Molecular Phenotyping of Mouse Mutant Resources by RNA Expression Profiling

J. Beckers[1]*, J. Hoheisel[2], W. Mewes[3], M. Vingron[4], and M. Hrabé de Angelis[1]

[1]GSF-Research Center for Environment and Health, Institute of Experimental Genetics, Ingolstädter Landstr.1, 85764 Neuherberg, Germany, [2]German Cancer Research Center, Functional Genome Analysis, 69120 Heidelberg, Germany, [3]GSF-Research Center for Environment and Health, Institute for Bioinformatics, Ingolstädter, Landstr. 1, 85764, Neuherberg, Germany, [4]MPI of Molecular Genetics, Theoretical Bioinformatics, 14195 Berlin, Germany

Abstract: Microarray-based techniques allow us to visualize and quantify the expression of every single gene in any population of cells. In yeast the true potential of large-scale transcriptome analysis in identifying regulatory units and understanding gene function has already been demonstrated by evaluating expression profiles of a comprehensive group of mutants. We discuss the potential of DNA-chip technologies for the analysis of gene expression in complex organisms. The usefulness of transcriptome analysis for clinical purposes und diagnosis of cancers is already well established. We argue that microarray-based expression profiling will also be a useful tool for the analysis of gene function and approaches complementary to classical phenotypic description in mammals, particularly in regard of the large resources of mutant models that are currently being generated by gene-targeting and mutagenesis of the mouse genome. Experimental requirements and potential future directions are discussed.

FROM MOUSE MUTANTS TO GENE FUNCTION

The human genome has almost completely been sequenced [1-**3] and the sequence of the mouse genome will be completed within the next years. A large fraction of genes is already mapped to their chromosomal region and the precision of such genetic maps is improving continuously (for review, e.g., [**4]). The focus of genomic research in the near future will be the systematic analysis of gene function within the complexity of the organism. This process will add meaning and interpretation to the genomic sequence. Driving force of this development is the interest in a better understanding of the basis of human genetic diseases, of biology and mammalian development.

The mouse is, for several reasons, the major model organism for human genetic disease and mammalian, developmental genetics [5, **6]. In its physiology and development, the mouse is very similar to humans. Second, extensive comparative linkage maps are available for mouse and man, and syntenic regions of both organisms are already described. More than 70.000 mouse UniGene clusters from ESTs (www.ncbi.nlm.nih.gov/UniGene/Mm.Home.html, www.dkfz-heidelberg.de/tbi/services/GeneNest) and full length cDNA clones (http://genome.rtc.riken.go.jp/) have been identified that are invaluable in determining gene structure, gene function and in generating radiation hybrid maps [7, 8]. Several projects are ongoing to develop high-throughput genotyping methods that do not require gel-electrophoresis and which are based on single nucleotide polymorphisms (SNPs) of inbred mouse strains [9]. The establishment of SNP maps for sets of different mouse strains will very much simplify the automation of genetic linkage analyses.

Finally, there is at the moment no other vertebrate than the mouse for which such versatile techniques to manipulate the genome are that well developed [10-12]. Homologous recombination in mouse embryonic stem cells allows the deletion or integration of chunks of DNA at almost any known locus in the genome [13, 14]. The generation of transgenic mice, e.g., by injecting DNA in the pronucleus of zygotes allows the ectopic addition of new sequences into the mouse genome - also across species [15]. Mutations may be designed such that they are inducible, e.g., at a particular developmental stage, in specific tissues or by an artificial activator [16-19]. Moreover, trans-allelic targeted meiotic recombination between homologous loci allows the combination of linked alleles as well as interchromosomal unequal exchanges [20].

In addition to such gene based approaches, forward genetic approaches such as chemical mutagenesis using, e.g., N-ethyl-N-nitrosourea (ENU) and insertional mutagenesis with, e.g., the gene trap technique can be performed at large scale and with high efficiency in the mouse [21]. To facilitate the mapping of mutations and to limit new alleles to genomic regions of particular interest, some projects combine the point-mutagen ENU with non-complementation tests using mice with large genomic deletions that may be generated, e.g., by X-irradiation or site-specific recombination [22-24]. Thus, the genetic tricks that are already available make the mouse an excellent model system

*Address correspondence to this author at the GSF – Research Center for Environment and Health, Institute of Experimental Genetics, Ingolstädter Landstr.1, 85764 Neuherberg, Germany; Tel: +49 89 3187 3513; Fax: +49 89 3187 3500; E-mail: beckers@gsf.de

1389-2029/02 $35.00+.00 © 2002 Bentham Science Publishers Ltd.

for functional genomics. For the first time, the systematic and comprehensive analysis of gene function and gene expression mechanisms in a mammal closely related to humans appears feasible.

NETWORKS FOR PHENOTYPING

At present there are still far less mouse mutants available than there are genes and, even more so, than there are informative alleles - a phenomenon that has been termed 'the phenotype gap' [25]. To narrow this gap, some recent projects successfully focus on large-scale phenotype driven approaches to isolate new mutants and allelic series that are identified independent of known genes. Large-scale projects have been initiated at the MRC, Harwell, United Kingdom and at the GSF, Munich, Germany, to systematically mutagenise the mouse genome either by means of chemical mutagenesis or by insertional/gene trap mutagenesis in embryonic stem cells [*26-28]. The implementation of such research facilities is now being extended to institutions in Australia, Japan, Canada, USA and other countries. The concept of these projects, however, is not an invention of mammalian geneticists and the human genome projects, but rather has been pioneered by researches that successfully implemented large scale approaches for the systematic mutagenesis of the genomes of Drosophila [29, 30], Caenorhabditis [31, 32], Arabidopsis [33] and Danio [34, 35]. These projects have been of tremendous value for the understanding of the function of genes within the respective organisms. The mutant resources generated in the past decades were the basis for the fundamental insight that genetic pathways have been remarkably conserved during evolution and that the multiple deployment of these pathways (or networks) for different functions is rather the rule than the exception [36-38].

The major argument in favour of such phenotype based approaches is, that a gene required for (a) particular biological function(s) does not need to be cloned prior to the functional analysis. Instead, new, yet uncharacterised genes and their mutant alleles are identified. In addition, mutagenesis screens may be extended to identify genes with partially redundant or non-essential functions by the implementation of sensitised screens involving classical complementation tests (i.e., crossing a new mutant over a known mutant and examining the phenotype) [39]. Using ENU, which is at the moment the most potent mutagen for the induction of point mutations in mouse spermatogonia, a variety of alleles can be generated from a given gene locus [40]. The synthetic compound generally induces point mutations (mostly A-T transversions and A-G transitions), which potentially range from loss-of-function alleles, over hypomorphic to gain-of-function (hypermorphic) alleles. The availability of such diverse alleles has proven to be very informative, for example, for the analysis of late gene functions or the identification of functional domains of gene products. A major advantage of the gene trap mutagenesis approach, on the other hand, is that the inserted sequence serves as a tag to identify the mutagenised gene.

The several thousand mouse mutants that have been generated by undirected and directed mutagenesis in small laboratories as well as in high-throughput facilities have provided proof-of-principle that systematic, genome wide mutagenesis is feasible in the mouse. The principle difference between mouse mutagenesis screens and the previous screens in non-mammalian organisms is the significantly larger space and the more cost-intensive facilities required for such an approach in a mammal.

The success of a genome-scale mutagenesis project strongly depends on the quality and extend of the screening and phenotyping procedures. An efficient genome wide mutagenesis screen requires that the expertise of specialists from a variety of biological and medical fields is implemented in the screen of every potential mutant mouse generated by ENU treatment. In this regard, the systematic, phenotypic analysis of mouse mutants poses a challenge for both small laboratories and large facilities.

In order to reach the goal to systematically and comprehensively annotate every gene of the mouse genome with at least one described function ("*one mutation in every gene*"), the International Mouse Mutagenesis Consortium (IMMC) has called for a coordinated world-wide endeavour to integrate the research efforts of academic and economical laboratories [**41]. Specifically, it has been recognised that the establishment of networks for phenotyping centres will be one of the major milestones of this enterprise. It will be necessary to establish standard operating procedures and detailed phenotyping protocols that are made public so that phenotyping conditions can be reproduced and are directly comparable in laboratories world-wide. Recently, a mouse phenome database has been opened at the Jackson Laboratory, USA, that may serve as a platform for the research community to collect phenotypic data of commonly used and genetically modified inbred strain (www.jax.org/phenome).

We argue here that one important tool for an unbiased and systematic phenotypic analysis of mouse mutant resources at the molecular level is expression profiling (see below). In addition, microarray based techniques for transcriptome analysis have the potential for extensive automation. This facilitates standardisation and high-throughput analyses.

PHENOTYPING AT THE MOLECULAR LEVEL: THE POTENTIAL OF TRANSCRIPTOME ANALYSIS FOR FUNCTIONAL GENOMICS IN MAMMALS

The recent technologies to perform genome wide expression analyses have widely been recognized as a complementary approach to the 'classical' phenotyping strategies. The characterisation of changes in as many aspects as possible of an organism is the fundamental principal of the phenotypic analysis of mutants. The development of techniques to monitor gene expression at the genomic level, i.e., of thousands of genes in a single experiment, brings the possibilities of phenotyping to a new, molecular level [42]. The feasibility to monitor genome wide gene expression allows an unbiased way to access changes that are induced by a mutation. Such highly parallel expression studies will detect phenotypes – at the molecular

level – that may otherwise not be detected in standard phenotypic screens that specialise, e.g., on external appearances of mutants, blood parameters, immunological alterations and so on. In particular, in combination with high-throughput mutagenesis projects, expression profiling will certainly further improve the efficiency of the phenotypical characterisation of existing mutants and the isolation of new mutants (Fig. **1**).

To date, most publications that apply microarray technologies deal with the classification of tumours, such as, breast [43, 44] and prostate [45] cancers or leukemias, embryonal tumours [46], as well as inflammatory diseases such as rheumatoid arthritis [47]. The molecular phenotyping at a genomic scale of tumour tissues helps improve drug efficiencies in at least two ways: Transcriptome analysis allows an unbiased and systematic approach to tumour classification based on genome wide expression data [48]. It enables to differentiate between tumours that are morphologically, histopathologically, cytogenetically etc. indistinguishable but differ in their response to therapy [49]. Secondly, the molecular description may open the possibility to specifically design drugs that alter pathway activities to a (closer to) normal level [50]. Several target-designed drugs, aiming at specific molecular pathways, are already on the commercial market. Below, two examples related to the classification of leukemias are given, that show how molecular data in form of expression profiles can be used as diagnostic tool and to select therapy.

The differentiation between acute leukemias derived from lymphoid precursor (acute lymphoblastic leukemia) and from myeloid precursors (acute myeloid leukemia) is critical for the successful chemotherapeutical treatment [51]. In the first example, using RNA samples from 38 acute leukemias and chips containing 6817 genes it was demonstrated that the expression levels of approximately 1100 of these genes were more likely to correlate with class distinction between acute lymphoblastic and acute myeloid leukemia than being random [52]. New samples from heterogeneous sources of acute leukemias were assigned to one of the two classes based on the expression of 50 informative genes. Out of 34 samples, strong predictions were made for 29 samples with 100% accuracy. Interestingly, the arbitrarily chosen informative genes not only included markers of the haematopoietic lineage but also genes related to cancer pathogenesis, i.e. genes that code for proteins involved in S-phase cell cycle progression, chromatin remodelling, transcription, cell adhesion as well as known oncogenes. Thus, expression profiling may also provide insight into cancer pathogenesis and pharmacology [52].

The second example provided evidence that the classification of cancers allows the design of therapeutic treatment to individual cases. It was shown that the morphologically indistinct group of diffuse large B-cell lymphoma (DLBCL) based on their RNA expression profiles can be classified in at least two, molecularly more homogenous groups. Based on the expression of B-cell marker genes, one group was related to germinal centre (GC) B cells and the second subtype expressed genes that are indicative of in vitro activated peripheral blood B cells. Interestingly, in this study GC B-like DLBCL patients had a

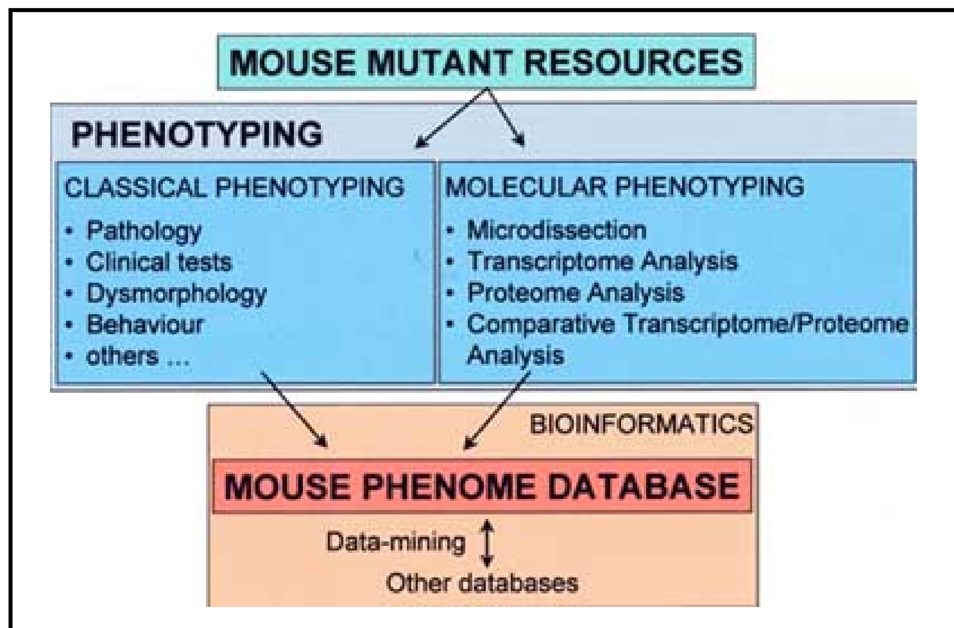

Fig. (1). For the analysis of mouse mutant resources phenotyping is key. We believe that besides the 'classical' phenotyping protocols molecular phenotyping, i.e., transcriptome and proteome analyses, will be an important tool for functional genomics and gene annotation. A comparative analysis of the transcriptome and proteome will allow distinguishing transcriptional and post-transcritpional regulation. The combination of innovative microdissection technologies with the molecular phenotyping techniques will pose an important challenge to improve sample homogeneity. The archiving of phenotypic data in mouse phenome databases serves as a virtual center for contributing laboratories. Phenome databases from different species are an important source for data-mining.

76% chance to survive the next five years following standard multi-agent chemotherapy, this chance was only 16% for patients with activated B-like DLBCL, whereas the average survival for all patients was 52% [53]. This study on the one hand demonstrates that applying more precise classifications of malignancies will allow a more selective use of therapy [54, 55]. On the other hand, this technique will also be helpful in designing new drugs. By precisely identifying the therapeutically induced molecular pathways in malignant cell lines upon chemotherapy, it may in the near future even be possible to design drugs – such as specific human antibodies – that target only those pathways that are clinically beneficial. Such an approach could help circumvent the strong side-effects of chemotherapy.

THE POTENTIAL OF LARGE DATA SETS FOR EXPRESSION ANALYSIS

The analysis of a 'compendium of expression profiles' from a large set of yeast mutants did establish the true potential of a comprehensive transcription analysis [56]. By comparing expression profiles of uncharacterised yeast mutants to a large and diverse set of reference profiles it was possible to match profiles of unknown mutants to profiles of mutants in known cellular pathways [*57]. As a first step towards standardizing profiling results in the study by Hughes et al., sets of genes with transcriptional fluctuations under apparently identical experimental conditions were identified. Based on these measurements an 'error model' was applied that valued the significance of a transcriptional change based also on the fluctuation in the control experiment. One very important finding in this study was that, based on their expression profiles mutants that are known to share a phenotype, generally clustered – or, in other words, mutants that affect the same cellular process, such as mitochondrial respiration, mating, or sterol pathway etc., often display related transcription profiles. In turn, it was possible to predict cellular functions of unknown genes based solely on their consistent affiliation to a group of co-regulated genes with known biological function. It was shown that the co-regulation of unknown genes with well characterized pathways can be indicative of a potential function of the unknown gene in a particular biological process. For example, a group of more than 100 genes was described that was co-regulated with components of the mitochondrial ribosome in 300 experiments. Approximately one third of them were unknown open reading frames; by mutagenising a selection of these genes and analysing respiratory deficiencies, it was demonstrated that clusters of co-regulated genes may be used to enrich for genes of a particular cellular function.

The analysis of transcript profiles during the metamorphosis of *Drosophila* has demonstrated that such correlations between characteristics of expression profiles and complex biological processes can also be established during developmental stages of multi-cellular organisms [58]. In this study co-regulated groups of genes that are involved, for example, in larval muscle breakdown, adult myogenesis, programmed cell death or cellular differentiation were identified which are known to be induced by ecdysone, an initiator signal of metamorphosis. These observations were made despite the restriction that entire organisms were used for this analysis of stage related expression profiles.

The next logical step in this type of analysis is, of course, the comparison of mutant versus wildtype transcript profiles. Whereas the current knowledge of gene function is usually limited to single pathways or a small set of target genes, transcription profiling of single gene mutants will allow the holistic analysis of regulatory interactions in global molecular networks. The large number of mouse mutants, that is currently generated, provides the resources needed to extend the analysis of transcription profiles in mammals to the same level as the 'compendium analysis' performed for yeast [57]. Similar to the study of yeast mutants and the recent progress in the classification of histologically and morphologically indistinguishable cancers, it will be possible to identify new categories within mutants of similar phenotypic traits. This will allow the determination of new molecular pathways that are associated with particular biological processes in mammals. In fact, the major difference between a differential transcriptome analysis of yeast and mice - besides the complexity of the genome (approximately 6.000 genes versus around 50.000 genes in mammals) – lies in the heterogeneity of tissue samples. Even if expression profiles are established for every discrete organ, it has to be considered that their cellular constitution is not homogenous, such that the transcript profile is the net result of different associated cell types. In addition, 'growth conditions' or environmental factors of higher animals are more complex than in vitro cultures of cells. They require stringent tests for the reproducibility of experiments. Important factors that ultimately affect gene expression and that have to be considered for the reproducibility of experiments are (or may be) parameters such as climate, nutrition, circadian rhythm, cage size, architecture of the cage, type of bedding, age of weaning and age at the time of the experiment, previous experimental procedures, number of siblings, the number of mice per cage and other parameters. However, our preliminary observations on expression profiles from organs of isogenic mouse strains that have been raised under specified pathogen free (spf) conditions and that have been treated using standardised operating protocols suggest, that highly reproducible expression profiles can be achieved under these conditions.

The storage and analysis of such collections of mouse mutant transcript profiles requires archiving in databases and software tools for the efficient analysis of expression patterns [59, 60]. One important aspect in this regard will be to link existing genetic and phenotypic information on mutants with comprehensive expression data. For the mouse, already one solution on how such data can be presented in a way that is useful for the scientific community has been put in place in the databases of The Jackson Laboratory (GXD) for 'classical' single gene expression data from in situ techniques, Northern analysis, or PCR based expression analysis (www.informatics.jax.org) [61, 62]. In particular, the integration in the GXD database of an 'Anatomical Dictionary Browser', which essentially represents an ontology of all structures, organs and tissues present in the mouse at each developmental stage, provides the basis for

efficient data-mining. Similarly, a common language for phenotypic descriptions must be developed [63, 64].

Several recent publications suggest that despite the complexity of mammalian organs, expression profiling is a useful tool to identify pathways associated with particular biological processes in mammalian organs. It was shown, for example, that ageing of the neocortex and cerebellum in mice is associated with changes in the expression of genes that are indicative, for example, of inflammatory response, oxidative stress and reduced neurotrophic support in these brain regions. This ageing related gene expression pattern was at least partially reversed by caloric restriction, suggesting that metabolic alterations have profound effects on brain ageing. Interestingly, the transcriptional response to ageing in the mouse brain has significant similarities to that in human neurodegenerative disorders, such as Alzheimer's disease [65, 66]. The differential gene expression in several brain regions and the response to seizure in two inbred mouse strains has also been analysed [67]. This study provided strong evidence for differential gene expression between isogenic mouse strains. This suggests that differences in gene expression may account for distinct phenotypes in inbred strains. Although these transcriptome studies in the mouse have clearly shown that important and interesting biological information can be obtained by analysing heterogeneous tissues, there is no doubt that the implication of innovative technologies to reduce tissue complexity, such as laser-microdissection, will further improve the interpretation of gene expression data [68, 69]. Whereas the microdissection technique is well established, the major challenge will be to improve the sensitivity of microarray hybridisations, for example, by establishing protocols for linear amplification of mRNA.

Another important by-product of a systematic analysis of expression profiles of uncharacterised mouse mutants is that it allows an efficient support of candidate gene approaches. By combining rough gene mapping data and the information on affected pathways in mutants, candidate genes can be selected from critical genomic intervals [70]. In addition, gene expression profiling of organs from uncharacterised mutagenised mice can be used as first line assay to identify new mutants. In this sense, expression profiling can be regarded as a complementary approach to phenotypic screens at a new molecular level ('molecular phenotyping'; Fig. **1**) [71]. This data will be particularly informative, once profiles of uncharacterised mutants can be correlated to large sets of expression data from well characterised mouse mutant lines. Such data mining approaches will shorten the route from highly-parallel approaches, genomic sequence and functional genomics to a new level of hypothesis driven experimental biology.

TO THE TRANSCRIPTOME AND BEYOND

Most biochemical processes within and between cells are put into effect by the interaction between proteins, or between proteins and their substrates. The proteome of a cell is the result of controlled biosynthesis, and hence largely (but not exclusively) regulated by gene expression. In turn, gene expression can be regarded as a sensitive read-out of the biochemical state of the cell, or in other words the proteome. Transcriptome and proteome feedback to each other in a highly complex and somehow controlled way. Thus, the regulatory context is a crucial part of gene function [72]. The understanding of this functional regulation is until today limited to isolated signalling or metabolic pathways. The regulatory interactions within the molecular network of the cell are far from being understood. However, evidence is accumulating that those single pathways that have been studied in diverse organisms are in fact mere components of complicated networks *"that integrate many inputs to generate the complex output that is cell behaviour"* [73]. When it will be achieved to integrate techniques to isolate homogenous populations of cells from complex tissues, such as microdissection, with techniques that allow comparative transcriptome and proteome analyses in mutant animals, then it will be possible - for the first time – to analyse gene function in the context of the molecular network of the cell (Fig. **1**). Such an holistic approach of molecular analysis would have important synergistic effects on the analysis of regulatory interdependencies that determine the molecular phenotype of the cell; it would also allow to distinguish between transcriptional and post-transcriptional regulation in a holistic approach.

What can we learn from comprehensive functional genomics and proteomics? At the end of this era we will probably have a good understanding on how components of the transcriptome or the proteome relate to each other on the cellular level, how the function of single gene products affects the molecular network of the cell, and how transcriptome and proteome regulate each other. We may also know for most genes what the consequences of changes in the biochemical status of cells are on the physiological, morphological and anatomical level – with some inherent restrictions due to the experimental system [74]. But what about epigenetic factors or environmental determination? Should it not be expect that if we largely understand the genetic determination of organisms (at some time in the future) – from embryonic development, over disease, to social and psychological traits – we will also improve our knowledge on those traits that are not genetically determined? Although there are several instances where specific alleles have been implicated in psychological traits or social behaviour in man, such as schizophrenia or male homosexuality [75, 76], it is probably obvious that a human being is not only the result of its genetic constitution. However, also these environmental or epigenetic factors will find their manifestation in the physiological state of the cell and thus are detectable in changes of the proteome or the transcriptome.

ACKNOWLEDGEMENT

We would like to thank Veronique Blanquet and colleagues from the lab for helpful comments on the manuscript.

REFERENCES

[1] Hattori, M.; Fujiyama, A.; Taylor, T.D.; Watanabe, H.; Yada, T.; Park, H.S.; Toyoda, A.; Ishii, K.; Totoki, Y.; Choi, D.K.; Soeda, E.; Ohki, M.; Takagi, T.; Sakaki, Y.; Taudien, S.; Blechschmidt, K.; Polley, A.; Menzel, U.; Delabar, J.; Kumpf, K.; Lehmann, R.; Patterson, D.; Reichwald, K.; Rump, A.; Schillhabel, M.; Schudy, A. The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium [see comments]. (2000) *Nature* **405**, 311-319.

[**2] IHGSC: Initial sequencing and analysis of the human genome. (2001) *Nature* **409**, 860-921.

See annotation to **3

[**3] Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor, Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di, Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X.The sequence of the human genome. (2001) *Science*, **291**, 1304-1351.

The articles by the International Human Genome Consortium and Venter et al. as well as other publications in these issues of Nature and Science present an initial analysis of the draft sequence of the human genome.

[**4] Denny, P.; Justice, M.J. Mouse as the measure of man? (2000) *Trends Genet.* **16**, 283-287.

See annotation to **6

[5] Rossant, J.; McKerlie, C. Mouse-based phenogenomics for modelling human disease. *Trends* (2001) *Mol. Med.* **7**, 502-507.

[**6] Malakoff, D. The rise of the mouse, biomedicine's model mammal [news]. (2000) *Science* **288**, 248-253.

The articles by Denny and Justice **4 and by Malakoff give brief overviews of the history of mouse genetics and the resources that have been generated in the last decades. The resources of information, clone banks, and inbred and mutant strains are the basis for functional genomics in a mammal.

[7] McCarthy, L.C.; Terrett, J.; Davis, M.E.; Knights, C.J.; Smith, A.L.; Critcher, R.; Schmitt, K.; Hudson, J.; Spurr, N.K.; Goodfellow, P.N. A first-generation whole genome-radiation hybrid map spanning the mouse genome. (1997) *Genome Res.* **7**, 1153-1161.

[8] Van Etten, W.J.; Steen, R.G.; Nguyen, H.; Castle, A.B.; Slonim, D.K.; Ge, B.; Nusbaum, C.; Schuler, G.D.; Lander, E.S.; Hudson, T.J. Radiation hybrid map of the mouse genome. (1999) *Nat. Genet.* **22**, 384-387.

[9] Lindblad-Toh, K.; Winchester, E.; Daly, M.J.; Wang, D.G.; Hirschhorn, J.N.; Laviolette, J.P.; Ardlie, K.; Reich, D.E.; Robinson, E.; Sklar, P.; Shah, N.; Thomas, D.; Fan, J.B.; Gingeras, T.; Warrington, J.; Patil, N.; Hudson, T.J.; Lander, E.S. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. (2000) *Nat. Genet.* **24**, 381-386.

[10] Jackson, I.J. Mouse mutagenesis on target. (2001) *Nat. Genet.* **28**, 198-200.

[11] Yu, Y.; Bradley, A. Engineering chromosomal rearrangements in mice. (2001) *Nat. Rev. Genet.* **2**, 780-790.

[12] Stanford, W.L.; Cohn, J.B.; Cordes, S.P. Gene-trap mutagenesis: past, present and beyond. (2001) *Nat. Rev. Genet.* **2**, 756-768.

[13] Doetschman, T.; Gossler, A.; Serfling, E.; Schaffner, W.; Marcu, K.; Stanton, L.; Kemler, R. Introduction of genes into mouse embryonic stem cells. (1986) *Prog. Clin. Biol. Res.* **??**, 47-50.

[14] Thomas, K.R.; Folger, K.R.; Capecchi, M.R. High frequency targeting of genes to specific sites in the mammalian genome. (1986) *Cell* **44**, 419-428.

[15] Hogan, B.; Costantini, F.; Lacy, E. Manipulating the mouse embryo, a laboratory manual. (1994) *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.*

[16] Dymecki, S.M. Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6191-6196.

[17] Gu, H.; Zou, Y.R.; Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. (1993) *Cell* **73**, 1155-1164.

[18] Lakso, M.; Sauer, B.; Mosinger, B.; Jr.; Lee, E.J.; Manning, R.W.; Yu, S.H.; Mulder, K.L.; Westphal, H. Targeted oncogene activation by site-specific recombination in transgenic mice. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6232-6236.

[19] Orban, P.C.; Chui, D.; Marth, J.D. Tissue- and site-specific DNA recombination in transgenic mice. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6861-6865.

[20] Herault, Y.; Rassoulzadegan, M.; Cuzin, F.; Duboule, D. Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). (1998) *Nat. Genet.* **20**, 381-384.

[21] Hrabe de Angelis, M.; Balling, R. Large scale ENU screens in the mouse: genetics meets genomics. (1998) *Mutat. Res.* **400**, 25-32.

[22] Rinchik, E.M.; Carpenter, D.A. N-ethyl-N-nitrosourea mutagenesis of a 6- to 11-cM subregion of the Fah- Hbb interval of mouse chromosome 7: Completed testing of 4557 gametes and deletion mapping and complementation analysis of 31 mutations. (1999) *Genetics* **152**, 373-383.

[23] Justice, M.J.; Zheng, B.; Woychik, R.P.; Bradley, A. Using targeted large deletions and high-efficiency N-ethyl-N-nitrosourea mutagenesis for functional analyses of the mammalian genome. (1997) *Methods* **13**, 423-436.

[24] Zheng, B.; Mills, A.A.; Bradley, A. A system for rapid generation of coat color-tagged knockouts and defined chromosomal rearrangements in mice. (1999) *Nucleic Acids Res.* **27**, 2354-2360.

[25] Brown, S.D.; Peters, J. Combining mutagenesis and genomics in the mouse--closing the phenotype gap. (1996) *Trends Genet.* **12**, 433-435.

[*26] Hrabe de Angelis, M.H.; Flaswinkel, H.; Fuchs, H.; Rathkolb, B.; Soewarto, D.; Marschall, S.; Heffner, S.; Pargent, W.; Wuensch, K.; Jung, M.; Reis, A.; Richter,T.; Alessandrini, F.; Jakob, T.; Fuchs, E.; Kolb, H.; Kremmer, E.; Schaeble, K.; Rollinski, B.; Roscher, A.; Peters, C.; Meitinger, T.; Strom, T.; Steckler, T.; Holsboer, F.; Klopstock, T.; Gekeler, F.; Schindewolf, C.; Jung, T.; Avraham, K.; Behrendt, H.; Ring, J.; Zimmer, A.; Schughart, K.; Pfeffer, K.; Wolf, E.; Balling, R. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. (2000) *Nat. Genet.* **25**, 444-447.

See annotation to [*27]

[*27] Nolan, P.M.; Peters, J.; Strivens, M.; Rogers, D.; Hagan, J.; Spurr, N.; Gray, I.C.; Vizor, L.; Brooker, D.; Whitehill, E.; Washbourne, R.; Hough, T.; Greenaway, S.; Hewitt, M.; Liu, X.; McCormack, S.; Pickford, K.; Selley, R.; Wells, C.; Tymowska-Lalanne, Z.; Roby, P.; Glenister, P.; Thornton, C.; Thaung, C.; Stevenson, J.A.; Arkell, R. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. (2000) *Nat. Genet.* **25**, 440-443.

The publications by Hrabé de Angelis *et al.* *26 and Nolan *et al.* demonstrated that systematic genome-wide mutagenesis screens could efficiently be performed in a mammal closely related to man. Similar mouse mutagenesis screens were subsequently launched world-wide.

[28] Wiles, M.V.; Vauti, F.; Otte, J.; Fuchtbauer, E.M.; Ruiz, P.; Fuchtbauer, A.; Arnold, H.H.; Lehrach, H.; Metz, T.; von Melchner, H.; Wurst, W. Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. (2000) *Nat. Genet.* **24**, 13-14.

[29] Gans, M.; Audit, C.; Masson, M. Isolation and characterization of sex-linked female-sterile mutants in Drosophila melanogaster. (1975) *Genetics* **81**, 683-704.

[30] Nusslein-Volhard, C.; Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. (1980) *Nature* **287**, 795-801.

[31] Brenner, S. The genetics of *Caenorhabditis elegans*. (1974) *Genetics* **77**, 71-94.

[32] Kemphues, K.J.; Priess, J.R.; Morton, D.G.; Cheng, N.S. Identification of genes required for cytoplasmic localization in early C. elegans embryos. (1988) *Cell* **52**, 311-320.

[33] Mayer, M. Molecular neurobiology. NMDA receptors cloned at last [news; comment]. (1991) *Nature* **354**, 16-17.

[34] Haffter, P.; Granato, M.; Brand, M.; Mullins, M.C.; Hammerschmidt, M.; Kane, D.A.; Odenthal, J.; van Eeden, F.J.; Jiang, Y.J.; Heisenberg, C.P.; Kelsh, R.N.; Furutani-Seiki, M.; Vogelsang, E.; Beuchle, D.; Schach, U.; Fabian, C.; Nusslein-Volhard, C. The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. (1996) *Development* **123**, 1-36.

[35] Driever, W.; Solnica-Krezel, L.; Schier, A.F.; Neuhauss, S.C.; Malicki, J.; Stemple, D.L.; Stainier, D.Y.; Zwartkruis, F.; Abdelilah, S.; Rangini, Z.; Belak, J.; Boggs, C. A genetic screen for mutations affecting embryogenesis in zebrafish. (1996) *Development* **123**, 37-46.

[36] Duboule, D.; Wilkins, A.S. The evolution of 'bricolage'. (1998) *Trends Genet.* **14**, 54-59.

[37] Jacob, F. Evolution and tinkering. (1977) *Science* **196**, 1161-1166.

[38] Raff, R.A. The shape of life. (1996) *The University of Chicago Press: genes, development, and the evolution of animal form*

[39] Matin, A.; Nadeau, J.H. Sensitized polygenic trait analysis. (2001) *Trends Genet.* **17**, 727-731.

[40] Russell, W.L.; Kelly, E.M.; Hunsicker, P.R.; Bangham, J.W.; Maddux, S.C.; Phipps, E.L. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5818-5819.

[**41] Nadeau, J.H.; Balling, R.; Barsh, G.; Beier, D.; Brown, S.D.; Bucan, M.; Camper, S.; Carlson, G.; Copeland, N.; Eppig, J.; Fletcher, C.; Frankel, W.N.; Ganten, D.; Goldowitz, D.; Goodnow, C.; Guenet, J.L.; Hicks, G.; de Angelis, M.H.; Jackson, I.; Jacob, H.J.; Jenkins, N.; Johnson, D.; Justice, M.; Kay, S.; Kingsley, D.; Lehrach,

H.; Magnuson, T.; Meisler, M.; Poustka, A.; Rinchik, E.M.; Rossant, J.; Russell, L.B.; Schimenti, J.; Shiroishi, T.; Skarnes, W.C.; Soriano, P.; Stanford, W.; Takahashi, J.S.; Wurst, W.; Zimmer, A. Sequence interpretation. Functional annotation of mouse genome sequences. (2001) *Science* **291**, 1251-1255.

This article is a call for a world-wide effort to functionally and comprehensively annotate the mammalian genome.

[42] Schena, M.; Shalon, D.; Davis, R.W.; Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [see comments]. (1995) *Science* **270**, 467-470.

[43] van 't Veer, L.J.; Dai, H.; van de Vijver, M.J.; He, Y.D.; Hart, A.A.; Mao, M.; Peterse, H.L.; van der Kooy, K.; Marton, M.J, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002, **415**, 530-536.

[44] Nacht, M.; Ferguson, A.T.; Zhang, W.; Petroziello, J.M.; Cook, B.P.; Gao, Y.H.; Maguire, S.; Riley, D.; Coppola, G.; Landes, G.M.; Madden, S.L.; Sukumar, S. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. (1999) *Cancer Res.* **59**, 5464-5470.

[45] Elek, J.; Park, K.H.; Narayanan, R. Microarray-based expression profiling in prostate tumors. *In Vivo* 2000, **14**, 173-182.

[46] Pomeroy, S.L.; Tamayo, P.; Gaasenbeek, M.; Sturla, L.M.; Angelo, M.; McLaughlin, M.E.; Kim, J.Y.; Goumnerova, L.C.; Black, P.M.; Lau, C.; Allen, J.C.; Zagzag, D.; Olson, J.M.; Curran, T.; Wetmore, C.; Biegel, J.A.; Poggio, T.; Mukherjee, S.; Rifkin, R.; Califano, A.; Stolovitzky, G.; Louis, D.N.; Mesirov, J.P.; Lander, E.S.; Golub, T.R. Prediction of central nervous system embryonal tumour outcome based on gene expression. (2002) *Nature* **415**, 436-442.

[47] Heller, R.A.; Schena, M.; Chai, A.; Shalon, D.; Bedilion, T.; Gilmore, J.; Woolley, D.E.; Davis, R.W. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 2150-2155.

[48] Wooster, R. Cancer classification with DNA microarraysis less more? (2000) *Trends Genet.* **16**, 327-329.

[49] Bittner, M.; Meltzer, P.; Chen, Y.; Jiang, Y.; Seftor, E.; Hendrix, M.; Radmacher, M.; Simon, R.; Yakhini, Z.; Ben-Dor, A.; Sampas, N.; Dougherty, E.; Wang, E.; Marincola, F.; Gooden, C.; Lueders, J.; Glatfelter, A.; Pollock, P.; Carpten, J.; Gillanders, E.; Leja, D.; Dietrich, K.; Beaudry, C.; Berens, M.; Alberts, D.; Sondak, V. Molecular classification of cutaneous malignant melanoma by gene expression profiling. (2000) *Nature* **406**, 536-540.

[50] Debouck, C.; Goodfellow, P.N. DNA microarrays in drug discovery and development. (1999) *Nat. Genet.* **21**, 48-50.

[51] Bishop, J.F. Adult acute myeloid leukaemia: update on treatment. (1999) *Med. J. Aust.* **170**, 39-43.

[52] Golub, T.R.; Slonim, D.K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J.P.; Coller, H.; Loh, M.L.; Downing, J.R.; Caligiuri, M.A.; Bloomfield, C.D.; Lander, E.S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. (1999) *Science* **286**, 531-537.

[53] Alizadeh, A.A.; Eisen, M.B.; Davis, R.E.; Ma, C.; Lossos, I.S.; Rosenwald, A.; Boldrick, J.C.; Sabet, H.; Tran, T.; Yu, X.; Powell, J.I.; Yang, L.; Marti, G.E.; Moore, T.; Hudson, J.; Jr.; Lu, L.; Lewis, D.B.; Tibshirani, R.; Sherlock, G.; Chan, W.C.; Greiner, T.C.; Weisenburger, D.D.; Armitage, J.O.; Warnke, R.; Staudt, L.M.; *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling [see comments]. (2000) *Nature* **403**, 503-511.

[54] Roden, D.M.; George, A.L. The genetic basis of variability in drug responses. (2002) *Nature Reviews* **1**, 37-44.

[55] Pinkel, D. Cancer cells, chemotherapy and gene clusters [news; comment]. (2000) *Nat. Genet.* **24**, 208-209.

[56] Gullans, S.R. Of microarrays and meandering data points. (2000) *Nat. Genet.* **26**, 4-5.

[*57] Hughes, T.R.; Marton, M.J.; Jones, A.R.; Roberts, C.J.; Stoughton, R.; Armour, C.D.; Bennett, H.A.; Coffey, E.; Dai, H.; He, Y.D.; Kidd, M.J.; King, A.M.; Meyer, M.R.; Slade, D.; Lum, P.Y.; Stepaniants, S.B.; Shoemaker, D.D.; Gachotte, D.; Chakraburtty, K.; Simon, J.; Bard, M.; Friend, S.H. Functional discovery via a compendium of expression profiles. (2000) *Cell* **102**, 109-126.

This paper discovers the true potential of large-scale transcriptome analysis in identifying regulatory units and understanding gene function by evaluating expression profiles of a substantial group of yeast mutants.

[58] White, K.P.; Rifkin, S.A.; Hurban, P.; Hogness, D.S. Microarray analysis of Drosophila development during metamorphosis. (1999) *Science* **286**, 2179-2184.

[59] Beissbarth, T.; Fellenberg, K.; Brors, B.; Arribas-Prat, R.; Boer, J.; Hauser, N.C.; Scheideler, M.; Hoheisel, J.D.; Schutz, G.; Poustka, A.; Vingron, M. Processing and quality control of DNA array hybridization data. (2000) *Bioinformatics* **16**, 1014-1022.

[60] Frishman, D.; Albermann, K.; Hani, J.; Heumann, K.; Metanomski, A.; Zollner, A.; Mewes, H.W. Functional and structural genomics using PEDANT. (2001) *Bioinformatics* **17**, 44-57.

[61] Ringwald, M.; Eppig, J.T.; Begley, D.A.; Corradi, J.P.; McCright, I.J.; Hayamizu, T.F.; Hill, D.P.; Kadin, J.A.; Richardson, J.E. The Mouse Gene Expression Database (GXD). (2001) *Nucleic Acids Res.* **29**, 98-101.

[62] Begley, D.A.; Ringwald, M. Electronic tools to manage gene expression data. (2002) *Trends Genet.* **18**, 108-110.

[63] Eppig, J.T. Algorithms for mutant sorting: the need for phenotype vocabularies. (2000) *Mamm. Genome* **11**, 584-589.

[64] Paigen, K.; Eppig, J.T. A mouse phenome project. (2000) *Mamm. Genome* **11**, 715-717.

[65] Lee, C.K.; Klopp, R.G.; Weindruch, R.; Prolla, T.A. Gene expression profile of aging and its retardation by caloric restriction [see comments]. (1999) *Science* **285**, 1390-1393.

[66] Lee, C.K.; Weindruch, R.; Prolla, T.A. Gene-expression profile of the ageing brain in mice. (2000) *Nat. Genet.* **25**, 294-297.

[67] Sandberg, R.; Yasuda, R.; Pankratz, D.G.; Carter, T.A.; Del, Rio, J.A.; Wodicka, L.; Mayford, M.; Lockhart, D.J.; Barlow, C. From the cover: regional and strain-specific gene expression mapping in the adult mouse brain. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 11038-11043.

[68] Schutze, K.; Clement-Sengewald, A. Catch and move--cut or fuse. (1994) *Nature* **368**, 667-669.

[69] Schutze, K.; Lahr, G. Identification of expressed genes by laser-mediated manipulation of single cells [see comments]. (1998) *Nat. Biotechnol.* **16**, 737-742.

[70] Wells, C.; Brown, S.D. Genomics meets genetics: towards a mutant map of the mouse. (2000) *Mamm. Genome* **11**, 472-477.

[71] Beckers, J.; Hrabe, de Angelis, M. Large-scale mutational analysis for the annotation of the mouse genome. (2001) *Current Opinion in Chemical Biology* **6**, 17-23.

[72] Fessele, S.; Maier, H.; Zischek, C.; Nelson, P.J.; Werner, T. Regulatory context is a crucial part of gene function. (2002) *Trends Genet.* **18**, 60-63.

[73] Peifer, M. Signal transduction. Neither straight nor narrow [news; comment]. (1999) *Nature* **400**, 213-215.

[74] Tautz, D. A genetic uncertainty problem. (2000) *Trends Genet.* **16**, 475-477.

[75] Brzustowicz, L.M.; Hodgkinson, K.A.; Chow, E.W.; Honer, W.G.; Bassett, A.S. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. (2000) *Science* **288**, 678-682.

[76] Hamer. D.H.; Hu, S.; Magnuson, V.L.; Hu, N.; Pattatucci, A.M. A linkage between DNA markers on the X chromosome and male sexual orientation [see comments]. (1993) *Science* **261**, 321-327.