2 family, which is one of the two genes spanning the Mom1 region [57–59]. This region is known to be associated with modifications of the germline *Apc* mutations that induce intestinal tumorigenesis [57]. A frameshift mutation of *Pla2g2a* has a strain distribution identical to that of the Mom1-susceptible allele [58]. However, when the human homologue was utilised for association studies to detect possible variant associations with the risk of colorectal cancer, no significant association was found [60]. This

Table 2
Mapped mouse cancer modifier loci

Locus (Gene)	Chromosome	Affected tissue
Scc3	1	Colon
Pas8	1	Lung
Skts8	1	Skin
Hcf2	1	Liver
Scc1 (*Ptprj*)	2	Colon
Sluk16	2	Lung
Skts13 (*Aurora*)	2	Skin
Moml (*Pla2g2a*)	4	Intestine
Hcr1	4	Liver
Pctr1 (*Cdkn2a*)	4	Plasmacytoma
Skts7	4	Skin
Scc13	6	Colon
Pas1 (*Lrmp or M22Rik*)	6	Lung
Sluc 19	7	Lung
Tlsm1	7	Lymphoma
Sluc9	8	Lung
Pas4	9	Lung
Scc9	10	Colon
Gct2	12	Ovary
Sluc25	15	Lung
Apmt1	15	Breast
Skts9	16	Skin
Lts	17	Lung
Sluc14	18	Lung
Pas3	19	Lung
Gtc4	X	Ovary

Loci for which the underlying cancer-susceptibility gene has been identified are marked in bold and the corresponding gene symbol is bracketed.

result highlights one of the possible limitations of using model organisms to model human pathology, i.e., that some gene effects are indeed species-specific. Another susceptibility gene is *Pthlh* [61], whose The_{166}-Pro amino acid polymorphism is linked with high and low susceptibility to skin carcinogenesis [61]. In addition, *Cdkn2a* encoding p16ink4a, the cyclin-dependent kinase inhibitor 2A [62], is a candidate for Pctr1, a locus affecting plasmocytoma susceptibility [63]. A coding polymorphism in this gene also distinguishes susceptible (BALB/c) from resistant (DBA/2) strains [64]. The human counterpart, *CDKN2A*, is mutated in familial melanoma, breast and pancreatic tumours [65]. Interestingly, a germline mutation in this gene is associated with a case of multiple myeloma that is supposed to be the mouse counterpart of plasmocytoma [66].

The fourth gene for which some information is available is *Ptprj*. This gene codes for a protein, tyrosine phosphatase receptor type 1, and is a candidate gene for the mouse Scc1 region that is affected and associated with colon tumorigenesis [67]. Amino acid substitutions in this protein allow discrimination between cancer-resistance and -susceptibility alleles [68]. Table 2 summarises some of the mouse cancer QTL according to chromosome location and type of tumour affected.

There are only few studies in this area because of several obstacles to gene isolation. First, all of the current genetic analyses allow mapping to intervals of 1–10 cM, sizes which are prohibitively large for gene isolation. Major technological gene-discovering approaches, including information from the mouse genome sequence and transcriptome analysis using microarrays as well as haplotype mapping, should help to overcome at least some of these drawbacks. Secondly, the limited genetic variation (locus variability) between the different inbred strains available with respect to their wild-type populations is well recognised. Thirdly, some, if not most, of the genetically mapped susceptibility loci seem to harbour several independently acting genes, which, together with the existence of clustered loci, represents a major problem for the ultimate identification of candidate genes [48,49]. Fourthly, even if we assume that the number of mouse susceptibility genes will increase, there will always be problems concerning the different life-spans of mice and men, the organ and site-specificity, and, above all, the fairly different exposures (far more complex in humans) to which the two species are subjected. Finally, the most common mapping strategies in the mouse (recombinant inbred, congenic strain etc.) are rather time-consuming and expensive.

6. Conclusions

As outlined above, geneticists use three main conceptual and experimental approaches to determine how millenia of evolution have helped shape complex diseases such as cancer.

These include the exploitation of the mouse as a cancer model organism, and two conflicting genetic approaches. One relies on high-throughput approaches, including improvements in genomic technology and the implementation of more powerful statistical tools for wide genome-scanning projects. The other is centred more on biological and genetic insights, focusing on a careful ascertainment of the study population and families in order to maximise real genetic determinants of risk. Furthermore, its supporters emphasise the importance of the complex correlations between genotypes and phenotypes and propose that simple, mathematically tractable models might be totally misleading in trying to gain a full understanding of the complexity of both allelic and non-allelic heterogeneity, interactions between genes and between genes and the environment, epistasis, pleiotropy and variable expression of different alleles even in the same gene.

We envisage that, even with the inevitable drawbacks of this latter approach, it is by far the most suitable to use to gain a realistic understanding of the biology that underlines complex diseases, nevertheless, this approach undoubtedly represents a real conceptual and experimental challenge for the genetic community in future research studies.

Conflict of interest statement

None declared.

References

1. King RA, Rotter JI, Motulsky AG. *The genetic basis of common diseases*. In Motulsky AG, Harper FS, Scriver C, Bobrow M. eds. vol. 20. Oxford University Press, 1992.
2. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*, 1516–1517.
3. Risch N. Searching for genetic determinants in the new millennium. *Nature* 2000, **404**, 847–856.
4. Lander E, Schork NJ. Genetic dissection of complex traits. *Science* 1994, **265**, 2037–2048.
5. Weiss KM. Perspective: in search of human variation. *Genome res*.
6. Sachidanandam R *et al*. A map of human sequence variation containing 1.42 millions single nucleotide polymorphisms. *Nature* 2001, **409**, 928–933.
7. Chakravarti A. It's raining SNPs, hallelujah. *Nat Genet* 1999, **19**, 216–217.
8. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. *Nat Genet* 2003, **33**, 238–244.
9. Rothman N, Wacholder S, Caporaso N, Garcia-Closas, Beutow JR, Fraumeni JR. The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim Biophys Acta* 2001, **1471**, C1–C10.
10. Caporaso N, Golstein A. Cancer genes: single and susceptibility: expressing the difference. *Pharmacogenetics* 1995, **5**, 59–63.
11. Kruglyak L, Nickerson D. Variation is the spice of life. *Nat Genet* 2001, **23**, 234–236.

12. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA. *Nat Genet* 2003, **33**, 518–521.
13. Wall JD, Pritchard K. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 2003, **4**, 587–597.
14. Webb T. SNPs: can genetic variants control cancer susceptibility?. *J Natl Cancer Inst* 2002, **94**, 476–479.
15. Strauberg RL, Simpson AJ, Wooster R. Sequence-based cancer genomics: progress, lessons and opportunities. *Nat Rev Genet* 2003, **4**, 408–418.
16. Hirschhorn JN, Lohmueller K, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002, **4**, 45–61.
17. Garcia-Closas M, Rothman N, Lubin J. Misclassification in case control studies of gene-environment interactions: assessment of bias and sample size. *Cancer Epidemiol Biomark Prev* 1999, **8**, 1043–1050.
18. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomark Prev* 2002, **11**, 513–520.
19. Thompson JR. Invited commentary: re: multiple comparisons and related issues in the interpretation of epidemiological data. *Am J Epidemiol* 1998, **147**, 801–806.
20. Little J, Bradley L, Bray MS, *et al*. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002, **156**, 300–310.
21. Wacholder S, McLaughlin J, Silverman D, Mandel JS. Selection of controls in case-control Studies! Principles. *Am J Epidemiol* 1992, **135**, 1019–1049.
22. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997, **278**, 1580–1581.
23. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past success for mendelian diseases, future approaches for complex diseases. *Nature Genet* 2003, **33**, 228–237.
24. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001, **17**, 502–509.
25. The International HapMap Consortium. The international HapMap project. *Nature* 2003, **426**, 789–796.
26. Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs. *Nat Genet* 2000, **26**, 151–157.
27. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases?. *Am J Hum Genet* 2001, **69**, 124–137.
28. Pritchard JK, Cox N. The allelic architecture of human disease genes: common disease-common variant o not?. *Hum Mol Genet* 2002, **11**, 2417–2423.
29. Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality?. *Curr Opin Biotechnol* 1998, **9**, 578–594.
30. Terwilliger JD, Haghighi F, Hiekkalinna TS, Goring HH. A biased assessment of the use of SNPs in human complex traits. *Curr Opin Genet Dev* 2002, **12**, 726–734.
31. Wright AF, Hastie N. Complex genetic diseases: controversy over the Croesus code Genome Biology.
32. Clark AG. Finding genes underlying risk of complex diseases by linkage disequilibrium mapping. *Curr Opin Genet Devel* 2003, **13**, 296–302.
33. Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 2002, **18**, 19–26.
34. Morton NE. Significance levels in complex diseases. *Am J Hum Genet* 1998, **62**, 690–697.
35. Wright AF, Charlesworth B, Rudan I, Carothers A, Campbell H. A polygenic basis for late-onset disease. *Trends Genet* 2003, **19**, 97–106.
36. Badano JL, Katsanis N. Beyond Mendelian evolving view of human genetic disease transmission. *Nat Rev Genet* 2002, **3**, 779–789.
37. Dipple KM, McCabe RB. Phenotypes of patients with simple mendelian disorders are complex traits: thresholds, modifiers and systemic dynamics. *Am J Hum Genet* 2000, **66**, 1729–1735.
38. Scrivers CR, Waters PJ. Monogenic traits are not simple. *Trends Genet* 1999, **15**, 267–272.
39. Pennisi E. A closer look at SNPs suggests difficulties. *Science* 1998, **281**, 1787–1789.
40. Couzin J. New mapping project splits the community. *Science* 2002, **296**, 1391–1392.
41. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* 2000, **72**, 63–132.
42. Terwilliger JD, Zollner S, Laan M, Paabo S. Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'drift mapping' in small populations with no demographic expansion. *Hum Hered* 1998, **48**, 138–154.
43. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet* 1999, **23**, 397–404.
44. Wright AF, Carothers AD, Campbell H. Gene-environment interactions – the bioBank UK study. *Pharmacogenom J* 2002, **2**, 75–82.
45. Balmain A, Harris CC. Carcinogenesis in mouse and human cells: parallels and paradoxes. *Carcinogenesis* 2000, **21**, 371–377.
46. Balmain A. Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. *Cell* 2002, **108**, 145–152.
47. de Koning JP, Mao JH, Balmain A. Novel approaches to identify low-penetrance cancer susceptibility genes using mouse models. *Recent Results Cancer Res* 2003, **163**, 19–27.
48. Mao JH, Balmain A. Genomic approaches to identification of tumor-susceptibility genes using mouse models. *Curr Opin Genet Dev* 2003, **13**, 14–19.
49. Demant P. Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nat Rev Genet* 2003, **4**, 721–734.
50. Flint J, Mott R. Finding molecular basis of quantitative traits: successes and pitfalls. *Nat Rev Genet* 2001, **2**, 437–445.
51. Barton NH, Keightley PD. Understanding quantitative genetic variation. *Nat Rev Genet* 2002, **3**, 11–17.
52. Dragani T. 10 years of mouse cancer modifier loci: human relevance. *Cancer Res* 2003, **63**, 3011–3018.
53. Hann B, Balmain A. Building 'validated' mouse models of human cancer. *Curr Opin Cell Biol* 2001, **13**, 778–784.
54. Van Dyke T, Tyler J. Cancer modeling in the modern era: progress and challenges. *Cell* 2002, **108**, 135–144.
55. Tuveson DA, Tyler J. Technologically advanced cancer modeling in mice. *Curr Opin Genet Dev* 2002, **12**, 105–110.
56. Meuwissen R, Jonker J, Berns A. Mouse models of sporadic cancer. *Exp Cell Res* 2001, **264**, 100–110.
57. MacPhee M, Chepenik KP, Liddell RA, Nelson KK, Siracusa AM, Buchberg AM. The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of ApcMin-induced intestinal neoplasia. *Cell* 1995, **81**, 957–966.
58. Cormier RT, Hong KH, Halberg RB, *et al*. Secretory phospholipase Pla2g2a confers resistance to intestinal tumorigenesis. *Nat Genet* 1997, **17**, 88–91.
59. Cormier RT, Bilger A, Lillich AJ, Halberg RB, Hong KH, Gould KA, *et al*. The Mom1AKR intestinal tumor resistance region consists of Pla2g2a and a locus distal to D4Mit64. *Oncogene* 2000, **19**, 3182–3192.
60. Tomlinson JP, Becknd NE, Neale K, Bodmer WF. Variants at the secretory phospholipase A2 locus: analysis of association with familial adenomatous polyposis and sporadic colorectal tumors. *Am J Hum Genet* 1996, **60**, 369–376.
61. Manenti G, Peissel B, Gariboldi M, *et al*. A cancer modifier role for parathyroid hormone-related protein. *Oncogene* 2000, **19**, 5324–5328.

62. Potter M, Mashinsky EB, Wax JS, Hartley J, Mock BA. Identification of two genes on chromosome 4 that determine resistance to plasmocytoma induction in mice. *Cancer Res* 1994, **54**, 969–975.
63. Zhang Z, Wang Y, Herzog CR, Liu G, Lee HW, DePinho RA, *et al*. A strong candidate gene for the Papg1 locus on mouse chromosome 4 affecting lung tumor progression. *Oncogene* 2002, **21**, 5960–5966.
64. Lynch HT, Brand RE, Hogg D, *et al*. Phenotypic variation in eight extended CDKN2A germline mutation familial atypical multiple mole melanoma-pancreatic carcinoma-prone families: the familial atypical mole melanoma-pancreatic carcinoma syndrome. *Cancer* 2002, **94**(1), 84–96.
65. Borg A, Sandberg T, Nilsson K, *et al*. High frequency of multiple melanomas and breast and pancreas carcinomas in CDKN2A mutation-positive melanoma families. *J Natl Cancer Inst* 2000, **92**(15), 1260–1266.
66. Mock BA, Krall MM, Dosik JK. Genetic mapping of tumor susceptibility genes in mouse plasmacytomagenesis. *Proc Natl Acad Sci USA* 1993, **90**, 9499–9503.
67. Moen CJ, Groot PC, Hart AA, Snoek M, Demant P. Fine mapping of colon tumor susceptibility (Scc) genes in the mouse, different from the genes known to be somatically mutated in colon cancer. *Proc Natl Acad Sci USA* 1996, **93**, 1082–1086.
68. Ruivenkamp CA, van Wezel T, Zanon C, *et al*. Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet* 2002, **4**, 295–300.