nature genetics

Mining for regulatory programs in the cancer transcriptome

Daniel R Rhodes[1–3], Shanker Kalyana-Sundaram[1], Vasudeva Mahavisno[1], Terrence R Barrette[1], Debashis Ghosh[2,4] & Arul M Chinnaiyan[1–3,5]

DNA microarrays have been widely applied to cancer transcriptome analysis. The Oncomine database contains a large collection of such data, as well as hundreds of derived gene-expression signatures. We studied the regulatory mechanisms responsible for gene deregulation in these cancer signatures by searching for the coordinate regulation of genes with common transcription factor binding sites. We found that genes with binding sites for the archetypal cancer transcription factor, E2F, were disproportionately overexpressed in a wide variety of cancers, whereas genes with binding sites for other transcription factors, such as Myc-Max, c-Rel and ATF, were disproportionately overexpressed in specific cancer types. These results suggest that alterations in pathways activating these transcription factors may be responsible for the observed gene deregulation and cancer pathogenesis.

Global gene expression profiling with DNA microarrays has been widely applied to human cancer, leading to the elucidation of complex gene-expression programs activated and repressed in various types and subtypes of cancer, a 'molecular taxonomy' of cancer. We and others have attempted to characterize large collections of cancer gene-expression data in terms of common 'signatures' of activation[1,2] or in terms of coordinately regulated processes or 'modules'[3]. But such efforts have not focused on the regulatory mechanisms responsible for observed gene-expression alterations in cancer. Some gene-expression patterns observed from microarray data probably represent a downstream read-out of a few genetic aberrations (mutations, amplifications, deletions, translocations, etc.) that led to the activation or inactivation of a few transcription factors. In some cases, cancer-causing genetic aberrations may not be directly apparent from these downstream gene-expression read-outs. For example, a mutation in the Rb tumor suppressor that leads to dissociation and activation of E2F1 would manifest not in the differential gene-expression of Rb or E2F1, but in the coordinate activation of E2F target genes. Global methods for inferring transcriptional regulatory mechanisms from gene-expression data have been widely applied to yeast gene expression and also to the human cell cycle[4] but not yet to human cancer. We searched for cancer regulatory programs that link transcription factors to target genes that are conditionally activated in specific cancer types and subtypes (**Fig. 1**).

We began by defining gene-expression signatures characteristic of a wide variety of cancer types and subtypes represented in the Oncomine database[5]. We used data from 65 independent studies including 6,732 microarray experiments and 70.8 million gene-expression measurements to derive 265 gene-expression signatures (**Supplementary Table 1** online). Signatures were defined as sets of genes with statistically significant ($Q < 0.10$) differential expression in cancer, either relative to normal tissue or relative to other types or subtypes of cancer. We also derived normal tissue signatures as sets of genes differentially expressed in a single normal tissue type relative to other normal tissue types. Gene-expression signatures ranged in size from 20 genes to 2,200 genes and represented nearly every major type of cancer and normal tissue.

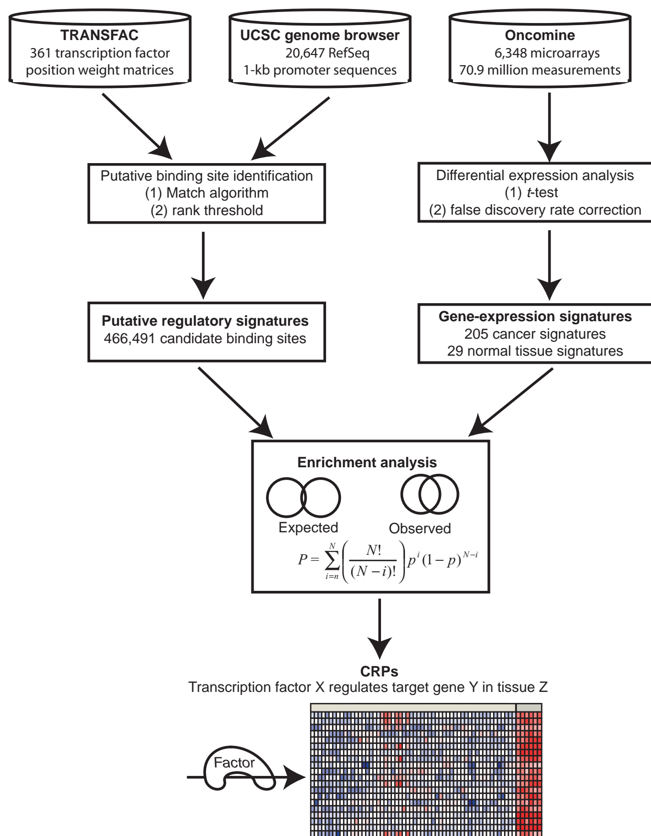
Next, we constructed a database of transcriptional regulatory signatures, relating transcription factors to candidate target genes by identifying putative transcription factor binding sites in the promoter sequences of human genes. We submitted all 1-kb human promoter sequences to the MATCH software program, which identifies and scores sequence matches to transcription factor binding site position weight matrices from the TRANSFAC database[6]. Although a high-scoring match does not constitute a definitive transcription factor binding site and regulatory interaction, we reasoned that the sets of genes with the highest scoring matches are likely to be enriched for true target genes. After we applied a match threshold and rank filter, our database contained 361 regulatory signatures, comprising 466,491 potential regulatory interactions that represent putative transcription factor binding sites in the promoters of candidate target genes (**Supplementary Table 2** online). There are several limitations to this approach, and we concede that our database is incomplete and probably contains false interactions. But the database is sufficient for large-scale enrichment analysis as well as initial hypothesis generation.

With a database of gene-expression signatures and transcriptional regulatory signatures in place, we sought to identify conditional regulatory programs (CRPs), consisting of a transcription factor that coordinately regulates a set of target genes in a particular tissue type. We identified candidate CRPs by searching for disproportionate overlap of regulatory signatures with gene-expression signatures. We reasoned that if a transcription factor is responsible for the coordinate regulation

[1]Department of Pathology, [2]Bioinformatics Program, [3]Comprehensive Cancer Center, and Departments of [4]Biostatistics and [5]Urology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. Correspondence should be addressed to A.M.C. (arul@umich.edu).

Published online 26 May 2005; doi:10.1038/ng1578

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics



$$P=\sum_{i=n}^{N}\left(\frac{N!}{(N-i)!}\right)p^{i}(1-p)^{N-i}$$

Figure 1 Overview of the method used to elucidate CRPs. Data were integrated from three sources: TRANSFAC[6], the University of California Santa Cruz (UCSC) genome browser and Oncomine[5]. Putative transcription factor regulatory signatures were compared with gene-expression signatures, and their overlap was assessed using the binomial distribution to derive CRPs.

of a set of genes in a given tissue type, then the candidate target genes of the transcription factor should be disproportionately over-represented in the gene-expression signature. We compared 265 gene expression signatures with 366 regulatory signatures. We counted the degree of overlap between all signature pairs and computed the significance of the overlaps by the binomial distribution. From this analysis, we defined 311 regulatory programs that showed highly significant overlap ($P <$ 0.00033) between a gene-expression signature and a regulatory signature (**Supplementary Tables 3–6** online). Given the total number of hypotheses tested, we would expect only 31 signature pairs to have such significant overlap by chance ($Q < 0.10$; **Fig. 2a**).

We first examined the 81 CRPs specific to normal human tissues. Several of the programs validated our method by specifically linking a transcription factor to the tissue type in which it is known to act. For example, CRPs 11 and 58 are composed of genes activated in normal liver tissue that have promoter binding sites for HNF4α and HNF1, respectively, two hepatocyte nuclear factors known to control liver-specific processes[7,8] (**Fig. 2b**). CRP 222 is composed of genes activated in muscle tissue that have promoter binding sites for MEF2A (also called RSRFC4), a transcription factor with a known role in myocyte differentiation[9]. CRP 24 is composed of genes activated in normal brain tissue that have promoter binding sites for the EGR (also called KROX) family of transcription factors[10]. CRPs 32, 136 and 307 are composed of genes activated in normal blood cells that have binding sites for IRF family of transcription factors, which are activated in white blood cells and have a role in host defense[11]. Similarly, binding sites for NF-κB, which has a central role in immune function, are also enriched in genes expressed in blood cells (CRPs 178 and 264)[12]. The gene-expression signature for early progenitor cells showed an enrichment for Oct/POU family transcription factors, which function in pluripotent stem cells[13,14]. Taken together, these results show that our approach can identify regulatory mechanisms responsible for gene regulation in human tissue. They also suggest sets of target genes for each of these tissue-specific regulatory programs. Other normal tissue CRPs may link transcription factors to tissue types in which they were not previously known to act. The full list of normal tissue CRPs is provided in **Supplementary Table 4** online and can be explored in detail at Oncomine.

Next, we examined the 232 CRPs involving human cancer. More than half (126) of these relate one of several variant E2F binding sites to target genes in one of many cancer types, including follicular lymphoma, Burkitt lymphoma, diffuse large B-cell lymphoma (DLBCL), acute lymphoblastic leukemia, glioblastoma, medulloblastoma, leiomyosarcoma, small cell lung cancer (SCLC), squamous cell lung cancer, hepatocellular carcinoma, salivary adenoid cystic carcinoma, adrenocortical carcinoma, high-grade astrocytoma and high-grade breast carcinoma. These results reaffirm that activation of the E2F pathway is a prevalent event in human cancer[15,16] and provide hundreds of putative E2F targets activated in specific human tumors. As others have found, we show through a second layer of enrichment analysis that E2F CRPs include genes involved in several cellular proliferation related processes such as the cell cycle, DNA replication and mRNA splicing. We also show that several E2F cancer regulatory programs, including those activated in high-grade breast cancer (CRP 99) and small cell lung cancer (CRP 96), are enriched for proteins involved in chromatin modification, including EZH2, JJAZ1 (SUZ12), CBX3, HMGA1 and BAF53A. The E2F pathway regulates EZH2 (ref. 17); perhaps its role in regulating chromatin modifying genes is more widespread. Regulatory programs linking NF-Y binding sites to cancer signatures were also common and usually coincided with E2F cancer regulatory programs. This is not surprising as the binding of the NF-Y transcription factor to certain promoters is necessary for E2F activity[16]. NF-Y transactivation is dependent on phosphorylation by CDK2, suggesting a potential therapeutic approach for repressing NF-Y and thus E2F cancer regulatory programs[18].

To confirm that the E2F cancer regulatory programs represent gene sets truly activated by E2F, we collected data from an independent study that identified transcriptional targets of the E2F family in an inducible cell line system[19]. In total, 558 genes were significantly overexpressed upon E2F activation. We reasoned that if at least a fraction of these results represented a physiologically relevant E2F signature, and if our E2F cancer regulatory programs represented valid programs activated by E2F in human cancer *in vivo*, then we should find substantial overlap between the two. To test this, we selected a representative E2F binding site and its ten respective cancer regulatory programs. In nine of the ten CRPs, we found a significant enrichment of genes from the *in vitro* E2F signature ($P < 0.005$), suggesting that our approach identified valid E2F targets activated in human cancers (**Supplementary Table 7** online).

To select the most promising candidate E2F targets among our CRPs, we identified target genes that are activated by E2F in the inducible cell culture system and are most common in E2F CRPs. Among the nine CRPs that showed significant overlap with the *in vitro* signature, eight contained three known E2F targets, including CCNE2, RRM2 (ref. 20) and EZH2 (ref. 17). Other known E2F targets activated in many cancer CRPs include TFDP1, CDC25, RPA1 and USP13. The near universal

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

activation of these genes by E2F in cancer suggests that they are crucial mediators of carcinogenesis. For example, the observed activation of cyclin E2 by E2F is part of an autoregulatory loop, as cyclin E2 activates CDKs, which further activate E2F[21]. Most of these E2F target genes have a role in cellular proliferation; therefore, their importance in E2F-mediated tumorigenesis is not surprising. Notably, however, EZH2 functions as a chromatin-modifying transcriptional repressor and is important in embryogenesis[22]. Functional studies showed that EZH2 promotes invasion in breast cancer cells[23] and is associated with a lethal phenotype in prostate cancer[24]. EZH2 is also associated with poorly differentiated cancers relative to their well-differentiated counterparts[2]. Perhaps the hyperactivated E2F pathway leads to EZH2 overexpression and concomitant repression of prodifferentiation genes, thus locking cells into an undifferentiated invasive phenotype. This raises the possibility that the E2F pathway is responsible for both cancer growth and dedifferentiation.

Our analysis uncovered several other cancer regulatory programs involving transcription factors other than E2F. For example, CRP 120 suggests that c-Rel activates hundreds of target genes in DLBCL, and CRP 45 suggests that c-Rel activity is most apparent in *de novo* DLBCL relative to transformed DLBCL (**Fig. 2c**). The enrichment of c-Rel binding sites in the promoters of genes activated in DLBCL is consistent with the observation of c-Rel amplification in DLBCL[25]. Although one report failed to find a link between c-Rel amplification status and downstream gene-expression changes[26], our results suggest that c-Rel activity is evident from DLBCL gene-expression patterns. Transformed and *de novo* DLBCL are markedly different at the gene-expression level[27], though morphologically indistinguishable. Our work suggests that a key difference may be the specific activation of the c-Rel regulatory program in *de novo* DLBCL. Upon examination of the target genes in the c-Rel–DLBCL program, we found an enrichment of genes involved in both cell proliferation and apoptosis. We found that E2F1 was among the target gene set; our analysis also identified an E2F1-DLBCL regulatory program. These results suggested that there may be a two-tiered regulatory mechanism beginning with c-Rel activation of target genes, which include E2F1, and then E2F1 activation of its target genes, many of which have a role in cellular proliferation (**Fig. 2c**).

Another regulatory program (CRP 240) details an abundance of c-Myc–Max binding sites among genes overexpressed in SCLC. This is consistent with the known amplification and overexpression of the Myc family of transcription factors in SCLC[28,29]. Enrichment analysis of this program identified a preponderance of genes involved in DNA metabolism and the cell cycle. Furthermore, we found that this program significantly overlapped with the E2F1-SCLC regulatory program (CRP 96; $P = 0.01$), suggesting that several genes are dually activated by E2F1 and Myc in SCLC. Myc binding sites were also common among genes activated in normal umbilical endothelial cells (CRP 186) as well as in adreocortical carcinoma (CRP 320). The final regulatory program that we explored suggested that ATF activates target genes in salivary carcinoma, of which a disproportionate number are involved in cell migration (CRP 177). ATF1 is activated as a fusion protein in metastatic melanoma[30], but no link between ATF and salivary carcinoma currently exists. If the ATF program is indeed overactivated in salivary carcinoma, then therapies targeting ATF in melanoma[31] may be useful. We observed that many of the transcription factors involved in cancer regulatory programs have oncogenic activity, suggesting that their predicted regulatory function in CRPs may be important in carcinogenesis. Transcription factors identified by our analysis with a causative role in cancer include E2F1 (ref. 32), Myc[33], c-Rel[34], ATF1 (ref. 30) and C-ETS-1 (ref. 35). All regulatory programs and their target genes can be explored through our web-based data-mining platform, Oncomine.

The identification of a CRP implies that a specific transcription factor is active in a specific tissue type and is responsible for the observed gene regulation or deregulation. Activation of a transcription factor can occur either as a downstream effect of a signaling cascade (*e.g.*, phosphoryla-

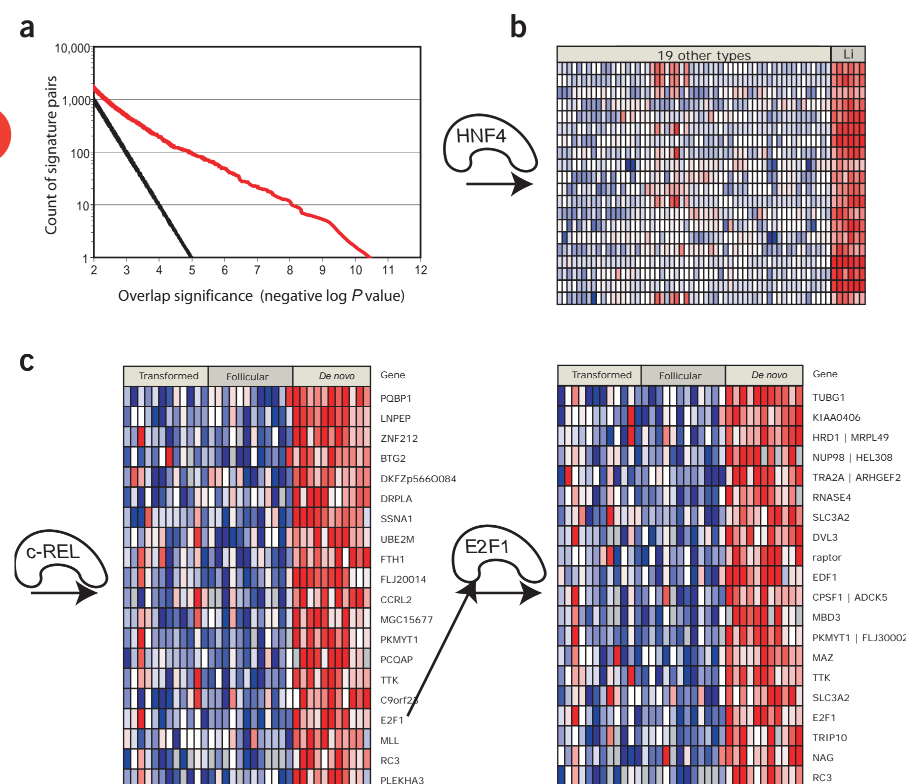


Figure 2 Regulatory programs encoded in gene-expression signatures. (**a**) Regulatory programs were inferred if a gene-expression signature significantly overlapped with a signature of candidate transcription factor targets. The number of significant overlaps observed (red) is compared with the number expected by chance (black). (**b**) A representative normal tissue regulatory program linking the HNF4 transcription factor to 78 target genes (20 shown) exclusively activated in normal liver tissue (Li) relative to several other normal tissues. (**c**) Two regulatory programs activated in *de novo* DLBCL relative to post-transformation DLBCL and follicular lymphoma. The transcription factor of the second program (E2F1) is a target gene in the first program (c-Rel), suggesting a two-tier regulatory mechanism. Red indicates relative overexpression of genes (rows) in the profiled samples (columns); blue indicates relative underexpression.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

tion, nuclear translocation) or by overexpression of the transcription factor itself. To identify those CRPs that may be regulated by the latter mechanism, we searched for concomitant overexpression of the transcription factors that bind to the enriched binding sites. We found overexpression of a respective transcription factor for 79 of the 311 identified CRPs (25%). This analysis may provide a level of validity to some CRPs and, in some cases, may identify the specific transcription factor among a family that is responsible for the target gene regulation. For example, the CREB-ATF binding site is enriched among genes overexpressed in salivary carcinoma (CRP 250). We found that both CREB1 and ATF5 are significantly overexpressed in salivary carcinoma, suggesting that they may be responsible for the CREB-ATF CRP (**Supplementary Fig. 1** online). We also observed concomitant overexpression of the expected transcription factor with several of the normal tissue CRPs, including HNF4A in liver, MEF2A in muscle, EGR3 and EGR4 in brain and multiple IRFs in blood (**Supplementary Table 8** online).

In summary, integrative bioinformatics analyses similar to those carried out by Segal *et al.*[3], our group and others[1,36] will generate new hypotheses about cancer progression. Previously, by carefully maintaining clinical annotations of the specific tissue specimens analyzed, we were able to identify gene alterations that were common to cancer regardless of tissue of origin as well as gene signatures characteristic of more aggressive dedifferentiated cancers[2]. In this report, our integrative approach of analyzing gene-expression signatures in the context of candidate regulatory signatures identified hundreds of normal tissue and cancer CRPs, of which we have highlighted only a few. Several of these regulatory programs link a transcription factor to a tissue type in which the transcription factor is thought to act, whereas several others suggest new regulatory mechanisms in cancer and normal tissue, such as ATF pathway activation in salivary carcinoma. Furthermore, the regulatory programs uncovered by our analysis suggest candidate target genes, such as E2F1 activation by c-Rel in *de novo* DLBCL and chromatin-modifying genes by E2F in high-grade breast cancer. Though powerful, our approach has several limitations: (i) the number of characterized transcription factor binding sites, (ii) the accuracy of the binding sites, (iii) the facts that we only scanned 1-kb promoters and that binding sites are likely to occur outside this region and (iv) the number of genes profiled in the microarray studies and the sensitivity of the various microarray platforms. Despite these limitations, we were able to discern several regulatory mechanisms encoded in gene-expression signatures. We anticipate that our approach will become more valuable as the accuracy and coverage of transcription factor target databases improves.

METHODS

Cancer signatures. We derived cancer signatures from the Oncomine cancer microarray database. We used 65 independent data sets comprising 6,348 samples (arrays) and 70.9 million gene-expression measurements. The samples spanned 26 normal and cancer tissue types. The 65 data sets measured an average of 6,376.5 (range 507–15,294) unique genes as determined by Entrez Gene. We analyzed differential expression with Student's *t*-test and false discovery rates to identify genes with significant differential expression between two classes of samples. We defined gene-expression signatures from analyses that resulted in 20 or more significant genes ($Q < 0.10$, mean difference > 0.5 Z-score units), for a total of 234 gene-expression signatures with an average size of 398 genes (range 20–2,997). Twenty-nine were normal human tissue signatures, and 205 were cancer signatures, of which 68 were derived from comparisons of a cancer type and other cancer types, 50 from comparisons of a cancer type and the respective normal tissue, 22 from comparisons of various molecular subtypes of cancer and 12 from comparisons of histologic subtypes of cancer.

Regulatory signatures. We defined regulatory signatures by scanning human gene promoter sequences for the presence of experimentally defined transcription factor binding sites. We downloaded 1-kb promoter sequences from 20,647 RefSeq reference sequences from the University of California Santa Cruz genome browser (August 2004). These reference sequences mapped to 15,665 unique genes (Entrez Gene). In cases with multiple reference sequences per gene, we analyzed each promoter sequence independently. We submitted sequences sequentially to MATCH, a component of the TRANSFAC Professional Suite, which scans a sequence for the presence of transcription factor binding sites as determined by a database of position weight matrices. We applied the following settings: 'group of matrices' was set to 'vertebrates'; 'use high quality matrices' was selected; 'cut-off selection' was set to 0.8 and 0.85 'as mat. sim and core sim. cutoff.' For each promoter sequence, the program output 'hits' designated by matrix identifier and factor name. For each hit, the position, strand, core match and matrix match were provided. In total, 366 distinct matrices were identified in the promoters of human genes, although many of the matrices represent variants of the same transcription factor binding site. With the aforementioned settings, 16,159,457 million hits were identified. Because in some cases, our lenient match threshold identified hits in nearly every promoter sequence, we filtered the hit list to contain only the top 2,000 hits per matrix sorted by the matrix similarity score. Five matrices with greater than 2,000 perfect matches (score = 1.0) were removed from the analysis. To ensure that our results would be robust to the selected hit threshold, the analysis was rerun with 1,500 and 2,500 hit thresholds. As expected, we obtained largely overlapping results (data not shown). After mapping reference sequences to Entrez Gene, we defined 466,491 potential regulatory interactions. Transcription factor matrices had an average of 1,292.2 potential gene targets (range 4–1,554).

Enrichment analysis. We assessed each gene-expression signature (S_G) for the significant enrichment of each regulatory signature (S_R). The possible set for each gene-expression signature (P_G) was defined as the set of measured genes in each respective data set. The possible set for regulatory signatures (P_R) was defined as the set of genes with available promoter sequences. We counted the number of genes intersecting a gene-expression signature and a regulatory signature: $n = c(S_G \cap S_R)$, where $c(A)$ denotes the number of elements in set A. We counted the number of genes in both the regulatory signature and the possible set for the gene-expression signature: $N = c(S_R \cap P_G)$. Next, we computed the background probability of observing a gene in a gene-expression signature by dividing the number of genes in both the gene-expression signature and the possible set for the regulatory signature by the number of genes in both possible sets:

$$p = \frac{c(S_G \cap P_R)}{c(P_G \cap P_R)}$$

Finally, we calculated the probability of observing an equal or larger intersection between the gene-expression signature and regulatory signature by chance by summing the binomial distribution probabilities for all intersections of equal or larger size:

$$P\text{ value} = \sum_{i=n}^{N} \left(\frac{N!}{(N-i)!} \right) p^i (1-p)^{N-i}$$

We applied the method of false discovery rates to adjust P values for multiple hypothesis testing. We calculated Q values as:

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

$$\text{Q value} = \frac{(N * P\ \text{value})}{R}$$

where N is the number of regulatory signatures tested against each gene-expression signature and R is the ascending order rank of the respective P value. We also calculated global Q values where N is the total number of hypotheses tested (all gene-expression signatures by all regulatory signatures). We used the global Q value to assess the significance of our entire study and used the signature-specific Q values to interpret the significance of the observed enrichments per gene-expression signature.

***In vitro* E2F analysis.** We collected target genes for E2F1, E2F2 and E2F3 from an *in vitro* E2F profiling study[19]. We created a composite signature of 588 target genes by combining all genes that were induced by any one of the E2F family members. We selected ten CRPs that corresponded to a single representative E2F binding site (V$E2F_Q4_01) for enrichment analysis. We carried out enrichment analysis by the binomial distribution exactly as described in the preceding section.

URLs. The Oncomine database is available at http://www.oncomine.org/. Promoter sequences from the University of California Santa Cruz genome browser are available at http://hgdownload.cse.ucsc.edu/goldenPath/hg17/bigZips/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
We thank D. Gibbs for hardware support and R. Varambally for database support. This research is supported in part by the National Institutes of Health through the University of Michigan's Cancer Center Support Grant, pilot funds from the Dean's Office and the Department of Pathology. D.R.R. was supported by the Medical Scientist Training Program and the Cancer Biology Training Program, and A.M.C. is a Pew Scholar.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

1. Ramaswamy, S., Ross, K.N., Lander, E.S. & Golub, T.R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).
2. 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).
3. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* **36**, 1090–1098 (2004).
4. Elkon, R., Linhart, C., Sharan, R., Shamir, R. & Shiloh, Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res.* **13**, 773–780 (2003).
5. Rhodes, D.R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
6. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
7. Sladek, F.M., Zhong, W.M., Lai, E. & Darnell, J.E. Jr. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev.* **4**, 2353–2365 (1990).
8. Xanthopoulos, K.G. *et al.* The different tissue transcription patterns of gene for HNF-1, C/EBP, HNF-3, and HNF-4, protein factors that govern liver-specific transcription. *Proc. Natl. Acad. Sci. USA* **88**, 3807–3811 (1991).
9. Black, B.L. & Olson, E.N. Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu. Rev. Cell Dev. Biol.* **14**, 167–196 (1998).
10. O'Donovan, K.J., Tourtellotte, W.G., Millbrandt, J. & Baraban, J.M. The EGR family of transcription-regulatory factors: progress at the interface of molecular and systems neuroscience. *Trends Neurosci.* **22**, 167–173 (1999).
11. Taniguchi, T., Ogasawara, K., Takaoka, A. & Tanaka, N. IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.* **19**, 623–655 (2001).
12. Caamano, J. & Hunter, C.A. NF-kappaB family of transcription factors: central regulators of innate and adaptive immune functions. *Clin. Microbiol. Rev.* **15**, 414–429 (2002).
13. Rosner, M.H. *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686–692 (1990).
14. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
15. La Thangue, N.B. The yin and yang of E2F-1: balancing life and death. *Nat. Cell Biol.* **5**, 587–589 (2003).
16. Zhu, W., Giangrande, P.H. & Nevins, J.R. E2Fs link the control of G1/S and G2/M transcription. *EMBO J.* **23**, 4615–4626 (2004).
17. Bracken, A.P. *et al.* EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**, 5323–5335 (2003).
18. Chae, H.D., Yun, J., Bang, Y.J. & Shin, D.Y. Cdk2-dependent phosphorylation of the NF-Y transcription factor is essential for the expression of the cell cycle-regulatory genes and cell cycle G1/S and G2/M transitions. *Oncogene* **23**, 4084–4088 (2004).
19. Muller, H. *et al.* E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev.* **15**, 267–285 (2001).
20. DeGregori, J., Kowalik, T. & Nevins, J.R. Cellular targets for activation by the E2F1 transcription factor include DNA synthesis- and G1/S-regulatory genes. *Mol. Cell Biol.* **15**, 4215–4524 (1995).
21. Keenan, S.M., Lents, N.H. & Baldassare, J.J. Expression of cyclin E renders cyclin D-CDK4 dispensable for inactivation of the retinoblastoma tumor suppressor protein, activation of E2F, and G1-S phase progression. *J. Biol. Chem.* **279**, 5387–5396 (2004).
22. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
23. Kleer, C.G. *et al.* EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc. Natl. Acad. Sci. USA* **100**, 11606–11611 (2003).
24. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624–629 (2002).
25. Gilmore, T.D., Kalaitzidis, D., Liang, M.C. & Starczynowski, D.T. The c-Rel transcription factor and B-cell proliferation: a deal with the devil. *Oncogene* **23**, 2275–2286 (2004).
26. Houldsworth, J. *et al.* Relationship between REL amplification, REL function, and clinical and biologic features in diffuse large B-cell lymphomas. *Blood* **103**, 1862–1868 (2004).
27. Lossos, I.S. *et al.* Transformation of follicular lymphoma to diffuse large-cell lymphoma: alternative patterns with increased or decreased expression of c-myc and its regulated genes. *Proc. Natl. Acad. Sci. USA* **99**, 8886–8891 (2002).
28. Nau, M.M. *et al.* L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* **318**, 69–73 (1985).
29. Wong, A.J. *et al.* Gene amplification of c-myc and N-myc in small cell carcinoma of the lung. *Science* **233**, 461–464 (1986).
30. Zucman, J. *et al.* EWS and ATF-1 gene fusion induced by t(12;22) translocation in malignant melanoma of soft parts. *Nat. Genet.* **4**, 341–345 (1993).
31. Jean, D. & Bar-Eli, M. Targeting the ATF-1/CREB transcription factors by single chain Fv fragment in human melanoma: potential modality for cancer therapy. *Crit. Rev. Immunol.* **21**, 275–286 (2001).
32. Johnson, D.G., Cress, W.D., Jakoi, L. & Nevins, J.R. Oncogenic capacity of the E2F1 gene. *Proc. Natl. Acad. Sci. USA* **91**, 12823–12827 (1994).
33. Schwab, M., Varmus, H.E. & Bishop, J.M. Human N-myc gene contributes to neoplastic transformation of mammalian cells in culture. *Nature* **316**, 160–162 (1985).
34. Sylla, B.S. & Temin, H.M. Activation of oncogenicity of the c-rel proto-oncogene. *Mol. Cell Biol.* **6**, 4709–4716 (1986).
35. Seth, A. & Papas, T.S. The c-ets-1 proto-oncogene has oncogenic activity and is positively autoregulated. *Oncogene* **5**, 1761–1767 (1990).
36. Lamb, J. *et al.* A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* **114**, 323–334 (2003).