© 1999 Nature America Inc. • http://genetics.nature.com

The genetics of cancer—a 3D model

Kristina A. Cole[1,2], David B. Krizman[3] & Michael R. Emmert-Buck[1,3]

[1]Pathogenetics Unit, Laboratory of Pathology, National Cancer Institute, Bethesda, Maryland 20892, USA. [2]Department of Pathology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [3]Cancer Genome Anatomy Project, Office of the Director, National Cancer Institute, Bethesda, Maryland 20892, USA. e-mail: mbuck@helix.nih.gov

Gene expression microarrays hold great promise for studies of human disease states. There are significant technical issues specific to utilizing clinical tissue samples which have yet to be rigorously addressed and completely overcome. Precise, quantitative measurement of gene expression profiles from specific cell populations is at hand, offering the scientific community the first comprehensive view of the *in vivo* molecular anatomy of normal cells and their diseased counterparts. Here, we propose a model for integrating—in three dimensions—expression data obtained using the microarray.

© 1999 Nature America Inc. • http://genetics.nature.com

New, high-throughput mRNA analysis platforms such as microarrays are capable of producing large gene expression data sets with the potential to provide novel insights into fundamental cancer biology at the molecular level[1–7] (Fig. 1). While initially exciting, data sets obtained from microarrays comparing normal and tumour samples become difficult to manage and immediately raise a number of questions: what is the validity and quantitative accuracy of the observed changes? Which genes should be prioritized for further study? How does one determine whether a given gene is a cause rather than a consequence of tumourigenesis? How can the histopathology and molecular information from multiple samples and many people be organized into a queryable database that allows biological questions to be asked and answered? In addition to specific considerations related to the acquisition and characterization of clinical samples (see page 26 (ref. 8)), there are numerous questions about how to best organize, interpret and gain insights from large amounts of gene expression data generated by complex biological systems.

Molecular fingerprinting and the CGAP

To promote the understanding of human cancer at a molecular level, the National Cancer Institute recently initiated the Cancer Genome Anatomy Project (CGAP, ref. 9; http://www.ncbi.nlm.nih.gov/ncicgap/). Two of its immediate goals are the development of a complete publicly available human UniGene set (see page 26 of this issue (ref. 8)) and the production and sequencing of cDNA libraries from five major human cancers (prostate, ovary, breast, lung and gastrointestinal tract). Future goals include a 'Molecular Fingerprinting' website, which should become available in the early months of 1999 as a component of the CGAP website (http://www.ncbi.nlm.nih.gov/ncicgap/). Its initial goal is to determine the key issues which need to be considered such that molecular analyses performed on specimens from human patients take into account the unique issues related to clinical studies. It will also focus on practical issues, such as the review of pertinent literature and the development of protocols, methods and a model approach for studying gene expression profiles in human neoplasia—using prostate cancer as a prototype.

Limitations of current surgical pathology methods

The standard surgical excision, processing and histopathological examination of tissue specimens is geared towards patient care. For example, accurate diagnosis and staging are first on the list when it comes to surgical resections for cancer[10–14]. Methods for fixing and embedding tissue samples were not developed with microarray-based studies in mind. Typically, the majority of tissue from surgical resections is discarded and the portions important for diagnosis and staging are processed through aldehyde-based fixatives (for example, formalin) which damage mRNA integrity[15]. If frozen tissue is collected at the time of surgery, it is possible to recover and study mRNA from dissected cell populations. However, frozen tissue sections are technically difficult to prepare, the histology is often severely compromised, and the tissue available to investigators typically contains only a limited portion of the tumour with few premalignant lesions or associated normal glands (see page 25 of this issue (ref. 8)) for further discussion of sample preperation). In short, current surgical pathology practice is ill-suited to high-throughput genetic technologies; new methods and strategies are required.

The third dimension

To study gene expression profiles in prostate cancer, our laboratory is taking a three-dimensional (3D) analytical approach to characterizing the entire prostate gland (Fig. 2). We think this is the best method for producing and querying multiplex data from human neoplasias and offers several advantages over the standard 'normal versus tumour' type of comparison. Consider the selection of normal epithelium as a baseline control against which to compare and contrast tumour gene expression profiles. Normal epithelium in prostatic ducts ranges from atrophic to resting to hyperplastic and each has a unique pattern of gene expression. Additionally, epithelium adjacent to tumour may not be 'normal'—although phenotypically 'normal' at the light microscopic level, it may be genotypically abnormal and/or exhibit an altered gene expression profile due to its proximity to the invading tumour[16–19]. Factors such as the degree of associated inflammation or proximity to the urethra may also have a significant impact. It is therefore necessary to profile a spectrum of normal and tumour cell populations from a series of patients to distinguish between alterations that are relevant to tumourigenesis and those reflecting the biological spectrum of the 'normal' prostate, or have occurred for reasons unrelated to transformation. The need to identify changes specific to the disease process is evident when one contemplates the size of microarray data sets, the substantial amount of time required to follow up a candidate gene, and potential for con-

© 1999 Nature America Inc. • http://genetics.nature.com

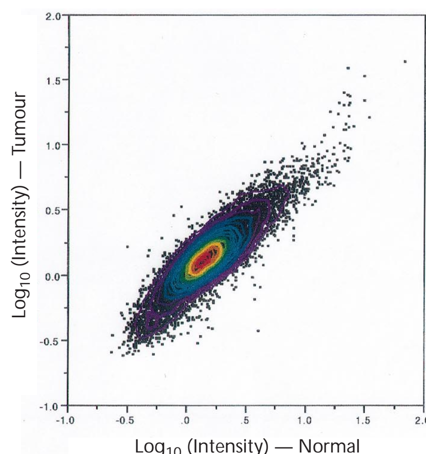

Fig. 1 Microarray experiment: normal versus tumour. Comparison of microdissected prostate epithelial cells and tumour cells from an individual patient. Approximately 2000 cells were dissected from a frozen tissue section, total RNA was recovered, and a target prepared by an RT-PCR-based method with incorporation of ^{33}P (refs 41–43). Normal and tumour targets were separately hybridized to nylon array filters (Genome Systems), scanned on a phosphorimager and analysed by P-SCAN (http://absalpha.dcrt.nih.gov:8008). Densitometry analysis revealed significant dysregulation of approximately 40 genes (0.2% of those represented on the microarray), including several known genes previously implicated in carcinogenesis[44–47].

Candidate Genes

upregulated in tumour
- serine/threonine kinase (*STK2*)
- β subunit of proteasome (*PSMB4*)
- CD36
- ribosomal protein L17 (*RPL17*)
- phospholipase A2 (*PLA2G1B*)
- ESTs

down-regulated in tumour
- desmoglein 2 (*DSG2*)
- tyro3 protein tyrosine kinase (*TYRO3*)
- type IV collagen (*COL4A4*)
- activating transcription factor 3 (*ATF3*)
- annexin 1 (*ANX1*)
- ESTs

founding interpretation by inclusion of changes unrelated to the pathology under study.

A second advantage of the 3D analytical approach is the opportunity to study the complete spectrum of tumour progression in an appropriate context. Resected prostatectomy specimens display a fascinating spectrum of phenotypic change, including hyperplasias, dysplasias and various grades of invasive tumour[20,21]. In comparison with other major human cancers, little is known of the genes or pathways that mediate the formation or progression of prostate tumours, and virtually nothing is known about how stromal cells, inflammatory cells, or the host–environment interaction affect this process[22–28]. A 3D reconstruction approach allows their effects to be gauged by permitting one to determine which dysplastic lesions develop into tumour and concurrent changes in gene expression profiles. Early dysplastic lesions may be especially useful in determining the fundamental molecular events that initiate prostate tumours[29,30]. Similarly, the 'normal' epithelium with hyperplastic change may serve as a particularly useful cell population for comparison, as its cells actively proliferate but do not progress to cancer, facilitating the identification of expression profiles unique to malignancy as opposed to those associated with benign cell growth.

Practical considerations

One of the important priorities for the 'Molecular Fingerprinting' effort is to field-test new strategies and methodologies for processing tissue samples. Ultimately, it is essential that a universal protocol is developed which is robust, inexpensive, easy to use, protects macromolecules (specifically, DNA, RNA and protein), and produces tissue sections with optimum histological detail for diagnosis. So far, we have processed 31 whole-mount prostatectomy specimens, using a variety of standard surgical pathology tissue methods as well as newly developed fixation and embedding strategies. This has allowed a rigorous comparison of methods and has produced a protocol which is sufficient for both patient diagnosis and high-throughput molecular analyses. Briefly, prostatectomy specimens are taken directly from the operating room and

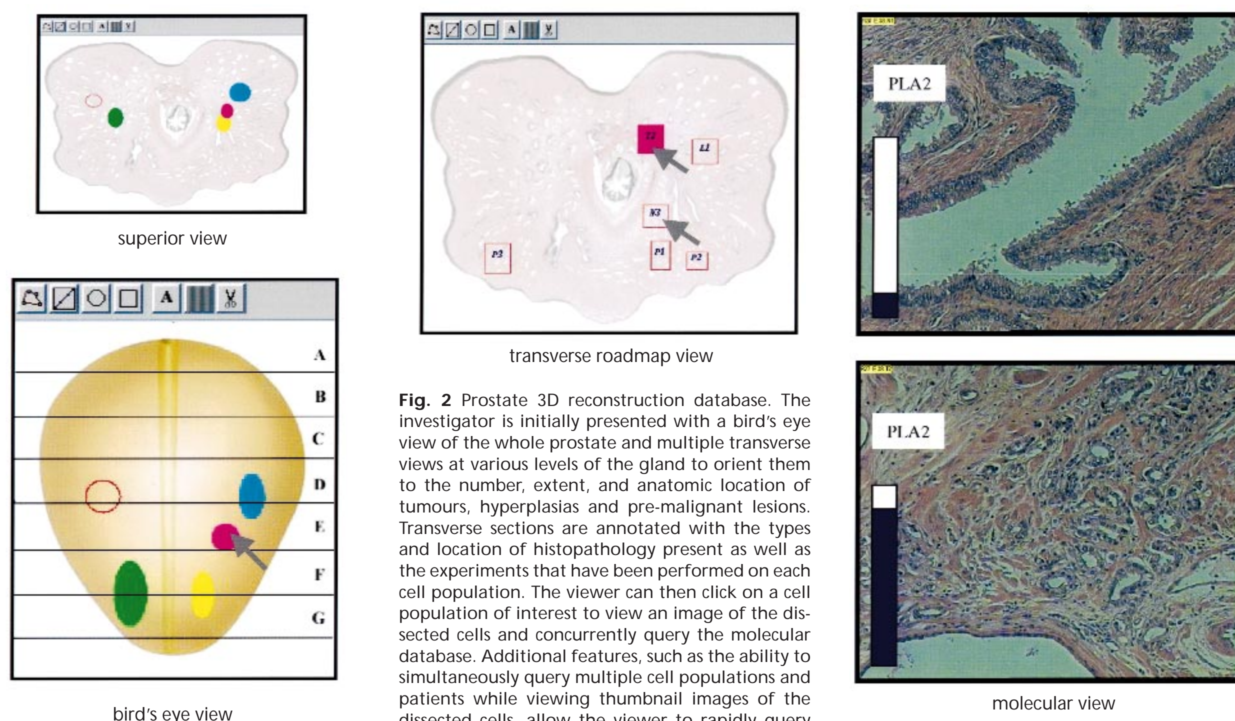

Fig. 2 Prostate 3D reconstruction database. The investigator is initially presented with a bird's eye view of the whole prostate and multiple transverse views at various levels of the gland to orient them to the number, extent, and anatomic location of tumours, hyperplasias and pre-malignant lesions. Transverse sections are annotated with the types and location of histopathology present as well as the experiments that have been performed on each cell population. The viewer can then click on a cell population of interest to view an image of the dissected cells and concurrently query the molecular database. Additional features, such as the ability to simultaneously query multiple cell populations and patients while viewing thumbnail images of the dissected cells, allow the viewer to rapidly query gene expression profiles across a spectrum of samples and/or patients.

© 1999 Nature America Inc. • http://genetics.nature.com

© 1999 Nature America Inc. • http://genetics.nature.com

processed with a non-aldehyde fixative (70% ethanol) and a low-temperature embedding compound (polyethylene glycol distearate) which preserves histological detail and protects mRNA and proteins for subsequent analysis. We cut whole-mount transverse cross-sections (representing the X and Y dimensions), such that the entire prostate gland, including the complete spectrum of normal epithelium and tumour progression, is available for viewing and microdissection. Several hundred adjacent 8-μm serial recut slides are prepared from the tissue blocks (8 μm is about half the diameter of an epithelial cell), revealing all of the normal histology and pathology in the Z-dimension. One may thereby determine the exact physical relationship of the normal ducts, premalignant lesions and tumour(s), and obtain an anatomical framework on which to overlay gene expression data.

Target preparation

We are currently comparing and contrasting both amplification-based and non-amplification-based methods to prepare microarray target from dissected cells. Each approach has unique strengths and weaknesses. For example, if 50,000–100,000 cells of a given type are present in a tissue section, then array experiments can be performed after dissection and mRNA recovery using direct labelling of first-strand cDNA. There are, however, several drawbacks to this approach—the fact that cells of interest are often limited in number; a substantial amount of time and effort is required to dissect over 50,000 cells; and genes of moderate or low abundance may not be detected on arrays without an amplification step.

Amplification-based methods offer the advantage of requiring substantially less than 50,000 cells. This reduces the time and effort involved in microdissection, particularly when an investigator wishes to study several different cell types. More importantly, this approach allows analysis of cell populations (hyperplasias and dysplasias, for example) which consist of only a few hundred to a few thousand cells. We typically use approximately 1000–2000 dissected cells to prepare microarray targets, obtaining reliable and reproducible results. We have not yet determined the lower limit of dissected cells needed for an array experiment although other groups have reported successful linear mRNA amplification from small numbers of cells[31]. The drawback, of course, with amplification-based schemes is the as-yet unknown transcript bias which is introduced. Ultimately, investigators will need to select their approach according to experimental parameters such as the number of available normal or lesional cells, the abundance of transcripts of interest, and the effect of amplification on mRNA representation.

As target from dissected samples is derived from a small amount of template, target bias, and experimental reproducibility are major concerns. All target preparation techniques induce some amount of bias, at the level of incorporation of label, reverse transcription, linear amplification, or PCR amplification. Key questions for experiments comparing two (or more) microdissected cell populations are as follows:

- What is the fold increase in transcripts produced by amplification-based methods?

- What is the overall reproducibility of array experiments with each target preparation method?

The answers to these questions are clearly central to the interpretation of microarray experiments.

- Is the induced bias reproducible?

As we are interested in the relative ratio of expression between comparison groups and are not attempting to determine the precise number of mRNA molecules in a cell population, an amplification-induced bias is acceptable if it occurs identically from sample to sample. This question is being addressed experimentally by comparing the reproducibility of array results from replicate experiments from a single mRNA source.

- Does the bias alter transcript levels to the extent that important differences between populations are lost?

This question is being addressed experimentally using a single mRNA source and separate targets prepared by amplification-based methods and direct labelling. Comparison of array results with northern blot analysis of the same mRNA source shows the effects of each approach on transcript representation.

The fourth dimension—integration

Integration of histopathology and high-throughput molecular analyses into a queryable database is not a trivial task. Investigators must be able to efficiently compare and contrast large datasets across multiple patients and samples, and make sense of the information in a biological context (see page 51 of this issue (ref. 32)). Our approach has been to create a web-based, visually-oriented system that allows efficient querying of gene expression profiles while viewing associated anatomy and histopathology. The viewer is intially presented with an overview of the prostate gland and is subsequently able to view multiple regions of the specimen. The database is queryable such that gene expression data can be overlayed with histopathology images. An important goal of the database is to add protein data to gene expression datasets such that a comprehensive analysis of the status of genes and gene products will be possible[33–36]. This should permit one to chart—in three dimensions—the sequence of events which follow, for example, the inactivation of a tumour suppressor gene, and to determine the anatomic location(s) and cell population(s) in which mutations and allelic deletions first occur[37–40]. It may also reveal the level, status (for example, with respect to phosphorylation) and binding partner(s) of the tumour suppressor protein at each stage of progression. Other effects can be explored, such as:

- the effect of loss of heterozygosity (LOH) on 'global' gene expression profiles
- the effect of allelic loss of genes within a deletion region on the corresponding genes on the homologous chromosome
- the timing and sequence of functional inactivation of all tumour suppressor genes relevant to a specific cancer

Clearly, there is a great deal of work to be done—enough to keep cancer biologists busy for a good while. It is, however, likely that knowledge gained from microarray studies of human neoplasias will be invaluable in furthering our understanding of the molecular events that underlie tumour development, and in producing new diagnostic, prognostic and therapeutic targets for the benefit of patients.

Acknowledgements

We gratefully acknowledge the contributions of the following team members and collaborators of the CGAP Molecular Fingerprinting effort. J.W. Gillespie, V.V. Prabhu, P.J. Munson, R.F. Bonner, A. Lash, D.K. Ornstein, J. Herring, L. Grouse, P.H. Duray, C.D. Vocke, J. Swalwell, M.R. Brown, C. Englert, E.F. Petricoin, S. Pan, J. Pfeifer, C. Johnson, R.D. Klausner, W. M. Linehan, L.A. Liotta and R. Strausberg.

© 1999 Nature America Inc. • http://genetics.nature.com

1. Nowak, R. Entering the postgenome era. *Science* **270**, 368–371 (1995).
2. Abbott, A. DNA chips intensify the sequence search. *Nature* **379**, 392 (1996).
3. Schena, M. *et al.* Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–469 (1995).
4. Velculescu, V. *et al*. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
5. DeRisi, J. *et al*. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet*. **14**, 457–460 (1996).
6. Schena, M. *et al.* Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16**, 301–306 (1998).
7. Ramsay, G. DNA chips: state of the art. *Nature Biotechnol.* **16**, 40–44 (1998).
8. Bowtell, D.L. Options available—from start to finish—for obtaining expression data by microarray. *Nature Genet.* **21**, 25–32 (1999).
9. Strausberg, R.L., Dahl C.A. & Klausner R.D. New opportunities for uncovering the molecular basis of cancer. *Nature Genet.* **16**, 415–416 (1997).
10. Murphy, G.P. *et al.* Histopathology of localized prostate cancer. Consensus conference on diagnosis and prognostic parameters in localized prostate cancer. *Scand. J. Urol. Nephrol. Suppl.* **162**, 7–42 (1994).
11. Epstein, J.I. Pathologic evaluation of prostatic carcinoma: critical information for the oncologist. *Oncology* **10**, 527–534 (1996).
12. Bostwick, D.G. & Montironi R. Evaluating radical prostatectomy specimens: therapeutic and prognostic importance. *Virchows Arch.* **430**, 1–16 (1997).
13. Small, E.J. Update on the diagnosis and treatment of prostate cancer. *Curr. Opin. Oncol*. **10**, 244–252 (1998).
14. Pettaway, C.A. Prognostic markers in clinically localized prostate cancer. *Tech. Urol*. **4**, 35–42 (1998).
15. Klimecki, W.T., Futscher, B.W. & Dalton, W.S. Effects of ethanol and paraformaldehyde on RNA yield and quality. *Biotechniques* **16**, 1021–1023 (1994).
16. Deng, G., Lu, Y., Zlotnikov, G., Thor, A.D. & Smith, H.S. Loss of heterozygosity in normal tissue adjacent to breast carcinomas. *Science* **20**, 2057–2059 (1996).
17. Chuaqui, R.F. *et al*. PB39: identification of a novel gene up-regulated in clinically aggressive human prostate cancer. *Urology* **50**, 302–307 (1997).
18. Zhuang, Z. *et al*. Barrett's esophagus: metaplastic cells with loss of heterozygosity at the APC gene locus are clonal precursors to invasive adenocarcinoma. *Cancer Res*. **56**, 1961–1964 (1996).
19. Hung, J. *et al.* Allele-specific chromosome 3p deletions occur at an early stage in the pathogenesis of lung carcinoma. *J. Amer. Med. A* **273**, 1908–1910 (1995).
20. Bostwick, D. & Brawer, M.K. Prostatic intra-epithelial neoplasia and early invasion in prostate cancer. *Cancer* **59**, 788–794 (1987).
21. Bostwick, D.G. Progression of prostatic intraepithelial neoplasia to early invasive adenocarcinoma. *Eur. Urol.* **30**, 145–152 (1996).
22. Isaacs, W. *et al*. genetic alterations in prostate cancer. *Cold Spring Harb. Symp. Quant. Biol.* **59**, 653–659 (1994).
23. Latil, A. & Lidereau, R. Genetic aspects of prostate cancer. *Virchows Arch.* **432**, 389–406 (1998).
24. Dong J.T., Isaacs, W.B. & Isaacs, J.T. Molecular advances in prostate cancer. *Curr. Opin. Oncol*. **9**, 101–107 (1997).
25. Lalani, E.N., Stubbs, A. & Stamp, G.W. Prostate cancer; the interface between pathology and basic scientific research. *Semin. Cancer Biol*. **8**, 53–59 (1997).
26. Hayward, S.W., Grossfeld, G.D., Tlsty, T.D. & Cunha G.R. Genetic and epigenetic influences in prostatic carcinogenesis. *Int. J. Oncol*. **13**, 35–47 (1998).
27. MacDougall, J. & Matrisian, L.M. Contribution of tumor and stromal matrix metalloproteinases to tumor progression, invasion and metastasis. *Cancer Metastasis Rev.* **14**, 351–362 (1995).
28. McCoy, K. *et al.* An acid protease secreted by transformed cells interferes with antigen processing. *J. Cell Biol.* **106**, 1879–1884 (1988).
29. Emmert-Buck, M.R. *et al*. Allelic loss on chromosome 8p12-21 in microdissected prostatic intraepithelial neoplasia (PIN). *Cancer Res.* **55**, 2959–2962 (1995).
30. Bostwick D.G. Prospective origins of prostate carcinoma. Prostatic intraepithelial neoplasia and atypical adenomatous hyperplasia. *Cancer* **78**, 330–336 (1996).
31. Eberwine, J. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. *Biotechniques* **20**, 584–591 (1996).
32. Bassett, D.E. Jr, Eisen, M.B. & Boguski, M.S. Gene expression informatics—it's all in your mine. *Nature Genet.* **21**, 51–55 (1999).
33. Emmert-Buck, M. *et al.* Increased gelatinase A and cathepsin B activity in invasive tumor regions of human colon cancer samples. *Am. J. Pathol*. **145**, 1285–1290 (1994).
34. Giometti, C.S., Williams, K. & Tollaksen, S.L. A two-dimensional electrophoresis database of human breast epithelial cell proteins. *Electrophoresis* **16**, 1187–1189 (1997).
35. Ching J., Voivodov, K.I. & Hutchens, T.W. Polymers as surface-based tethers with photolytic triggers enabling laser-induced release/desorption of covalently bound molecules. *Bioconjug. Chem.* **7**, 525–528 (1996).
36. Kuwata, H. *et al.* Bactericidal domain of lactoferrin: detection, quantitation, and characterization of lactoferricin in serum by SELDI affinity mass spectrometry. *Biochem. Biophys. Res. Commun.* **245**, 764–773 (1998).
37. Suzuki, H. *et al.* Interfocal heterogeneity of PTEN/MMAC1 gene alterations in multiple metastatic prostate cancer tissues. *Cancer Res.* **58**, 204–209 (1998).
38. Wiltshire, R. *et al.* Direct visualization of the clonal progression of primary cutaneous melanoma: application of tissue microdissection and comparative genomic hybridization. *Cancer Res.* **55**, 3954–3957 (1995).
39. Vocke, C. *et al*. Analysis of 99 microdissected prostate carcinomas reveals high frequency of allelic loss on chromosome 8p12-21. *Cancer Res.* **56**, 2411–2416 (1996).
40. Bova, G.S. & Isaacs, W.B. Review of allelic loss and gain in prostate cancer. *World J. Urol.* **14**, 338–346 (1996).
41. Emmert-Buck, M.R. *et al*. Laser capture microdissection. *Science* **274**, 998–1001 (1996).
42. Bonner, R.F. *et al.* Laser capture microdissection: molecular analysis of tissue. *Science* **278**, 1481–1483 (1997).
43. Krizman, D.B. *et al*. Construction of a representative cDNA library from prostatic intraepithelial neoplasia (PIN). *Cancer Res.* **56**, 5380–5383 (1996).
44. Levedakou, E.N. *et al.* Two novel human serine/threonine kinases with homologies to the cell cycle regulating Xenopus MO15, and NIMA kinases: cloning and characterization of their expression pattern. *Oncogene* **9**, 1977–1988 (1994).
45. Kallajokl, M., Alanen, K.A., Nevalainen, M. & Nevalainen, T.J. Group II phsopholipase A2 in human male reproductive organs and genital tumors. *Prostate* **35**, 263–272 (1998).
46. Imajoh-Ohmi, S. *et al*. Lactacystin, a specific inhibitor of the proteasome, induces apoptosis in human monoblast U937 cell. *Biochem. Biophys. Res. Commun.* **217**, 1070–1077 (1995).
47. Taylor, I.C. *et al*. Mouse mammary tumors express elevated levels of RNA encoding the murine homology of SKY, a putative tyrosine kinase. *J. Biol. Chem.* **270**, 6872–6880 (1995).

© 1999 Nature America Inc. • http://genetics.nature.com