Common markers of proliferation

Michael L. Whitfield‡, Lacy K. George*¶, Gavin D. Grant*¶ and Charles M. Perou§*

Abstract | When normal tissue and tumour samples are compared by microarray analysis, the biggest differences most often occur in the expression levels of genes that control cell proliferation. However, this difference is detected whenever mRNA samples that are taken from two cell populations with different proliferation rates are compared. Although the exact genes that comprise this 'proliferation signature' often differ, they are almost always genes that are involved in the fundamental process of cell proliferation. Can the proliferation signature be used to improve our understanding of the cell cycle and cancer pathogenesis, as well as being used as a biomarker for cancer diagnosis and prognosis?

Ki67-labelling index
Immunohistochemisty is performed on paraffin-embedded tumour samples for the nuclear antigen Ki67. The percentage of tumour cells that are positive for nuclear Ki67 labelling is determined, and this percent-positive frequency is typically predictive of patient outcomes. In general, Ki67-high patients show poor outcomes and Ki67-low patients show better outcomes.

*Department of Genetics and ‡Norris Cotton Cancer Center, Dartmouth Medical School, 7400 Remsen, Hanover, New Hampshire 03755, USA.
§Departments of Genetics and Pathology, Lineberger Comprehensive Cancer Center, CB number 7295, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.
Correspondence to M.L.W. e-mail: michael.whitfield@dartmouth.edu
¶These authors contributed equally to this work.
doi:10.1038/nrc1802

DNA microarrays are used to measure the transcript abundance, or levels of 'gene expression', of mRNA transcribed from each gene in the genome. This technology has provided a window into the genome-wide expression patterns of different cancers. Because defects in cell-cycle regulation are a fundamental feature of cancer pathogenesis, it is not surprising that genes involved in cell growth and proliferation have been observed to be expressed at high levels in almost all tumour microarray datasets. In most cases, increased expression of these genes is associated with poor outcomes in patients[1–4]. One study showed a correlation between this gene set and the proliferation rate of cells in culture[5], whereas a second study reported an association with tumour mitotic grade and proliferation-associated genes[6]. The large number of published results from tumour-profiling studies has allowed the comprehensive comparison of the expression patterns of these cell-cycle-associated genes across different tumour types, and has shown that the 'proliferation signature' — genes whose expression correlates with tumour mitotic grade, cell-cycle status, or doubling times — is a common and conserved pattern of expression in cancer cells[7]. In fact, this signature is one of the most prominent gene-expression patterns observed in tumour datasets, regardless of the tissue from which the tumour was derived. It is our hope and expectation that we can we use this signature as a defined marker of cell proliferation to better understand cancer pathogenesis and ultimately to improve therapeutics.

Proliferation-associated genes

The proliferation signature was first identified through both the gene expression profiling of human tumour samples and through a comparison of normal cell lines that had undergone cell-cycle arrest with those that were proliferating. This signature was found to be characterized by genes whose increased expression was correlated with increased proliferation rates in the cells under analysis. In 1999, the high level of expression of a specific group of genes was identified in highly proliferative breast tumour cells that were compared with samples of normal breast tissue. This analysis revealed a set of genes (which included the cell cycle transcriptional regulator *E2F1* as well as the replication-initiation complex proteins minichromosome maintenance 2–6 (MCM2–MCM6)), whose increased expression patterns were correlated with tumour cell proliferation rates, as assessed by the Ki67-labelling index[6]. Soon after, a study that analysed the gene expression profiles of diffuse large B-cell lymphomas identified a large set of genes that also included many cell-cycle-regulated genes[8]; later, it was shown that this gene set could be used to predict outcomes in patients with diffuse large B-cell lymphoma and mantle-cell lymphoma[3,9]. At the same time, microarray analysis showed that a similar group of genes was correlated with the doubling times of the 60 distinct cell lines used by the National Cancer Institute (NCI) (the NCI60 panel of cell lines) to screen for anti-cancer drugs[5]. Also, the simultaneous analysis of primary breast tumours and cell-line samples showed that the proliferation gene set was shared across tumours and cell lines, demonstrating its conserved nature (Supplementary information S1 (figure))[5,10].

Since these initial publications, examples of the proliferation signature have been identified in many different tumour datasets derived from distinct cell types. The proliferation signature is now commonly associated with the expression of a core set of genes (*MYBL2*, *BUB1* and polo-like kinase 1 (*PLK1*)), along with other cell-cycle-regulated genes such as *CCNE1*, *CCND1* and *CCNB1*(the genes that encode cyclin E1, cyclin D1 and cyclin B1, respectively). There will probably be differences in the exact genes

© 2006 Nature Publishing Group

At a glance

- The genes that are most often found when tumour and normal tissues are compared by gene expression are those involved in proliferation.
- Increased expression of the proliferation genes in tumours is often associated with poor prognoses in cancer patients.
- Cell-cycle-regulated genes have been identified by microarray analysis of cells in culture. Contained in this list are genes that are expressed in each cell cycle phase: G1/S, S, G2 and M phases.
- Comparing the proliferation genes found in tumours and cell-cycle-regulated genes found in cultured cells *in vitro* shows that there is a significant overlap between these two signatures.
- The cell-cycle-regulated genes identified by microarray analysis provide biomarkers of proliferation in both normal cells and tumours. The observation that the proliferation signature is so clearly identifiable indicates that it could be a component of genomic-based clinical diagnostics for cancer patients.

that are included in the proliferation signatures found in different tumours. These differences might be a function of the cell or tissue type analysed or the microarray platform used, but in all cases there is a core set of genes whose highest levels of expression always correlate with rapid cell proliferation. Using these criteria and a scan of the literature, we found that the proliferation signature was present in tumours of the breast[6,10,11], lung[12,13], ovary[14], prostate[15,16], liver[17] and stomach[18], as well as gliomas[19], lymphomas[3,8], and head and neck squamous cell carcinomas[20].

Further evidence for the universal nature of this pattern comes from the simultaneous analysis of multiple tumour data sets. In a simple comparison of the gene expression patterns of breast and lung carcinoma samples that were profiled using the same platform, Chung *et al.* showed that the proliferation signature — as defined by the cluster of genes that contained forkhead box M1 (*FOXM1*), *BUB1*, topoisomerase IIα (*TOP2A*) and *PLK1* — was present in both of these distinct tumour types. The signature was most commonly found in lung squamous cell carcinomas and basal-like tumours of the breast, which have many phenotypic similarities[21].

The most convincing demonstration, so far, of the universal nature of the proliferation signature comes from a meta-analysis of 40 published cancer microarray data sets that represents approximately 3,700 cancer samples[7] (see the Oncomine web site). These authors showed that the proliferation signature was present in tumour cell samples when compared with normal cell samples, and in high-grade tumour samples when compared with low-grade tumour samples. In addition, these authors extended the analysis to include a computational analysis of the promoter regions of the proliferation-associated genes. As expected, many of these genes contained binding sites for the transcription factor E2F1 (REF. 7), which (in part) controls the start of the cell-cycle programme. *E2F1* itself is also a member of the proliferation-associated gene-expression pattern, which indicates that transcription factors and their transcriptional targets are often co-regulated and therefore group together in a hierarchical clustering analysis[22,23].

Fourier analysis
A mathematical analysis of waves, discovered by the French mathematician and physicist Joseph Fourier (1768–1830). The mathematical technique can represent the data from a time series as a set of mathematical coefficients that are the amplitudes of a set of sine waves of different frequencies. Using such computational methods, one can select genes that show sinusoidal patterns as cells traverse multiple cell cycles.

Cell-cycle-regulated genes

The systematic characterization of the cell cycle on a genome-wide scale by microarray analysis has identified genes that are regulated during cell division in organisms that range from bacteria to humans[22–29], and many of these genes are also found in the proliferation signature. Therefore, an important question to be addressed is what types of genes comprise a proliferation signature and what is their relationship to the cell-cycle genes? To answer this question, the genes that are regulated during the cell division cycle *in vitro* were systematically characterized, directly comparing their expression levels to those of genes that are expressed in tumour and normal samples[22]. We will refer to these as 'cell-cycle-regulated' genes. Microarray-based screens to identify cell-cycle-regulated genes have taken several different forms, although a single type of experimental design now seems to predominate. The most common design makes use of the observation that a cell-cycle-regulated gene's expression levels should peak only once each cell division cycle. So, when mRNA expression levels or protein concentrations are measured across multiple cell cycles (FIG. 1), the expression of a cell-cycle-regulated gene would oscillate with a period that corresponds to the length of the cell cycle.

In preparing mRNA samples for analysis of cell-cycle-regulated genes, cells are typically collected at time intervals as cells traverse multiple, synchronous cell cycles, and RNA is then isolated and hybridized to DNA microarrays or analysed using traditional methods that measure the expression level of individual genes. This produces gene expression patterns that oscillate (FIG. 1). This property of the genes can be exploited to select genes mathematically by using Fourier analysis, which can be used to identify genes that show periodicity in expression. Using this approach, genes can be sorted based on the point in the cell cycle in which they show peak levels of gene expression (for example, the temporal order in which they are expressed across the cell cycle), and we are presented with a 'rolling wave' of gene expression. Genes that are found to be periodically expressed are considered to be cell-cycle-regulated by virtue of their expression levels peaking at the same point during multiple, repeated cell cycles.

Microarray analysis experiments pioneered in the budding yeast, *Saccharomyces cerevisiae*, provided the first look into the genes that are expressed during each cell division cycle in a eukaryotic organism[25,27]. These studies identified approximately 800 genes (out of 6,000) that were periodically expressed during the cell cycle, and showed that they were involved in essential, basic cell-cycle processes such as DNA replication, chromosome segregation, spindle assembly and bud formation[25,27]. Recently, several studies identified cell–cycle-regulated genes in the fission yeast, *Schizosaccharomyces pombe*, by DNA microarray analysis[28,30,31]. The study by Rustici *et al.* identified 407 periodically expressed genes[28], Peng *et al.* identified 747 cell-cycle-regulated genes[31] and Oliva *et al.* found 750 periodically expressed genes[30]. A comparison of the genes that were regulated in all

© 2006 Nature Publishing Group

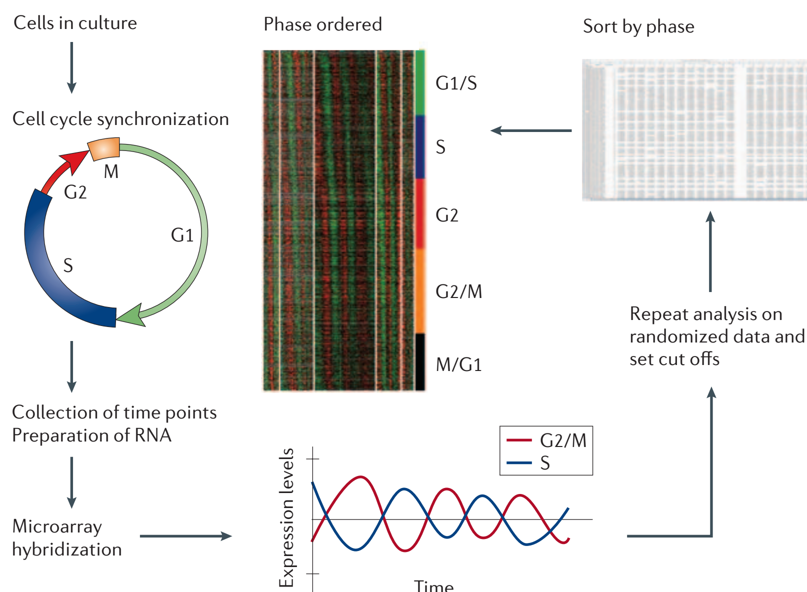


Figure 1 | **The identification of cell-cycle-regulated genes.** This diagram outlines the experimental design scheme used for the identification of cell-cycle-regulated genes. Cells in culture are synchronized using a cell-cycle-arrest agent (BOX 1), released from the synchronizing arrest, and collected at intervals. RNA is prepared and hybridized to DNA microarrays. Genes that are cell-cycle-regulated (sinusoidal patterns of expression levels over time), such as those that are upregulated during G2/M phase (red) or S phase (blue), are selected mathematically using Fourier analysis and autocorrelations. To estimate a false-positive rate, the data is randomized and the entire analysis repeated to establish the number of genes that would receive similar scores by chance, and a cutoff is established that gives a 1–5% false-positive rate. The genes are then sorted in a table and organized by which phase of the cell cycle their expression is at its highest level. Because it is difficult for the human eye to spot trends in this table, each gene expression value is converted to a coloured intensity value (highest expression levels in red, lowest expression levels in green). Using this format, expression of a gene that is cell-cycle-regulated can be easily observed to 'oscillate', peaking once each cell division cycle. When the data are displayed in the pseudo-colour format, we observe a banding pattern that is indicative of cell-cycle-regulated genes The 1,134 cell cycle-regulated genes that were identified by Whitfield *et al.*[22] are shown. The exact position of each gene can be found in the supplemental material of REF. 22 (the Human HeLa Cell-Cycle Study)

of the three *S. pombe* datasets found 170 genes that were cell-cycle-regulated[30]. Based on a comparison of genes that were found to be cell-cycle-regulated in both *S. pombe* and *S. cerevisiae*[31], the number of genes that regulate 'core' cell-cycle machinery in yeast might include 140–170 genes.

The identification of cell-cycle-regulated genes in mammalian cells has proven more difficult than in yeast. In yeast, convenient, temperature-sensitive mutants, as well as drugs and physical methods, can be used to study cell-cycle genes. In mammalian cells, less physiologically relevant methods are used to study cell-cycle-gene regulation. The first glimpse into the gene expression patterns associated with cell division in human cells came from serum stimulation experiments using fibroblasts[29]. These cells were first induced to enter a resting state, termed G0, through culture in low growth-factor conditions, and were then stimulated to re-enter the cell cycle by the addition of growth factors or serum. mRNA was then collected at different time points and expression levels were analysed by microarray hybridization.

G2 DNA damage checkpoint
DNA damage checkpoints exist to monitor the accuracy of replication fidelity and genome instability that might be caused by errors in replication, metabolic by-products, exogenous agents or extrinsic sources such as ultraviolet light or ionizing radiation. The checkpoint delays cell-cycle progression to allow the cell time for repair of the damaged DNA. The G2 DNA damage checkpoint arrests cells in G2 phase, stopping the cells from entering mitosis.

Although this method did synchronize cells, serum stimulation induces additional biological processes, such as a prominent wound-healing response. This makes it difficult to differentiate genes that are regulated by the cell cycle from those that are involved in other cell functions[29,32]. However, recent studies have shown that the expression of serum-induced genes that are not cell-cycle-regulated can also be associated with tumour progression, so various markers that have been identified through this method of cell synchronization have been useful predictors of cancer prognosis[32,33].

Several groups have performed studies of cell-cycle-regulated gene expression using serum-independent methods. Cells can be synchronized using drugs that induce arrest at specific points in the cell cycle, followed by release of the cells back into the cell cycle and the collection of samples as they progress through multiple, synchronous cell cycles (FIG. 1; BOX 1). These drugs include metabolic inhibitors that arrest cells in S phase, such as thymidine[34], hydroxyurea[35] or aphidicolin[36]; drugs that arrest cells in mitosis, such as nocodazole[37]; and physical methods that are used to select cells at specific points in the cell cycle, such as mitotic selection[38,39] or centrifugal elutriation[40] (BOX 1). These methods of synchronization avoid the confounding responses generated by serum stimulation. It has been suggested, however, that these methods do not actually synchronize cells, and the oscillatory gene-expression patterns that result are artificial and are not physiologically relevant — this topic has been extensively debated in the literature[41–45]. Nevertheless, the correlation between the genes that were identified as cell-cycle-regulated and those that were found to be highly expressed in rapidly growing tumours is remarkable and will probably be explained by a common and important biology[22].

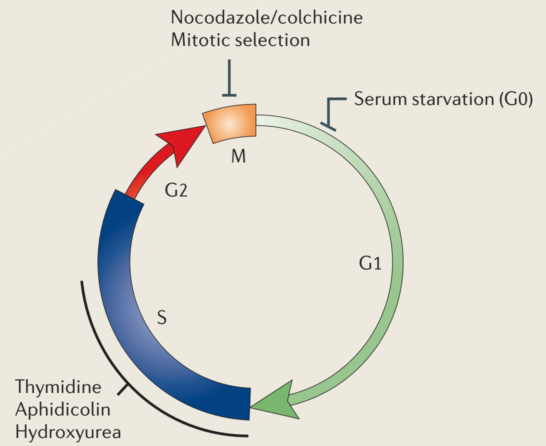
How many and which genes are cell-cycle-regulated in mammalian cells? A study of gene expression patterns that oscillate during the cell division cycle in the cervical carcinoma cell line HeLa identified more than 850 genes that show periodic expression across multiple, synchronous cell cycles[22] (FIG. 1). Additionally, approximately 700 cell-cycle-regulated genes were identified in human foreskin fibroblasts[26], and 578 cell-cycle-regulated genes were identified in mouse embryonic fibroblasts[23]. Surprisingly, there is little overlap between the genes identified in the HeLa epithelial cancer cell line[22] and those identified in normal human fibroblasts[26]; this could result from differences in cell-type-specific cell-cycle-regulated gene expression, differences in the analysis methods, or differences in the microarray platform used. A study of the G2 DNA damage checkpoint in HeLa cells found a delay in the expression of G2-phase genes, probably resulting from a delay of entry into mitosis by the cells under study[46]. Many of the genes affected by the ionizing radiation treatment were expressed at G2/M phase in the study of HeLa cells[22]. A comprehensive re-analysis of all of these gene expression datasets using the same computational algorithm is needed.

Each of the cell-cycle-regulated genes identified in the above studies show peak expression during a specific phase of the cell cycle, such as during G1/S phase,

© 2006 Nature Publishing Group

Box 1 | **Identification of cell-cycle-regulated genes**

To identify cell-cycle-regulated genes, cells must first be synchronized, and then collected for analysis at different time points. Multiple methods exist to synchronize mammalian cells, and these are different from those used to synchronize yeast cells. In mammalian cells, serum starvation, which puts cells into G0 (shown in diagram), is followed by stimulation with serum or growth factors that induce cells to re-enter the cell cycle. Serum stimulation, however, also induces expression of genes that are associated with processes other than cell-cycle progression, such as the wound-healing response. Other methods can be used to synchronize cells that avoid this confounding serum response. Agents such as thymidine, aphidicolin and hydroxyurea arrest cells in S phase (shown in diagram), whereas drugs such as nocodazole or colchicine arrest cells in mitosis (shown in diagram). Physical methods can also be used to arrest cells, including mitotic detachment (mitotic selection), and do not involve the application of drugs. During mitosis, adherent epithelial cells detach from their substrate before undergoing cytokinesis — mitotic cells can be isolated during this brief window of time and re-plated as a synchronous population of cells. The synchrony of a cell population can be determined through the use of flow cytometry, propidium iodide staining or bromodeoxyuridine labelling.



S phase, G2 phase or G2/M phase. A novel group of genes showed peak expression during M/G1 phase[22,25] (FIG. 1). In each phase, genes are expressed that are involved in the essential and basic processes necessary to duplicate a cell, such as cell-cycle control, DNA replication, DNA repair, spindle assembly and chromosome segregation. There are also genes that are periodically expressed but are not involved in these basic cell-cycle processes. In human epithelial cells, these genes are involved in other aspects of cell division such as cell adhesion, cell motility and vesicle transport. For example, adherent epithelial cells lose contact with their substrate or with other cells during M phase; this trait is exploited in mitotic shake-off synchronization experiments (BOX 1). Genes that regulate cell adhesion therefore show peak expression during M phase[47,48]. This observation indicates that such cell-type-specific cell-cycle regulation might be more significant than previously realized.

Comparing cell-cycle and proliferation genes

With a list of cell-cycle-regulated genes, it is possible to directly test the hypothesis that tumour-proliferation-associated genes are cell-cycle-regulated. A comparison of cell-cycle-regulated genes, identified by microarray analysis of cells grown in culture, versus the proliferation signature found in tumours (Supplementary information S1 (figure)) showed a significant overlap between the two groups of genes[22] (FIG. 2). The overlap between these gene sets clearly shows that proliferation-associated genes expressed by tumours are mainly cell-cycle-regulated, and therefore, the cell-cycle-regulated gene set is not only valuable for *in vitro* analysis of proliferation, but is also a marker of the proliferation of tumour cells *in vivo*. The number of genes in this list is large and cannot be easily conveyed in the space constraints of this review. Nevertheless, important genes that have been included in both lists include Ser/Thr kinase 6 (*STK6*), *PLK1*, *E2F1*, *FOXM1*, *MKI67* and the MCM genes. Supplementary information S1 (figure) shows the expression patterns associated with these genes in breast tumours[10].

Although the comparison of HeLa cell-cycle-regulated genes with the breast tumour profiles was based on a hierarchical clustering analysis, other statistical methods can be used to examine relationships between gene sets. For example, it is possible to determine whether specific gene ontology (GO) terms are statistically over-represented using software such as GO::TermFinder (REF. 49). As shown in FIG. 2, we used this tool to calculate the biological processes most commonly associated with the products of the genes included in the proliferation signature of breast tumour samples[10] (FIG. 2, Supplementary information S2 (table)). It is also possible to compare a list of genes to proliferation signatures identified in breast tumours[6,10] or lung tumours[13], or to the cell-cycle-regulated-gene signature[22,26] set using analytical tools such as Gene Set Enrichment Analysis. This tool allows researchers to determine if there is a statistically significant enrichment of a particular gene set in a list of genes that has been ordered by statistical significance[50,51].

These findings have several important implications. First, cell-cycle genes are probably regulated in such a way that they are expressed only when cells have entered the cell division programme and not at any other time. So, their prominent levels of expression in tumours might simply be a consequence of the fact that tumours contain more cycling cells relative to other tumours or to normal tissues. This is not surprising, as during transformation to malignancy, control of the cell cycle is deregulated; by its inherent nature, a tumour contains more cycling cells than normal tissue. This same scenario presents itself when proliferating cells and resting cells grown *in vitro* are compared. When one examines a population of proliferating cells, the proliferation genes are highly expressed; and when one examines a population of

© 2006 Nature Publishing Group

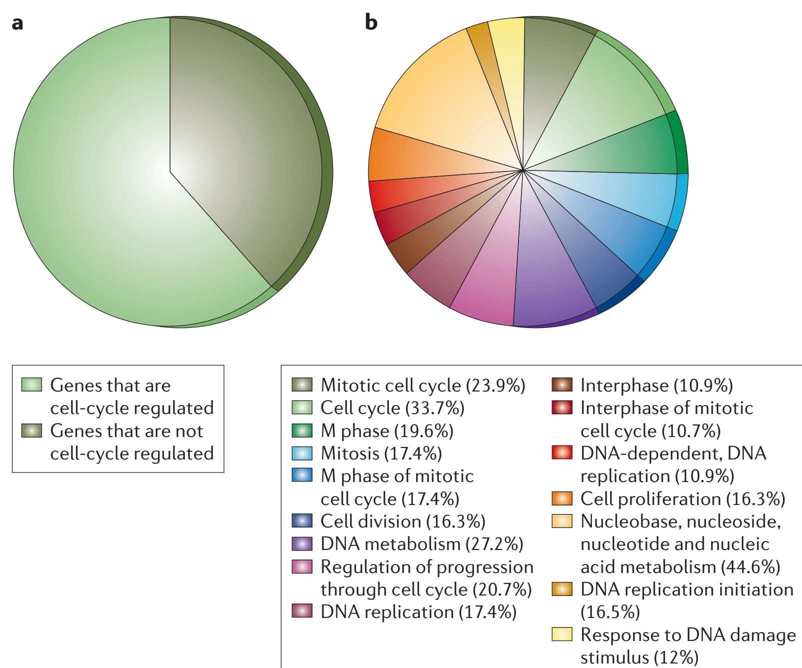


Figure 2 | **The proliferation-associated genes are cell-cycle-regulated. a** | A comparison of the cell-cycle-regulated genes[22] that were identified in HeLa cells with genes included in the breast tumour proliferation signature[10] shows that 62% (69/112) of the breast-tumour-proliferation-signature genes are cell-cycle-regulated. **b** | Through analysis of the Gene Ontology *Biological Process* annotations, 92 of the 112 genes in the proliferation cluster were examined using GO::TermFinder (see REF. 49). Forty-two terms were significantly enriched with Bonferroni-corrected *p*-values less than 0.01 (percentage enrichment indicated in brackets)[49]. Based on this analysis, the 15 gene ontology biological process terms that had the most statistically significant alterations in the breast tumour proliferation signature are graphed in the pie chart. Each section of the pie chart represents the number of genes that map to each term. All significantly enriched biological process annotations can be found in Supplementary information S2 (table). The definitions of all terms can be found on the Gene Ontology web site.

resting cells or a population with fewer cycling cells in which the non-cycling cells are in G0, the expression of these genes is turned off.

As might be expected, not all cell-cycle-regulated genes are correlated with tumour cell proliferation. Inspection of the gene expression patterns of all the genes identified as cell-cycle-regulated in HeLa cells[22] in the same study of breast tumours described above[6] showed that some cell-cycle-regulated genes were expressed in a non-correlated, or even anti-correlated, manner; these include genes for cell adhesion (for example, *CNN2*, the gene that encodes calponin 2, and *VCL*, the gene that encodes vinculin), growth control (for example, *CDKN2C*, the gene that encodes cyclin-dependent kinase inhibitor 2C (also known as p18)) and apoptosis (for example, *CASP3*, the gene that encodes caspase 3). These results indicate that the regulation of some genes is more complex than the simple restriction of transcription to certain phases of the cell cycle and could reflect cell line- or cell type-specific patterns of expression. Nonetheless, a significant fraction of *in vitro*-defined cell-cycle-regulated genes are present in the tumour-proliferation-associated profiles.

Clinical implications

Because the proliferation signature is easy to identify, it could be a component of genomic-based clinical diagnostics for cancer patients. A histologically based proliferation test has been a standard part of many pathologists' routine clinical assessment. In fact, the grading of breast cancers as low grade versus high grade is in part based on the number of mitotic cells present within a given sample, although this can be a subjective assay with significant levels of inter-observer variation[2]. Gene expression analysis therefore provides some significant advantages over the pathological assessment of proliferation, including the fact that it is more quantitative, objective, and can be automated. The use of cell-cycle-regulated genes as prognostic markers is not new. The Ki67-based labelling index is commonly used in clinical assessment, and the expression levels of proliferating cell nuclear antigen (*PCNA*) can also be used. Proteins encoded by the MCM genes have also been proposed as useful markers of proliferation, and high levels of expression of these genes is indicative of poor prognosis[52–55]. All of these genes are cell-cycle-regulated and are found among the genes associated with proliferation in tumours.

If measuring the expression level of a single cell-cycle-regulated gene (*MKI67* or the MCM genes) is a useful prognostic indicator and marker of proliferation, might measuring all of the genes associated with proliferation make the assay more robust? As expected, in almost all studies in which a multi-gene proliferation signature was correlated with patient outcome, it predicted poor patient outcome based on endpoints such as overall survival, relapse-free survival and/or the presence of metastases[1–4,9]. A recent study reported that the proliferation signature was a better predictor of outcome in patients with breast cancer than a tumour grade assay used by pathologists[2]. An 11 gene 'stem cell/*BMI1*-oncogene-associated' expression pattern has been identified in metastatic prostate cancer cells that contains a subset of the proliferation signature, including increased expression of *MKI67*. This pattern was associated with poor outcomes in patients with a variety of tumour types[56]. Gene expression analysis has therefore shown that the proliferation signature is a common feature of tumours and might provide a robust and objective assay for proliferation in the clinic[3,4,9,57].

In terms of therapy, almost all cytotoxic chemotherapeutic drugs arrest dividing cells by causing DNA damage or targeting products of proliferation-signature genes (TABLE 1). Some arrest cells in S phase by interfering with ribonucleotide synthesis or disrupting the basic DNA replication machinery. Others, such as cisplatin, cause DNA damage, which leads to the arrest of cycling cells in G1 or early S phase. Drugs that arrest cells in mitosis, such as the taxanes, interfere with mitotic spindle organization. Some drugs have been designed to target the CDKs that, along with their cyclin partners, are required to drive the cell through the various stages of the cycle[58,59]. Genes that encode CDKs are transcribed constitutively and post-translational processes regulate their activity, so the CDKs tend to be absent from the

© 2006 Nature Publishing Group

Table 1 | **Proliferation targets of chemotherapeutic agents**

Drug	Target	Mechanism of action	Phase affected	References
5-fluorouracil	Thymidylate synthetase	Ribonucleotide depletion	S	64
Hydroxyurea	Ribonucleotide reductase	Ribonucleotide depletion	S	64
Methotrexate	Dihydrofolate reductase	Ribonucleotide depletion	S	64
Doxorubicin	Topoisomerase II	Stabilization of Topoisomerase II–DNA complex, which leads to replication arrest	S	65
Etoposide	Topoisomerase II	Stabilization of Topoisomerase II–DNA complex, which leads to replication arrest and strand breakage	S or G2/M	65,66
Cisplatin	DNA	Intra-strand crosslinking, which leads to replication arrest	G1/S and G2/M	67
Taxol	Tubulin	Stabilization of microtubules	G2/M	58
Flavopiridol	CDK1, CDK2, CDK4, CDK6 and CDK7	Inhibition of CDKs by interfering with ATP-binding	Several	59
Staurosporine (UCN-01)	CDK1	Inappropriate activation of CDK1 by phosphorylation	G2/M abrogated	58,59

CDK, cyclin-dependent kinase

proliferation signature. Any drug that targets gene products of the proliferation signature or products of cell-cycle-regulated transcripts will probably affect not only cancer cells but also any normal dividing cell, which leads to the side effects of traditional chemotherapeutics. So, drugs developed against genes or products of the proliferation cluster could also result in toxicity to normal tissues. As targeted therapies become more specific for particular features of tumours, considerations for the presence of these targets in normal cells must be taken into account.

The number of microarray datasets of gene expression in tumours is increasing rapidly, and a complete comparison of all cell-cycle-regulated genes versus those that are deregulated in tumours is beyond the scope of this review. Which tumour-suppressor genes and oncogenes are present in the proliferation signature? A strictly curated census of cancer genes has recently been published that carefully examined the evidence that somatic mutation, amplification or chromosomal translocation in a particular gene resulted in cancer development[60] (see the Cancer Gene Census web site). Comparison of the genes in that database with proliferation signature genes, or with cell-cycle-regulated genes, shows that oncogenes or tumour-suppressor genes are not necessarily differentially expressed. For example, the tumour suppressors breast cancer 1 (*BRCA1*) and *TP53* are notably absent from the list of proliferation signature genes. This is probaby because tumour-suppressor gene expression or function is typically lost during tumour growth and progression, and are therefore not included in the proliferation signature. Alternatively, the genetic alterations that activate oncogenes typically result in expression of the proliferation genes — so even though oncogene mutations can upregulate the proliferation signature, these genes are not usually included in it.

Future directions

The proliferation signature of tumour cells is comprised of cell-cycle-regulated genes whose pattern of expression is altered because a tumour contains more proliferating cells than normal. The challenge lies in using this expression pattern to better understand how these genes become upregulated in cancer cells and the biological function of these gene products. Studies of the proliferation signature might also identify new drug targets as well as new diagnostic and prognostic uses. Compared with the assays for tumour cell proliferation rates performed by pathologists, gene expression analysis could result in clinical diagnostic assays that are more objective, quantitative and automated. In fact, the first gene-expression-based clinical assay that contains a proliferation-gene-expression profile has recently made it to the clinical setting as an assay for diagnosing patients with breast cancer[2].

It will also be important to uncover the roles of the more than 450 cell-cycle-regulated genes whose functions have not been identified that are included in the proliferation signature. It might be possible to identify a ‘core set’ of cell-cycle-regulated genes that is required for cell-cycle progression in all eukaryotic cells. Many cell-cycle genes are functionally conserved across eukaryotic organisms. For example, *CDC28* (cell division cycle 28) from *S. cerevisiae*[61] and *CDC2* from humans can both complement the phenotype of a *cdc2*-null strain of *S. pombe*[62], demonstrating the functional conservation of such core cell-cycle genes. Computational methods are being developed that can predict the function of gene products, based on expression patterns, analysis of the gene promoter sequence, and protein–protein interactions of gene products[63]. Using these types of experimental and computational approaches in model systems, we might be able to better understand the roles of proliferation-signature-gene

© 2006 Nature Publishing Group

products in tumour biology and, ultimately, design improved anticancer treatments.

Most importantly, analysis of proliferation signatures has revealed a striking commonality between cell-cycle-regulated genes and genes that are commonly upregulated in cancer cells. These analyses reiterate a common theme in genomic research — that there is an amazing conservation of genes and gene product function across organisms and among diseases. The continued systematic characterization of tumours and other diseases by microarray analysis, combined with the analysis of pathways in cell culture and model systems, will provide rich sets of data for further analysis. The proliferation signature will probably continue to be an important feature of cancer biology.

1. Dai, H. *et al.* A cell proliferation signature is a marker of extremely poor outcome in a subpopulation of breast cancer patients. *Cancer Res.* **65**, 4059–4066 (2005).
2. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
One of the first clinical assays based on gene expression levels.
3. Rosenwald, A. *et al.* The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell* **3**, 185–197 (2003).
4. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
5. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.* **24**, 227–235 (2000).
6. Perou, C. M. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96**, 9212–9217 (1999).
The first study to correlate a gene expression profile with proliferation (the KI67-labelling index) in normal and tumour tissues. It reported that the proliferation profile was conserved across cell lines and primary tumour types.
7. Rhodes, D. R. *et al.* Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl Acad. Sci. USA* **101**, 9309–9314 (2004).
This study showed the universality of the proliferation signature by showing that it is a common feature across more than 40 distinct tumour data sets, comparing low-grade tumours with high-grade tumours.
8. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
9. Rosenwald, A. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346**, 1937–1947 (2002).
10. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
This study correlated tumour mitotic grade with the proliferation signature.
11. van't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
12. Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl Acad. Sci. USA* **98**, 13790–13795 (2001).
13. Garber, M. E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl Acad. Sci. USA* **98**, 13784–13789 (2001).
14. Welsh, J. B. *et al.* Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl Acad. Sci. USA* **98**, 1176–1181 (2001).
15. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA* **101**, 811–816 (2004).
16. LaTulippe, E. *et al.* Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res.* **62**, 4499–4506 (2002).
17. Chen, X. *et al.* Gene expression patterns in human liver cancers. *Mol. Biol. Cell* **13**, 1929–1939 (2002).
18. Hippo, Y. *et al.* Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res.* **62**, 233–240 (2002).
19. Rickman, D. S. *et al.* Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res.* **61**, 6885–6891 (2001).
20. Chung, C. H. *et al.* Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* **5**, 489–500 (2004).
21. Chung, C. H., Bernard, P. S. & Perou, C. M. Molecular portraits and the family tree of cancer. *Nature Genet.* **32** (Suppl.), 533–540 (2002).
22. Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13**, 1977–2000 (2002).
A comprehensive study of the cell-cycle-regulated genes during the cell divison cycle in a human cancer cell line. This study was the first to show that cell-cycle-regulated gene sets overlap with proliferation signatures in tumour samples.
23. Ishida, S. *et al.* Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol. Cell. Biol.* **21**, 4684–4699 (2001).
24. Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M. & Shapiro, L. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**, 2144–2148 (2000).
25. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
26. Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nature Genet.* **27**, 48–54 (2001).
27. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
28. Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nature Genet.* **36**, 809–817 (2004).
29. Iyer, V. R. *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83–87 (1999).
30. Oliva, A. *et al.* The cell cycle-regulated genes of *Schizosaccharomyces pombe*. *PLoS Biol.* **3**, e225 (2005).
31. Peng, X. *et al.* Identification of cell cycle-regulated genes in fission yeast. *Mol. Biol. Cell* **16**, 1026–1042 (2005).
32. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, e7 (2004).
33. Chang, H. Y. *et al.* Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl Acad. Sci. USA* **102**, 3738–3743 (2005).
34. Bootsma, D., Budke, L. & Vos, O. Studies on synchronous division of tissue culture cells initiated by excess thymidine. *Exp. Cell. Res.* **33**, 301–309 (1964).
35. Reichard, P. & Ehrenberg, A. Ribonucleotide reductase — a radical enzyme. *Science* **221**, 514–519 (1983).
36. Huberman, J. A. New views of the biochemistry of eucaryotic DNA replication revealed by aphidicolin, an unusual inhibitor of DNA polymerase α. *Cell* **23**, 647–648 (1981).
37. Zieve, G. W., Turnbull, D., Mullins, J. M. & McIntosh, J. R. Production of large numbers of mitotic mammalian cells by use of the reversible microtubule inhibitor nocodazole. Nocodazole accumulated mitotic cells. *Exp. Cell. Res.* **126**, 397–405 (1980).
38. Eliassen, K. A., Baldwin, A., Sikorski, E. M. & Hurt, M. M. Role for a YY1 binding site in replication-dependent mouse histone gene expression. *Mol. Cell. Biol.* **18**, 7106–7118 (1998).
39. Schneiderman, M. H., Dewey, W. C., Leeper, D. B. & Nagasawa, H. Use of the mitotic selection procedure for cell cycle analysis. Comparison between the X-ray and cycloheximide G2 markers. *Exp. Cell. Res.* **74**, 430–438 (1972).
40. Grabske, R. J., Lake, S., Gledhill, B. L. & Meistrich, M. L. Centrifugal elutriation: separation of spermatogenic cells on the basis of sedimentation velocity. *J. Cell. Physiol.* **86**, 177–189 (1975).
41. Cooper, S. Rejoinder: whole-culture synchronization cannot, and does not, synchronize cells. *Trends Biotechnol.* **22**, 274–276 (2004).
42. Cooper, S. Is whole-culture synchronization biology's 'perpetual-motion machine'? *Trends Biotechnol.* **22**, 266–269 (2004).
43. Spellman, P. T. & Sherlock, G. Final words: cell age and cell cycle are unlinked. *Trends Biotechnol.* **22**, 277–278 (2004).
44. Spellman, P. T. & Sherlock, G. Reply: whole-culture synchronization — effective tools for cell cycle studies. *Trends Biotechnol.* **22**, 270–273 (2004).
45. Shedden, K. & Cooper, S. Analysis of cell-cycle-specific gene expression in human cells as determined by microarrays and double-thymidine block synchronization. *Proc. Natl Acad. Sci. USA* **99**, 4379–4384 (2002).
46. Crawford, D. F. & Piwnica-Worms, H. The G2 DNA damage checkpoint delays expression of genes encoding mitotic regulators. *J. Biol. Chem.* **276**, 37166–37177 (2001).
47. Rabouille, C. & Jokitalo, E. Golgi apparatus partitioning during cell division. *Mol. Membr. Biol.* **20**, 117–127 (2003).
48. Shorter, J. & Warren, G. Golgi architecture and inheritance. *Annu. Rev. Cell Dev. Biol.* **18**, 379–420 (2002).
49. Boyle, E. I. *et al.* GO::TermFinder — open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715 (2004).
50. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
51. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
52. Gonzalez, M. A., Tachibana, K. E., Laskey, R. A. & Coleman, N. Control of DNA replication and its potential clinical exploitation. *Nature Rev. Cancer* **5**, 135–141 (2005).
53. Gonzalez, M. A. *et al.* Geminin predicts adverse clinical outcome in breast cancer by reflecting cell-cycle progression. *J. Pathol.* **204**, 121–130 (2004).
54. Gonzalez, M. A. *et al.* Minichromosome maintenance protein 2 is a strong independent prognostic marker in breast cancer. *J. Clin. Oncol.* **21**, 4306–4313 (2003).
55. Chatrath, P. *et al.* Aberrant expression of minichromosome maintenance protein-2 and Ki67 in laryngeal squamous epithelial lesions. *Br. J. Cancer* **89**, 1048–1054 (2003).

© 2006 Nature Publishing Group

56. Glinsky, G. V., Berezovska, O. & Glinskii, A. B. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J. Clin. Invest.* **115**, 1503–1521 (2005).
57. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl Acad. Sci. USA* **100**, 8418–8423 (2003).
58. Johnson, D. G. & Walker, C. L. Cyclins and cell cycle checkpoints. *Annu. Rev. Pharmacol. Toxicol.* **39**, 295–312 (1999).
59. Senderowicz, A. M. The cell cycle as a target for cancer therapy: basic and clinical findings with the small molecule inhibitors flavopiridol and UCN-01. *Oncologist* **7** (Suppl.), 12–19 (2002).
60. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
61. Beach, D., Durkacz, B. & Nurse, P. Functionally homologous cell cycle control genes in budding and fission yeast. *Nature* **300**, 706–709 (1982).
62. Lee, M. G. & Nurse, P. Complementation used to clone a human homologue of the fission yeast cell cycle control gene *cdc2*. *Nature* **327**, 31–35 (1987).
63. Troyanskaya, O. G., Dolinski, K., Owen, A. B., Altman, R. B. & Botstein, D. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc. Natl Acad. Sci. USA* **100**, 8348–8353 (2003).
64. Allegra, C. J. & Grem, J. L. in *Cancer: Priciples and Practice of Oncology* (eds. DeVita, V. T., Hellman, S. & Rosenberg, S. A.) 432–452 (Lippincott–Raven, Philadelphia, 1997).
65. Hurley, L. H. DNA and its associated processes as targets for cancer therapy. *Nature Rev. Cancer* **2**, 188–200 (2002).
66. Stewart, C. F. & Ratain, M. J. in *Cancer: Principles and Practice in Oncology* (eds. DeVita, V. T., Hellman, S. & Rosenberg, S. A.) 452–467 (Lippincott–Raven, Philadelphia, 1997).
67. O'Dwyer, P. J., Johnson, S. W. & Hamilton, T. C. in *Cancer Principles and Practices of Oncology* (eds. DeVita, V. T., Hellman, S. & Rosenberg, S. A.) 418–432 (Lippincott–Raven, Philadelphia, 1997).
68. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nature Genet.* **36**, 1090–1098 (2004).

Acknowledgements

The authors would like to acknowledge the following people and institutions: Gavin Sherlock and Howard Y. Chang for critical reading of the manuscript, and members of the Whitfield laboratory for helpful discussions. M.L.W. is supported by Howard Hughes Medical Institute Biomedical Research award to Dartmouth College, the American Cancer Society, and by grants from the V Foundation for Cancer Research and the Scleroderma Research Foundation. M.L.W. is a V Scholar of the V Foundation for Cancer Research. C.M.P. is supported by the National Cancer Institute Breast SPORE programme, by the National Institute of Environmental Health Sciences and by the National Cancer Institute.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
BUB1 | *E2F1* | *MYBL2* | *PLK1* | *STK6*
National Cancer Institute: http://www.cancer.gov
breast tumour | lung cancer | lymphoma

SUPPLEMENTARY INFORMATION

See Online article: S1 (figure) | S2 (table)

FURTHER INFORMATION

Cancer Gene Census at the Sanger Centre: http://www.sanger.ac.uk/genetics/CGP/Census/
Gene Ontology: http://www.geneontology.org/
Gene Set Enrichment Analysis: http://www.broad.mit.edu/gsea/
GO::TermFinder: http://search.cpan.org/dist/GO-TermFinder/
Human HeLa Cell-Cycle Study: http://genome-www.stanford.edu/human-cellcycle/HeLa/
Human Fibroblast Cell-Cycle Study: http://www-sequence.stanford.edu/human_cell_cycle/index.html
Microarray analysis of the NCI60 cell lines: http://discover.nci.nih.gov/
Molecular Portraits of breast cancer: http://genome-www.stanford.edu/breast_cancer/molecularportraits/
Oncomine: http://www.oncomine.org
Stanford Breast Cancer Portal: http://genome-www.stanford.edu/breast_cancer/
The NCI60 Cancer Microarray Project: http://genome-www.stanford.edu/nci60/
Yeast Cell-Cycle Study: http://genome-www.stanford.edu/cellcycle/
Access to this links box is available online.

© 2006 Nature Publishing Group