nature biotechnology

Decoding global gene expression programs in liver cancer by noninvasive imaging

Eran Segal[1], Claude B Sirlin[2], Clara Ooi[4], Adam S Adler[5], Jeremy Gollub[6], Xin Chen[8], Bryan K Chan[2], George R Matcuk[7], Christopher T Barry[3], Howard Y Chang[5] & Michael D Kuo[2]

Paralleling the diversity of genetic and protein activities, pathologic human tissues also exhibit diverse radiographic features. Here we show that dynamic imaging traits in noninvasive computed tomography (CT) systematically correlate with the global gene expression programs of primary human liver cancer. Combinations of twenty-eight imaging traits can reconstruct 78% of the global gene expression profiles, revealing cell proliferation, liver synthetic function, and patient prognosis. Thus, genomic activity of human liver cancers can be decoded by noninvasive imaging, thereby enabling noninvasive, serial and frequent molecular profiling for personalized medicine.

Since Roentgen produced the first radiograph in 1895 (ref. 1), scientists and clinicians have used noninvasive imaging to study the physical and structural composition of living matter. Assessing the genetic and biochemical makeup of living tissue through noninvasive imaging is a key goal of current research. Recent development of genomic and proteomic methods have enabled molecular profiling of biological specimens by simultaneously revealing the expression level of thousands of genes and proteins. Gene expression patterns of cancer can reveal its etiology, prognosis and response to therapy[2–4]. However, current methods of molecular profiling require invasive surgeries for tissue procurement and specialized equipment, thus limiting their routine use. Moreover, current profiling methods provide only single snapshots in time because they are destructive; cells must be disintegrated to extract nucleic acids or proteins for analysis. Human tissues also exhibit diverse distinctive traits on noninvasive radiographic images, many of which currently have no known significance (**Fig. 1a**). Because imaging traits of tissues reflect the dynamic and physiologic interplay of parenchymal cells, blood vessels and stroma, we reasoned that imaging traits may be used to predict gene expression patterns in human cancers.

We hypothesized that the global gene expression patterns of human cancers may systematically correlate with their dynamic imaging features. To relate gene expression to imaging, we needed to first define 'units of distinctiveness', termed 'traits', from qualitative imaging features, and likewise define coherent patterns of variation from gene expression profiles. Second, imaging traits are likely to correlate with gene expression patterns in a complex manner, and methods of relating imaging to gene expression need to account for combinatorial and conditional logical relationships such as AND and OR.

To address these challenges of relating gene expression to imaging, we followed a three-step strategy to create an 'association map' between imaging features on three-phase contrast enhanced CT scans and gene expression patterns of 28 human hepatocellular carcinomas (HCC)[4] (**Fig. 1b**). First, we defined and quantified 138 distinctive imaging traits present in one or more HCCs. To identify the most informative traits, we filtered traits based on their frequency and prominence in the data, interobserver agreement between two radiologists, and independence from other traits as determined by Pearson correlation among the traits ($r = 0.9$). Thirty-two imaging traits were extracted by these criteria and used for subsequent analysis (**Supplementary Table 1** online). For instance, many tumors displayed channels of radio-dense signal within certain tumors on the arterial phase of the CT scan[5], and this trait was termed 'internal arteries' (**Fig. 1a**). Second, we adapted the module networks algorithm[6] to systematically search for associations between expression levels of 6,732 well-measured genes determined by microarray analysis[4] and combinations of imaging traits. The algorithm identifies groups of genes, termed modules, with coherent variation in expression across multiple samples. The algorithm also identifies combinations of imaging traits that can explain the expression levels of gene modules, where each combination explains the expression level of one gene module. Third, we validated the statistical significance of the association map by comparison with permuted data sets, and by testing the prediction of the association map in an independent set of tumors.

The association map of imaging traits and gene expression revealed that a large fraction of the gene expression program can be reconstructed from a small number of imaging traits (**Fig. 2a,b**). The expression variation in 6,732 genes was captured by 116 gene modules,

[1]Dept. of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Dept. of Radiology and [3]Dept. of Surgery, University of California at San Diego, 200 W. Arbor Dr., San Diego, California 92103, USA. [4]Dept. of Radiology, University of Hong Kong, 21 Sasson Road, Pokfulam, Hong Kong DAR, China. [5]Program in Epithelial Biology, 269 Campus Dr., Stanford, CCSR2155c, California 94305, USA. [6]Dept. of Biochemistry, 279 Campus Dr., Stanford, California 94305, USA. [7]Dept. of Radiology, University of Southern California, 1200 N. State St., Los Angeles, California 90033, USA. [8]Dept. of Biopharmaceutical Science, University of California at San Francisco, 513 Parnassus Ave., San Francisco, California 94143, USA. Correspondence should be addressed to M.D.K. (mkuo@ucsd.edu) or H.Y.C. (howchang@stanford.edu).

Received 8 January; accepted 24 April; published online 21 May 2007; doi:10.1038/nbt1306

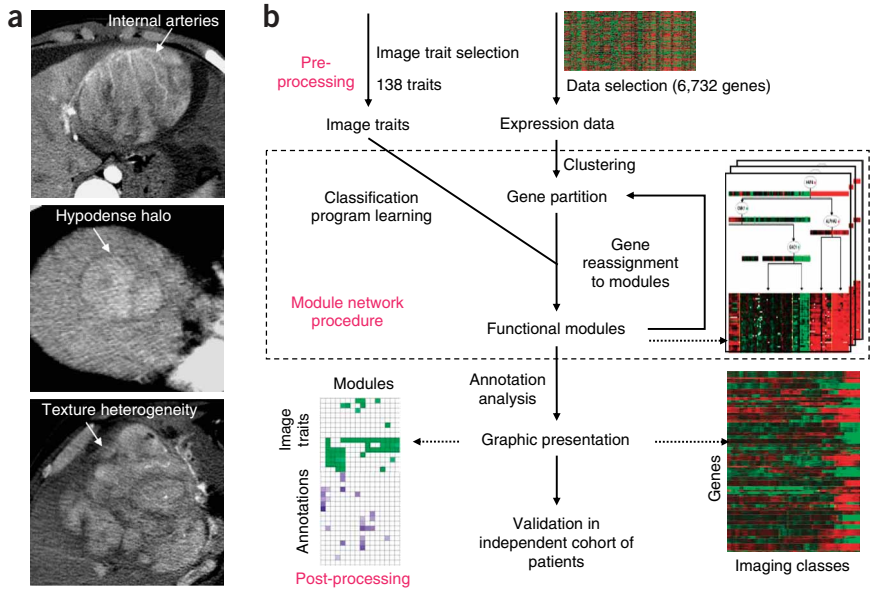

Figure 1 Linking imaging traits and global gene expression. (**a**) Examples of imaging traits in HCC. (**b**) Strategy for constructing an association map between imaging traits and gene expression.

each of which was associated with specific combinations of imaging traits. For each module, the presence or absence of combinations of imaging traits explained the aggregate expression level of genes within the module (**Fig. 2a**). The combinations of relevant imaging traits are depicted in decision trees: each split in the tree is specified by variation of an imaging trait; each terminal leaf in the tree is a cluster of samples that share a similar expression pattern of module genes. Thus, the association map allowed us to reconstruct the relative expression level of a gene (by mapping it to a module) in a given HCC sample (by mapping it to a cluster). Across all 116 gene modules capturing 6,732 genes in our model, the difference in the level of expression of member genes from their cognate module averages is 1.36- ± 1.33-fold (mean ± s.d.). Thus, the expression level of individual genes can be reconstructed from imaging features with an average deviation of about twofold, within the experimental determination level allowed by microarray analysis.

Notably, the combinations of only 28 imaging traits were sufficient to reconstruct the variation of all 116 gene modules. Only nine traits were sufficient to reconstruct the expression patterns of 50% of the transcriptome, and the model plateaus to >80% of the transcriptome with >23 traits (**Fig. 2b**). For each gene, the number of traits needed to predict its variation is on average three and no more than four in any instance. The association of imaging traits and gene expression was highly significant by several independent statistical criteria. Specification of the entire module network involved 355 splits based on imaging traits. The average gene expression levels between two sides of each split was significantly different in 299 of 355 splits ($P < 0.05$ after a Bonferroni correction for multiple hypotheses), accounting for 5,282 of 6,732 input genes (78.5%). Comparison of the observed association map of imaging traits and gene expression with maps derived from data sets with permuted sample labels confirmed that it was highly unlikely that the predictive power of imaging traits for expression patterns was due to chance alone. The log-likelihood was –18 per microarray, compared to only –23 ± 0.1 expected by chance (ten permutations; $P < 10^{-50}$). Thus, the variation in gene expression was densely correlated with a small number of imaging traits. Once identified, such 'coding' of imaging traits can be used to translate visual images into global gene expression programs.

To further validate the association map of imaging traits and global gene expression, we tested the predictive power of the map in an independent group of 19 prospectively collected patients with HCC. We predicted the global gene expression pattern of these 19 tumors based on the patients' preoperative CT scans, and then performed microarray analysis on their resected tumors to determine whether our predicted gene expression profiles were correct. We found that 71 of the 116 gene modules, comprising 4,996 out of 6,732 (74%) genes of the transcriptome under consideration, were significantly predicted by their cognate imaging traits, that is, the combinations of imaging traits predicted for each module split the module-genes into distinct patterns of expression ($P < 0.01$, Student's *t*-test; **Supplementary Table 2** online). A separate permutation test also confirmed the strong link between imaging traits and global gene expression. These results provide additional support that the association map of imaging traits and gene modules is robust and can be used to predict the global expression profiles of a large fraction of the transcriptome in independent sets of patients.

Imaging traits predictive of the expression level of specific genes are directly revealed using the association map, and the potential physiologic significance of many imaging traits can be inferred from their associated genes. We found that the distribution of genes into modules defined by imaging traits was not random, but was highly enriched for specific and diverse biological functions and processes. Comparison of gene membership in modules versus the published Gene Ontology annotation[7] revealed significant overlaps ($P < 0.05$, $FDR < 0.05$) (**Fig. 2c**), allowing many key physiologic properties of tumors to be gleaned from CT images. For example, three image traits reconstructed the expression level of module 697, which is highly enriched for genes involved in cell proliferation, including *PCNA*, *CCNA2*, *MCM5*, *MCM6* and *GMNN* (**Fig. 3a**). In addition, the expression level of *VEGF*, an important driver of tumor angiogenesis and target of the approved antibody therapeutic bevacizumab[8] (Avastin), covaries with these cell cycle genes and is predicted by the same imaging traits (**Fig. 3a**). This association immediately suggests a method for non-invasively delineating a molecularly distinct subset of tumors for a targeted therapeutic strategy. The expression of liver synthetic enzymes in HCC reflects the extent of tumor dedifferentiation[9]; this information is evident in module 595, which details the expression level of *ALB*, *PKLR*, *TFR2*, as well as revealing clotting function (*F2*, *F5*, *F10*), and detoxification activity (*GSTO1*, *CYP27A1*, *EPHX2*) (**Fig. 3b**). Conversely, identity of genes in a module can reveal the physiologic basis of an imaging trait. The imaging trait 'tumor margin score, minimum' denotes tumors that show an ill-defined transition zone between tumor and surrounding liver tissue; we found that the presence of this trait was associated with elevated expression of a group of genes associated with extracellular matrix remodeling, such as *MMP2*, *MMP7*, *COL3A1*, *COL6A2*, and *THBS1* and *THBS2* (**Fig. 3c**). Several of these genes, notably *MMP2* (refs. 10,11) and *THBS1* and *THBS2* (refs. 1,12), are known to increase tumor

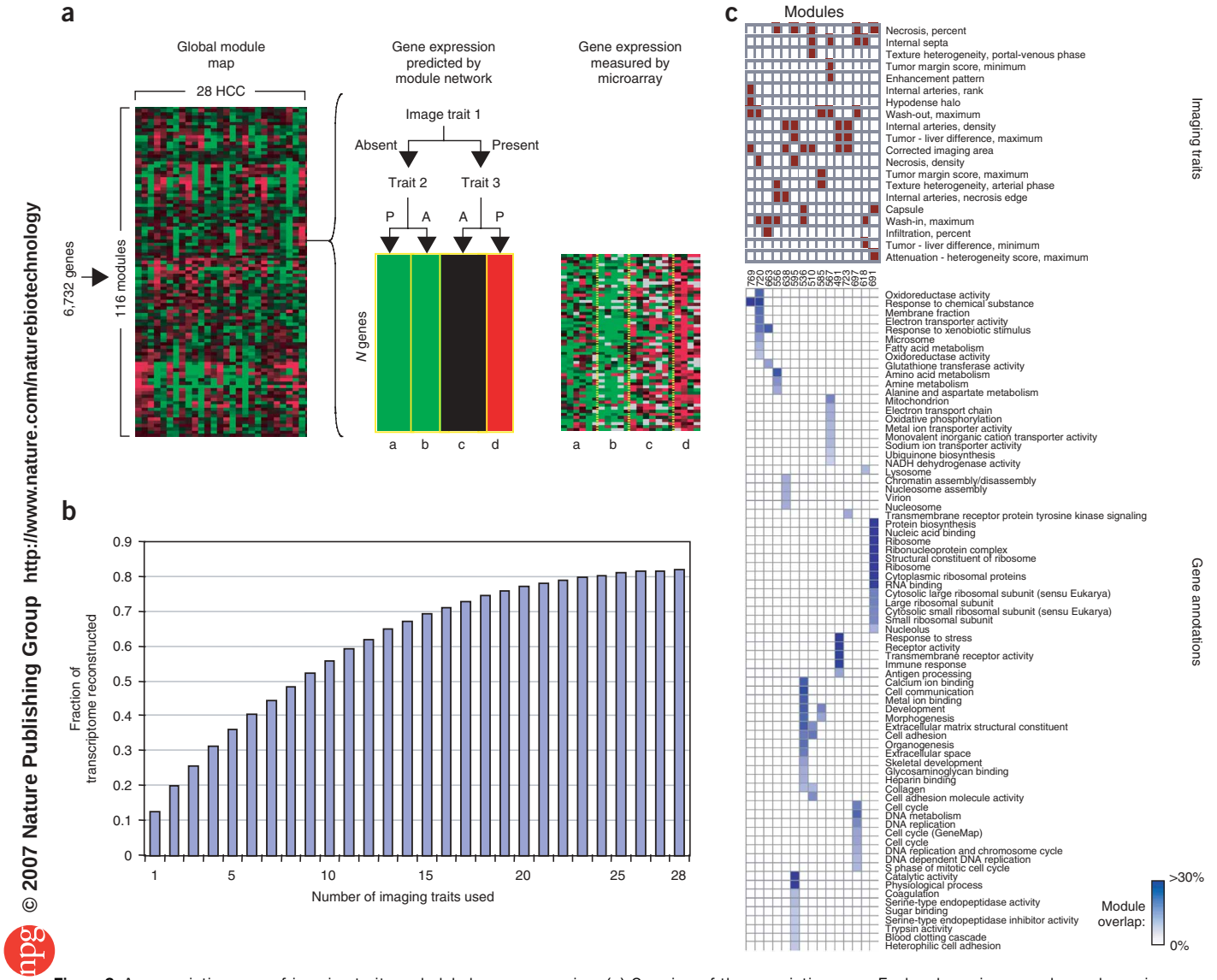

Figure 2 An association map of imaging traits and global gene expression. (**a**) Overview of the association map. Each column is a sample; each row is a module. For each module, a decision tree of imaging traits is associated with variation in the expression level of module genes. Knowledge of the imaging traits thus allows an approximate reconstruction of the gene expression pattern. (**b**) Dense encoding of gene expression variation by imaging traits. Shown is the cumulative fraction of gene expression variation across the transcriptome that is reconstructed by the number of imaging traits in the model. (**c**) Matrix of modules, associated imaging traits, and their enriched GO annotations. Only modules and annotations with significant enrichment (false discovery rate <0.05 after accounting for multiple hypothesis testing) are shown.

invasiveness into surrounding stroma, which may lead to the poor demarcation of tumor margins on CT images.

The association map also enabled systematic mapping of a predetermined group of genes to their corresponding imaging traits. Previously, a group[4] working with the same tumors imaged in this study identified expression variation in a group of 91 genes that was associated with microscopic venous invasion, a well-established sign of poor prognosis[9] that is extremely difficult to predict using conventional imaging methods in the absence of gross venous invasion. The 'venous invasion signature' comprises genes involved in cell proliferation (*CDK4*, *CDC20*, *MCM5*) and matrix invasion (*ADAMTS1*, *MMP14*, *SPARCL1*)—genes known to be induced in other HCC data sets and indeed in many types of cancers[13]. We found that the 91 genes in the 'venous invasion signature' were enriched in seven modules and associated with two predominant imaging traits—the presence of 'internal arteries' and absence of 'hypodense halos' (**Fig. 4a** and **Supplementary Table 3** online). Therefore, we tested whether this pair of imaging traits, as observed during the preoperative CT scan, predicted the occurrence of microscopic venous invasion by histologic analysis. In 30 patients with HCC, tumors with this combination of imaging traits had a 12-fold increased risk of microscopic venous invasion ($P = 0.004$). Importantly, the predictive value of the two-trait predictor of venous invasion was validated in an independent set of 32 patients that were not used for training the association map (**Fig. 4a**, $P = 0.03$), despite the fact that CT imaging, patient characteristics and even tumor sizes were not exactly matched with those of the training set. Moreover, the presence of the trait 'internal arteries' in the preoperative CT scan of HCCs was a significant

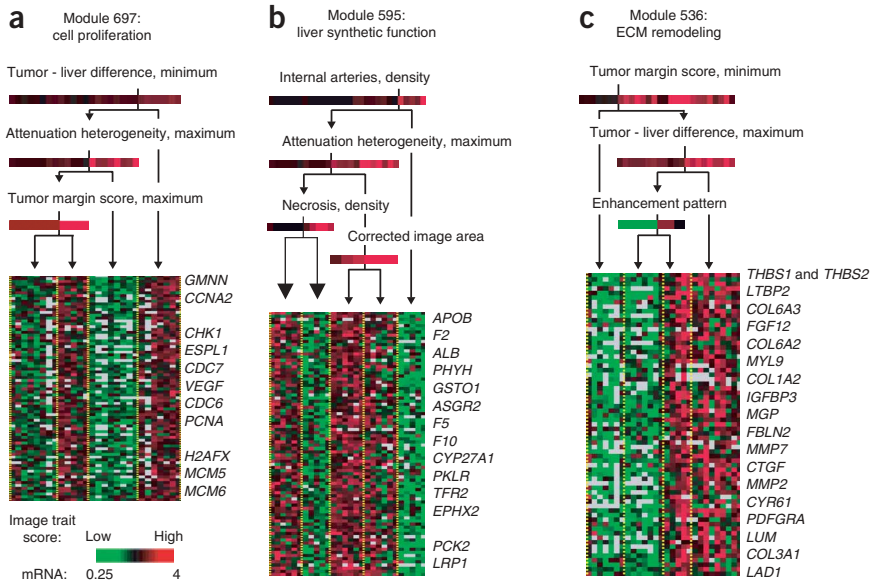

Figure 3 Molecular portraits of HCC from imaging traits. Modules associated with: (**a**) HCC proliferation; (**b**) liver synthetic function; and (**c**) extracellular matrix remodeling. Each column is tumor sample; each row is a gene. Imaging traits specifying each module are outlined on top; expression pattern of genes within the module as distinguished by imaging traits are shown on bottom.

($P < 0.05$) univariate predictor of overall survival in both groups of patients (**Fig. 4b**). Thus, the association map can identify novel imaging traits corresponding to gene expression signatures and provide useful information to guide clinical decision making.

In summary, we show that much of the global gene expression profiles of liver cancer were systematically correlated with their imaging features. The systematic association between imaging traits and gene expression allowed useful inference from both directions. On the one hand, the association map identified biological processes, based on specific gene expression programs, which underlie specific imaging traits. On the other hand, the association map enabled the use of imaging traits to reconstruct the global gene expression programs of cancer, thereby creating a noninvasive molecular portrait of the tumor (**Fig. 3**). We illustrate the utility of this approach by identifying and validating a two-trait predictor of venous invasion in HCC (**Fig. 4**). Moreover, the 'internal arteries' trait that emerged from this analysis was a significant predictor of survival in two independent groups of patients. These results demonstrate that existing imaging technology may be used to reconstruct the molecular anatomy of human liver cancer and potentially other diseases in a noninvasive fashion. Although our study involved two small sets of liver cancers, our success with these limited samples illustrates the robustness of the method.

Our algorithm for associating imaging features and gene expression is generalizable, and in principle may be applied to any disease state and imaging modality (such as positron emission tomography (PET), magnetic resonance imaging (MRI), or other imaging methods). However, an important caveat is that HCCs are generally large and well-perfused tumors, which permit detailed radiographic examination using computed tomography, a modality widely used in daily clinical practice. For smaller tumors or lesions of other disease states, CT scans may not be suitable and other imaging modalities, such as MRI or PET, may need to be employed. In the current study, we analyzed imaging traits as features over the entire tumor and systematically correlated them with mRNA levels extracted from resected whole tumors. Our method offers the possibility to identify imaging features that predict region-specific gene expression signatures within the tumor in the proper anatomic context of the patient. This development will require potentially higher resolution imaging modalities, targeted tissue specimens from radiographically diverse regions of the same or different tumors, and a larger number of samples to account for an additional level of multiple hypothesis testing.

We and others have shown that gene expression signatures, each comprised of dozens to hundreds of genes, can significantly improve the diagnostic classification, prognostication and prediction of therapeutic response in cancer[2–4,14]. For instance, in breast cancer, a 21-gene assay, termed Oncotype Dx, is commercially available to determine prognosis and predict response of primary breast tumors to

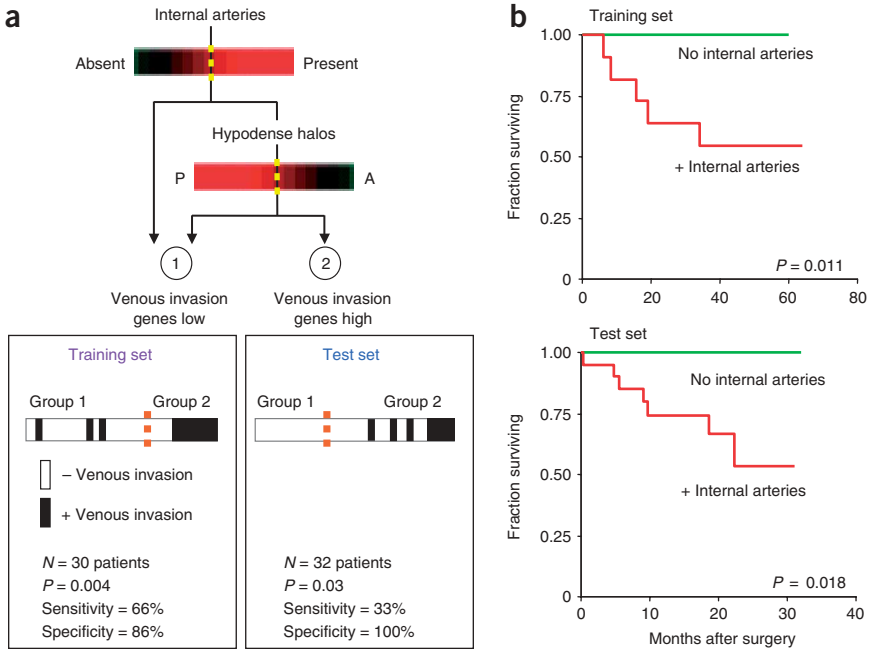

Figure 4 Imaging traits predict venous invasion and survival. (**a**) A two-trait decision tree associated with a gene expression signature of venous invasion predicts histologic venous invasion. (**b**) Kaplan-Meier survival curves of HCC patients with and without "internal arteries" imaging trait are shown.

chemotherapy[15,16]. Similarly, a 70-gene signature[17,18] is undergoing a multicenter clinical trial in Europe as the basis of undergoing adjuvant chemotherapy in primary breast cancer and is also commercially sold as Mammaprint. Gene expression signatures that can predict prognosis or therapeutic response in lung cancer[19], leukemias[20] and prostate cancers[21] have also been identified. Microarray analysis is also standard practice in many ongoing clinical trials in order to discover potential pharmacogenomic links between drugs and patient response. These activities suggest that gene expression profiling will become important for daily clinical practice in the near future. Our method potentially provides gene expression profiling that is noninvasive (and therefore with no risk of pain, infection or complications that accompany tissue procurement for microarray analysis), fast, repeatable and in the native anatomic context of the patient. Just as radiographic exams (CT, MRI, angiography) have replaced physical diagnosis in many clinical settings as the gold standard (e.g., the diagnosis of HCC in patients with a history of hepatitis virus infection), gene expression profiling by noninvasive imaging may supplement histologic examination for cancer diagnosis. In the future, canonical association maps constructed from large representative series of tumors may enable routine noninvasive diagnosis of genetically heterogeneous tumors, reveal their prognosis and allow serial profiling of tumors during therapy. Finally, this type of imaging-based molecular profiling could speed up the development of personalized medicine.

METHODS

Imaging traits. We defined 138 distinct imaging traits that were present in at least one tumor sample and were scored across all tumor samples. Traits were selected a priori based on intrinsic radiological interest (e.g., internal arteries and hypodense halos). Traits were also filtered based on their frequency and prominence in the data, interobserver agreement and independence from other traits based on Pearson correlation (cut-off value of 0.9). We used 32 imaging traits as input in the Bayesian model, and 28 of 32 were found to be informative of gene expression (**Supplementary Table 1**).

Microarray data. Training set: gene expression profiles[4] of 28 imaged HCCs were downloaded from Stanford Microarray Database (http://genome-www5.stanford.edu/). Data from array elements that had hybridization signals 1.5 times the background signal in both Cy5 and Cy3 channels and present in 70% of samples were centered by means across samples. Data from replicate probes representing the same gene (as determined by Locus Link ID) were averaged. We used the 6,732 genes that met these criteria for data quality for subsequent analysis. Validation set: 19 independent HCCs were imaged and subsequently excised and extracted for total RNA. Tumor RNA was amplified, labeled with Cy5, and competitively hybridized with Cy3-labeled reference RNA to Stanford cDNA microarrays, as previously described[4].

Module network. We adapted and applied the module network procedure that we previously developed[6] to construct an association map between imaging traits and gene expression profiles. The module network procedure takes as input gene expression data and a set of potential regulatory input, and attempts to partition the expression data into distinct and mutually exclusive modules, such that the gene assigned to each module can be well predicted by a small decision tree of regulatory inputs. In our application, we set the regulatory inputs to be the real-valued imaging traits and applied them to the expression data described above.

Validation of module network in independent set of patients. We predicted the global gene expression program of 19 prospectively collected, independent HCC samples based on their preoperative three-phase CT images. By comparing the global gene expression program predicted by CT images according to our previous model to the actual microarray profiles that we measured, we found a 74% validation rate, which is notable considering that the patients and microarray platforms of the training and validation sets were not the same. Specifically, in our training set, we identified 28 imaging traits that predicted the variation of expression in 5,282 of 6,732 (78%) genes in the transcriptome. We took the imaging trait corresponding to the top split in each gene module, and tested whether variation in that imaging trait also predicted variation of expression in the corresponding gene module in the new validation set of tumors. We found that 71 of the 116 gene modules, comprising 4,996 out of 6,732 (74%) genes of the transcriptome under consideration, were significantly predicted by their cognate imaging traits ($P < 0.01$, Student's *t*-test). The gene modules, their cognate imaging traits, and the *P*-values of the difference in gene expression as dictated by the variance in imaging are listed in **Supplementary Table 2**.

In addition, we performed a permutation analysis to more rigorously test the association between imaging traits and gene expression in the validation set. We permuted the values of the imaging trait across the new 19 arrays 100 times for each module, and compared the *P*-value of the gene module variation dictated by the real imaging trait to that obtained in the 100 permutations. Note that the 19 samples in the validation set have substantially less power than our training set, and are thus only able to validate the strongest relationships identified in the original training set of 28 samples. Nevertheless, the imaging trait–gene module pairs identified by our model still have outstanding *P*-values for many gene modules. 1,567 out of 6,732 transcribed genes (23%) were predicted by imaging traits that beat more than 95% of the permuted data ($P < 0.05$). These results provide further confidence that the association map of imaging traits and gene modules is highly robust and can be used to predict the expression level of thousands of genes in independent sets of samples.

Module enrichment in functional annotations. Significance of overlap between genes in modules and annotations in GO, GeneMap and KEGG are calculated by comparison to the degree of overlap expected by chance alone using the hypergeometric distribution. We account for multiple hypothesis testing by calculating a false discovery rate and present results with $FDR < 0.05$.

Mapping venous invasion genes to imaging traits. To find imaging traits that correspond to the set of 91 genes associated with venous invasion[4], we identified seven modules that are significantly enriched for these gene using the hypergeometric distribution as described above. The associated imaging trait trees of the seven modules were analyzed (**Supplementary Table 3**), and two traits, internal arteries and halos, were found to be overrepresented among the top splits. To identify the consensus threshold of applying these traits for this purpose, we calculated the *P*-value weighted average of the splits from the seven image trait trees. The consensus thresholds were used for the imaging trait decision tree of **Figure 4a**.

Clinical data analysis. Microscopic venous invasion status on histologic analysis was available for 30 patients in the training set and 32 patients in the test set. Within each data set, patients were partitioned into two groups based on the two trait decision trees ('internal arteries' and 'hypodense halos' on the CT scan, **Fig. 4a**). Significance of association between the two trait imaging groups and histologic venous invasion was calculated using two-by-two contingency tables and a chi-squared test. Overall survival data were available for 23 patients in the training set and 32 patients in the test set; only patients with tumor-free margins after HCC resection were used in this analysis. Within each data set, patients were partitioned based on the presence or absence of the 'internal arteries' trait on the CT scan, and survival analysis by the method of Kaplan and Meier for the two groups of patients was implemented in Winstat (R. Fitch software).

URLs. The 116 networks linking imaging traits to specific gene modules can be interactively searched at http://genie.weizmann.ac.il/pubs/imaging06/. Microarray data of the newly generated validation set of HCCs can be downloaded at Stanford Microarray Database (http://genome-www5.stanford.edu/) and GEO (http://www.ncbi.nlm.nih.gov/geo/).

Requests for materials. M.D.K. (mkuo@ucsd.edu) and H.Y.C. (howchang@stanford.edu).

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

ACKNOWLEDGMENTS

Supported by grants from UC San Diego (M.D.K.), the Israel Science Foundation (E.S.), and the National Institutes of Health (X.C., H.Y.C.). E.S. is the incumbent of the Soretta and Henry Shapiro career development chair. H.Y.C. is the Kenneth G. and Elaine A. Langone Scholar of the Damon Runyon Cancer Research Foundation. M.D.K. is the Bracco Diagnostics Research Scholar of the Radiological Society of North America.

AUTHOR CONTRIBUTIONS

E.S., H.Y.C. and M.D.K. conceived of the project, analyzed the data and wrote the paper. All other authors provided data or analysis tools.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology/.

Published online at http://www.nature.com/naturebiotechnology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Eisenberg, R.L. *Radiology: An Illustrated History* (Mosby Year Book, St. Louis, 1991).
2. Chung, C.H., Bernard, P.S. & Perou, C.M. Molecular portraits and the family tree of cancer. *Nat. Genet.* **32** Suppl, 533–540 (2002).
3. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to models: understanding cancer using microarrays. *Nat. Genet.* **37** Suppl, S38–S45 (2005).
4. Chen, X. *et al.* Gene expression patterns in human liver cancers. *Mol. Biol. Cell* **13**, 1929–1939 (2002).
5. Nino-Murcia, M. *et al.* Focal liver lesions: pattern-based classification scheme for enhancement at arterial phase CT. *Radiology* **215**, 746–751 (2000).
6. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat. Genet.* **34**, 166–176 (2003).
7. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
8. Kerr, D.J. Targeting angiogenesis in cancer: clinical development of bevacizumab. *Nat. Clin. Pract. Oncol.* **1**, 39–43 (2004).
9. Thomas, M.B. & Abbruzzese, J.L. Opportunities for targeted therapies in hepatocellular carcinoma. *J. Clin. Oncol.* **23**, 8093–8108 (2005).
10. Giannelli, G. *et al.* Clinical role of MMP-2/TIMP-2 imbalance in hepatocellular carcinoma. *Int. J. Cancer* **97**, 425–431 (2002).
11. Qin, L.X. & Tang, Z.Y. The prognostic molecular markers in hepatocellular carcinoma. *World J. Gastroenterol.* **8**, 385–392 (2002).
12. Poon, R.T. *et al.* Clinical significance of thrombospondin 1 expression in hepatocellular carcinoma. *Clin. Cancer Res.* **10**, 4150–4157 (2004).
13. Whitfield, M.L., George, L.K., Grant, G.D. & Perou, C.M. Common markers of proliferation. *Nat. Rev. Cancer* **6**, 99–106 (2006).
14. Chang, H.Y. *et al.* Robustness, scalability, and integration of a wound response gene expression signature in predicting survival of human breast cancer patients. *Proc. Natl. Acad. Sci. USA* **102**, 3738–3743 (2005).
15. 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).
16. Paik, S. *et al.* Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.* **24**, 3726–3734 (2006).
17. van 't Veer, L.J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
18. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
19. Potti, A. *et al.* A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N. Engl. J. Med.* **355**, 570–580 (2006).
20. Bullinger, L. *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N. Engl. J. Med.* **350**, 1605–1616 (2004).
21. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. USA* **101**, 811–816 (2004).