M:CC Bioinformática 2010/2011

Decoding global gene expression programs in liver cancer by noninvasive imaging

Summary

- Intro
- Gene expression
- Bayesian Networks
- Module Networks
- Decoding global gene expression programs in liver cancer by noninvasive imaging
- Conclusion

Intro

- Nowadays cancer is a very serious health problem that affects a great percentage of people around the world.
- The causes of cancer can be divide into two groups: the environmental cause group and the hereditary genetic cause group.
- Liver cancer is known as a cause of death in at least 30 cases per 100,000 inhabitants in most of the world, with higher rates observed in parts of Africa and eastern Asia.

Intro

- In 2007 cancer caused about 13% of all human deaths worldwide (7.9 million). Rates are rising as more people live to an old age and lifestyles change in the developing world.
- Liver cancer or hepatic cancer is properly considered to be a cancer which starts in the liver, as opposed to a cancer which originates in another organ and migrates to the liver, known as a liver metastasis.

Gene Expression

- Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product.
- These products are often proteins, but in non-protein coding genes such as ribosomal RNA (rRNA) genes or transfer RNA (tRNA) genes, the product is a functional RNA.
- The process of gene expression is used by all known life - eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea) and viruses - to generate the macromolecular machinery for life.

Gene Expression

- Several steps in the gene expression process may be modulated:
  - ***Transcription*** (the process of creating a complementary RNA copy of a sequence of DNA)
  - ***Splicing*** (a modification of an RNA after transcription, in which introns are removed and exons are joined)
  - ***Translation*** (the cellular process in which proteins are produced by decoding, or translating, particular genetic information of the DNA using a messenger RNA (mRNA) intermediate as the template)
  - ***Post–translational modification*** (the chemical modification of a protein after its translation)

Bayesian Networks

- These graphical structures are used to represent knowledge about an uncertain domain.
- Bayesian Networks became extremely popular models in the last decade.
- They have been used for applications in various areas, such as machine learning, text mining, natural language processing, speech recognition, signal processing, bioinformatics, error-control codes, medical diagnosis, weather forecasting, and cellular networks.

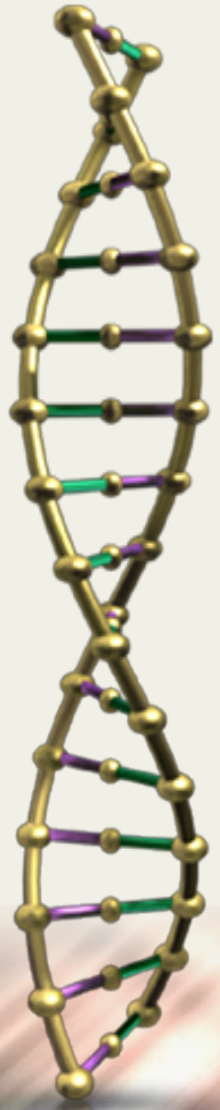
Bayesian Networks

- A Bayesian network (also called belief network) is an augmented directed acyclic graph, represented by the pair V , E where:
  - V is a set of vertices
  - E is a set of directed edges joining vertices. No loops of any length are allowed
- Each vertex in V contains the following information:
  - The name of the random variable
  - A probability distribution table indicating how the probability of this variable's values depends on all possible combinations of parental values

Bayesian Networks

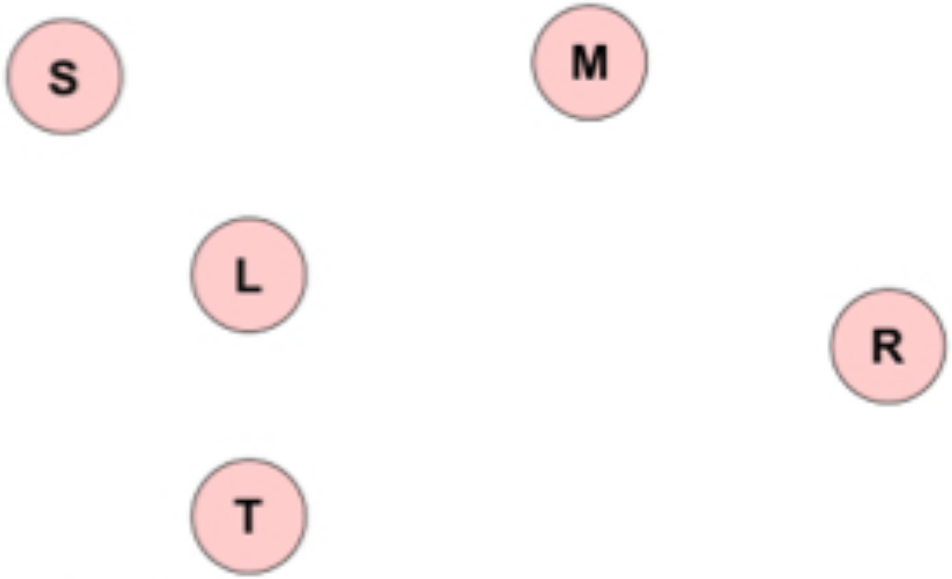
- Steps for building a Bayesian Network:
  1. Choose a set of relevant variables
  2. Choose an ordering for them
  3. Assume they are called X1 .. Xm (where X1 is the first in the ordering, X2 is the second, etc)
  4. For l = 1 to m:
     - (a) Add the Xi node to the network
     - (b) Set Parents(Xi) to be a minimal subset of X1...Xi-1 such that we have conditional independence of Xi and all other members X1...Xi-1 given Parents(Xi)
     - (c) Define the probability table of P(Xi = k — Assignments of Parents(Xi))

Bayesian Networks



Making a Bayes net

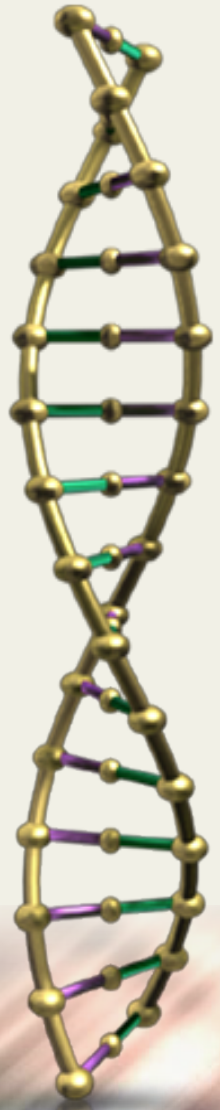
T: The lecture started by 10:35
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Manuela
S: It is sunny



Step One: add variables.

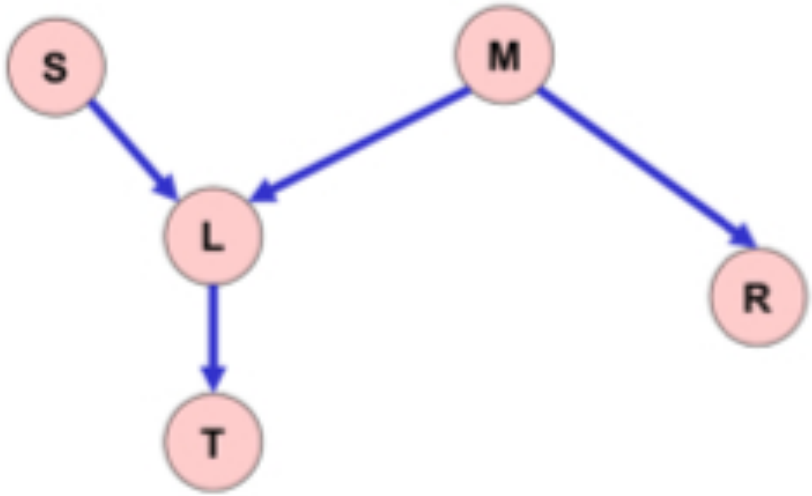
- Just choose the variables you'd like to be included in the net.

Bayesian Networks



Making a Bayes net

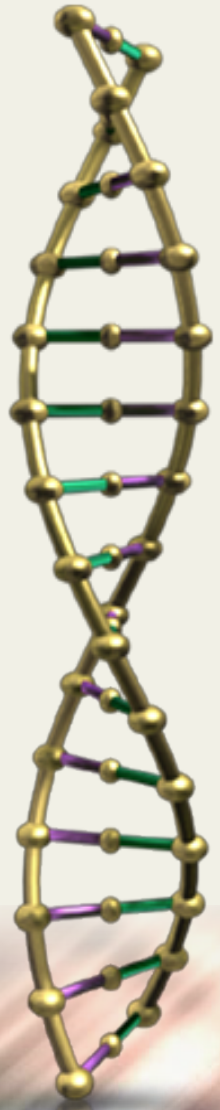
T: The lecture started by 10:35
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Manuela
S: It is sunny



Step Two: add links.

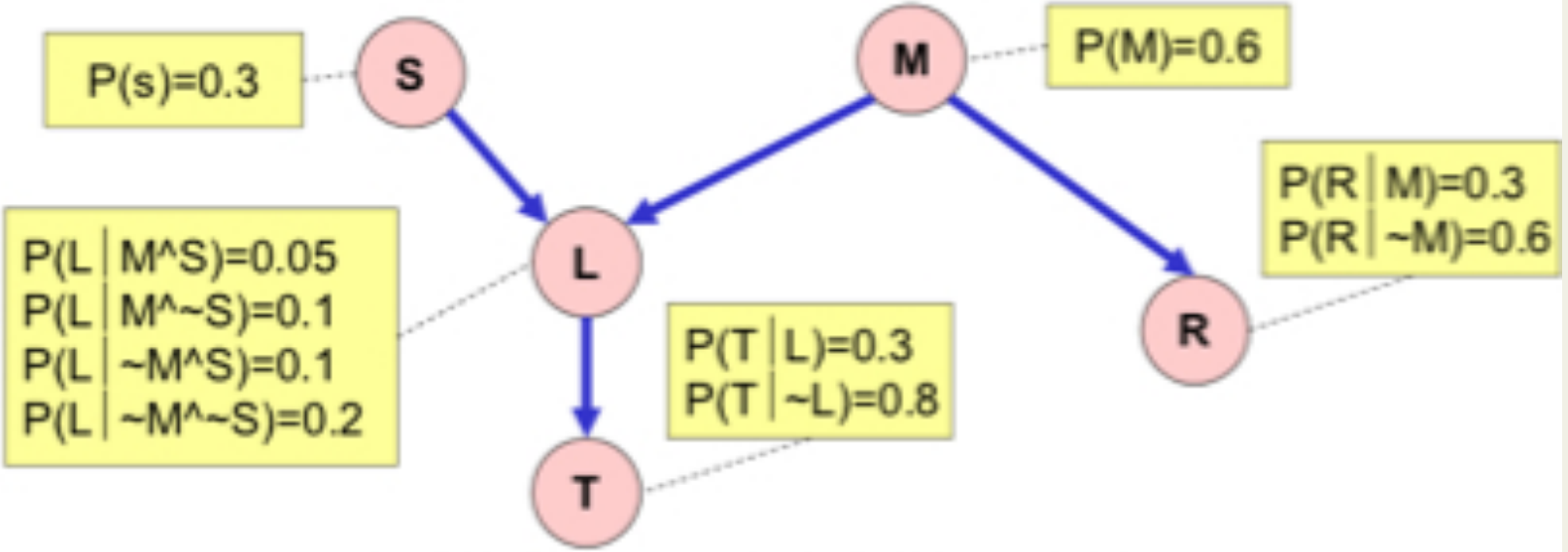
- The link structure must be acyclic.
- If node X is given parents $Q_1,Q_2,..Q_n$ you are promising that any variable that's a non-descendent of X is conditionally independent of X given $\{Q_1,Q_2,..Q_n\}$

Bayesian Networks



Making a Bayes net

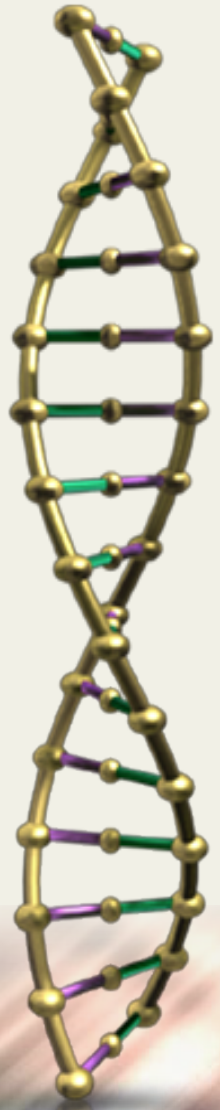
T: The lecture started by 10:35
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Manuela
S: It is sunny



Step Three: add a probability table for each node.

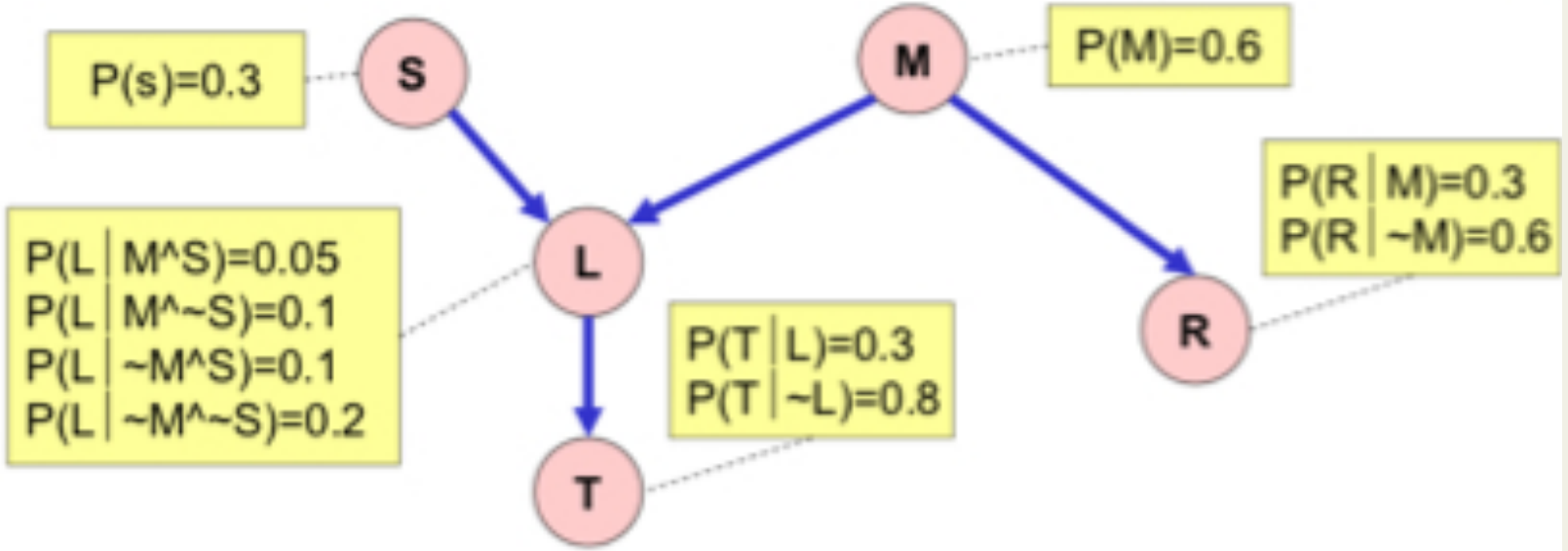
- The table for node X must list P(X|Parent Values) for each possible combination of parent values

Bayesian Networks



Making a Bayes net

T: The lecture started by 10:35
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Manuela
S: It is sunny



- Two unconnected variables may still be correlated
- Each node is conditionally independent of all non-descendants in the tree, given its parents.

Module Networks

- Methods for learning Bayesian Networks can discover dependency structure between observed variables
- Although these methods are very useful, they run into statistical and computational problems in domains that involve a large number of variables
  - i.e. modeling the dependencies among expression levels of all the genes in a cell, or even among changes in stock prices

Module Networks

- One of the more important problems is that in complex domains, the amount of data is almost always not enough to robustly learn a model of underlying distribution
- Generaly in these situations it will lead to spurious dependencies, due to the statistical noise, resulting in models that significantly overfit the data
- For these reasons and some other, there was a need for a new approach to address these issues

Module Networks

- In many large domains, the variables can be partitioned into sets, so that, to a first approximation, the variables within each set have a similar set of dependencies and therefore exhibit a similar behavior
- A new representation called module network was defined, it explicitly partitions the variables into modules
- Each module represents a set of variables that have the same statistical behavior (i.e. they share the same set of parents and local probabilistic model)

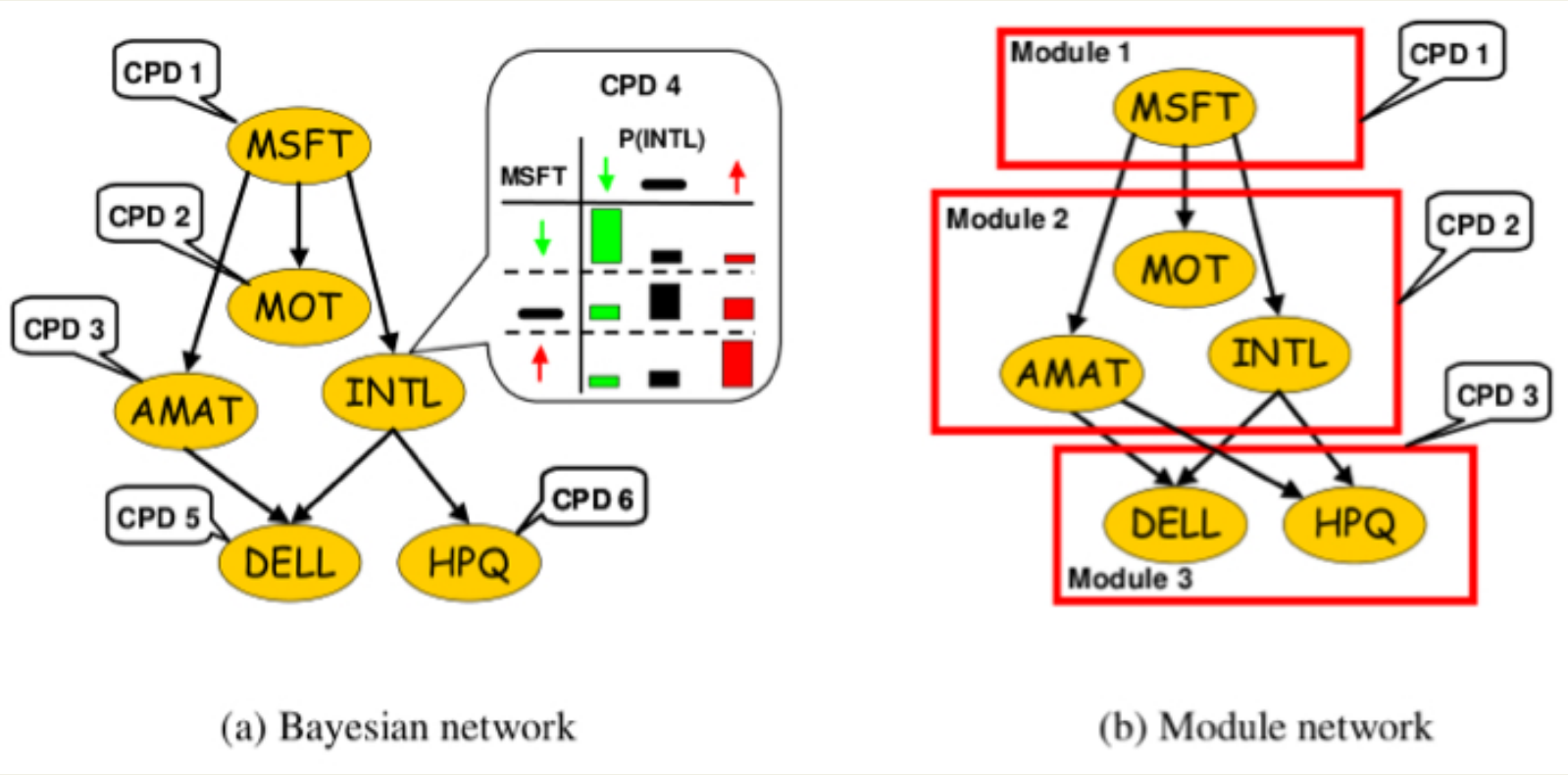
Module Networks

- By enforcing this constraint on the learned network, the complexity of the model space reduces significantly as well as the number of parameters
- These reductions lead to more robust estimation and better generalization on unseen data
- By making the modular structure explicit, the module network representation provides insight about the domain that are often obscured by the intricate details of a large Bayesian Network structure

Module Networks

- A module network can be seen as a Bayesian Network in which variables in the same module share parents and parameters
- Results show that the learned module network (the one used for the tests) generalizes to unseen test data much better that a Bayesian Network. It also ilustrates the ability of the learned module network to reveal high-level structure that provides important insights.

Module Networks



(a) Bayesian network (b) Module network

(a) A simple Bayesian network over stock price variables; the stock price of INTL is annotated with a visualization of its CPD, described as a different multinomial distribution for each value of its influencing stock price MSFT. (b) A simple module network; the boxes illustrate modules, where stock price variables share CPDs and parameters. Note that in a module network, variables in the same module have the same CPDs but may have different descendants.

Module Networks

- A module network is defined by:
  - a specified number of modules
  - an assignment of each variable to a module
  - a shared CPD for the variables in each module
- The learning task thus entails:
  - determining the assignment of variables to modules
  - inducing a CPD for each module

Decoding global gene expression programs in liver cancer by noninvasive imaging

- Non invasive imaging to study the physical and molecular composition of living matter has been used for already a long time now
- Gene expression patterns of cancer can reveal its etiology, prognosis and response to therapy
- The downside of current methods of molecular profiling is that these methods require invasive surgeries for tissue procurement and specialized equipment, leading to a limitation in their routine use
- Current profiling methods provide only single snapshots in time because they are destructive

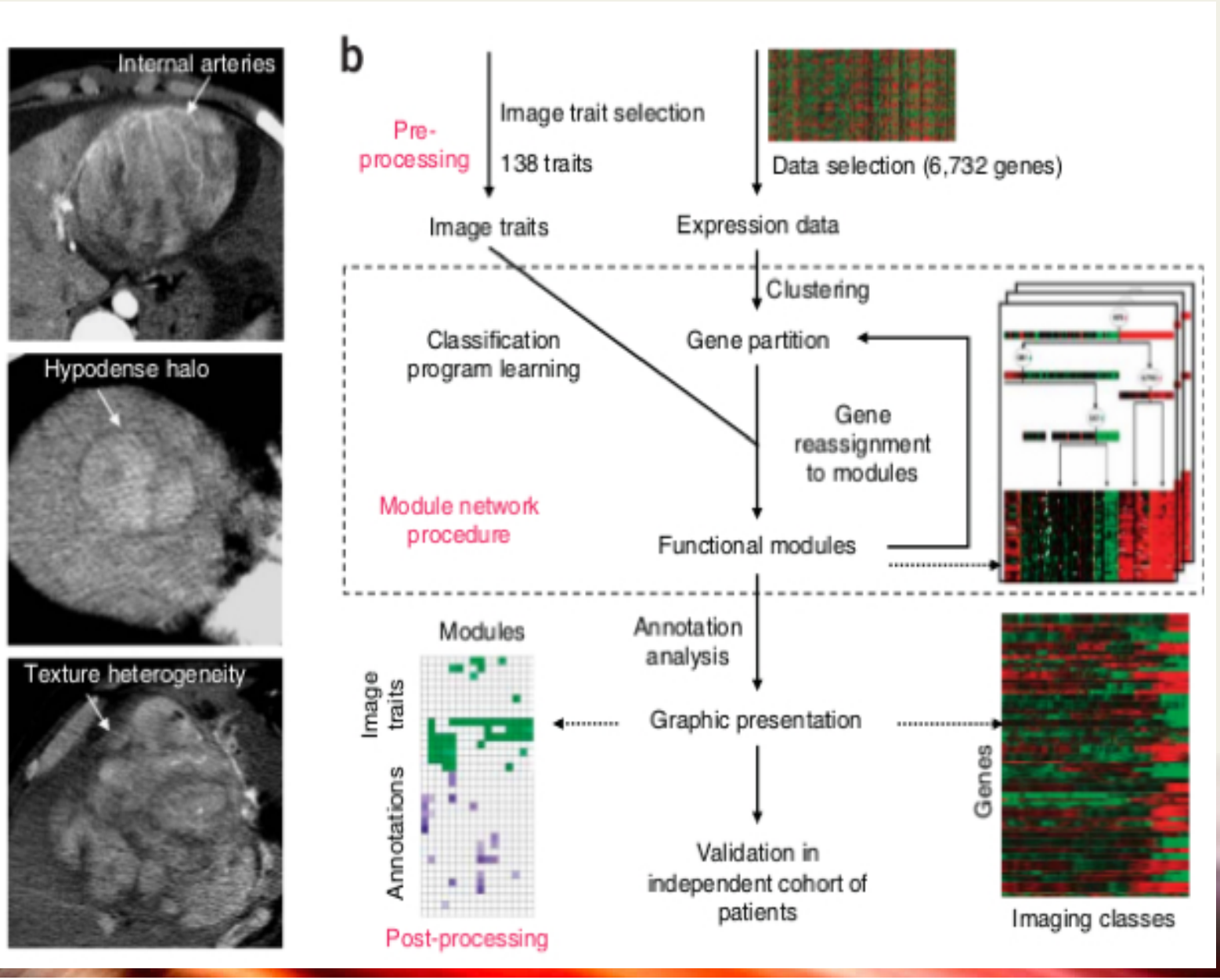
Decoding global gene expression programs in liver cancer by noninvasive imaging

- Human tissues also exhibit diverse distinctive traits on non invasive radiographic images
- To relate gene expression to imaging there are two challenges:
  - the need to define ‘units of distinctiveness’ (traits), from qualitative imaging features, and likewise define coherent patterns of variation from gene expression profiles
  - 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 logic relationships such as AND and OR

Decoding global gene expression programs in liver cancer by noninvasive imaging

- A three step strategy was created in order to address these challenges
- This strategy consists of creating an 'association map' between imaging features on CT scans and gene expression patterns of 28 HCC's
- The first thing done in this strategy was to define and quantify 138 distinctive imaging traits present in one or more HCC's
- The second thing was to adapt the module networks algorithm to systematically search for associations between expression levels of 6732 well-measured genes determined by microarray analysis and combinations of imaging traits
- Third, the validation of the statistical significance of the association map in an independent set of tumors

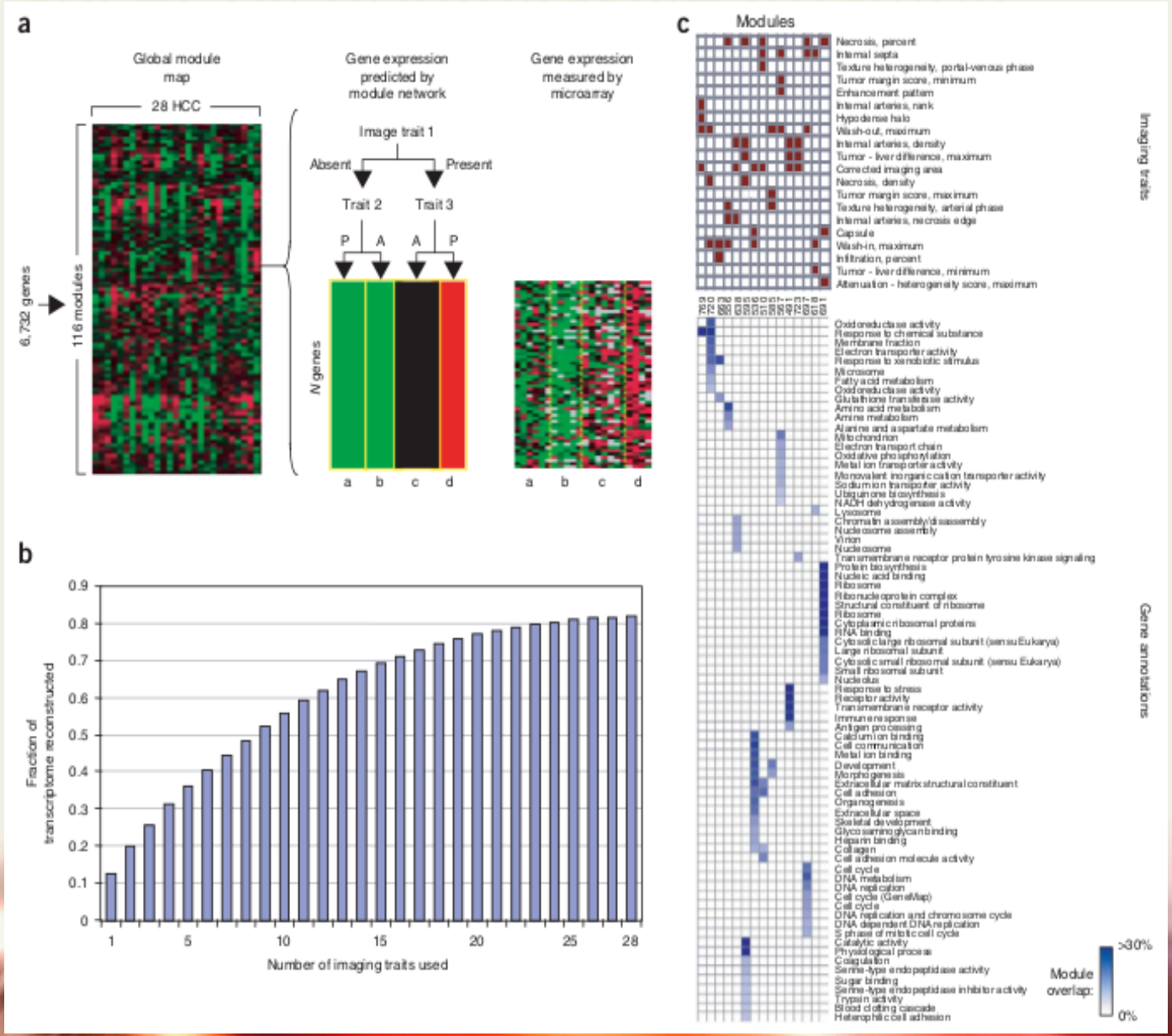
Decoding global gene expression programs in liver cancer by noninvasive imaging



Decoding global gene expression programs in liver cancer by noninvasive imaging

- 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
- The expression variation in 6732 well-measured genes was captured by 116 gene modules, each of which was associated with specific combinations of imaging traits
- The combination of relevant imaging traits are seen 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

Decoding global gene expression programs in liver cancer by noninvasive imaging



Decoding global gene expression programs in liver cancer by noninvasive imaging

- The association map allowed to reconstruct the relative expression level of a gene in a given HCC sample
- The combination of only 28 imaging traits was enough to reconstruct the variation of all 116 gene modules
- For each gene, the number of traits needed to predict its variation is on average three and no more than four in any instance

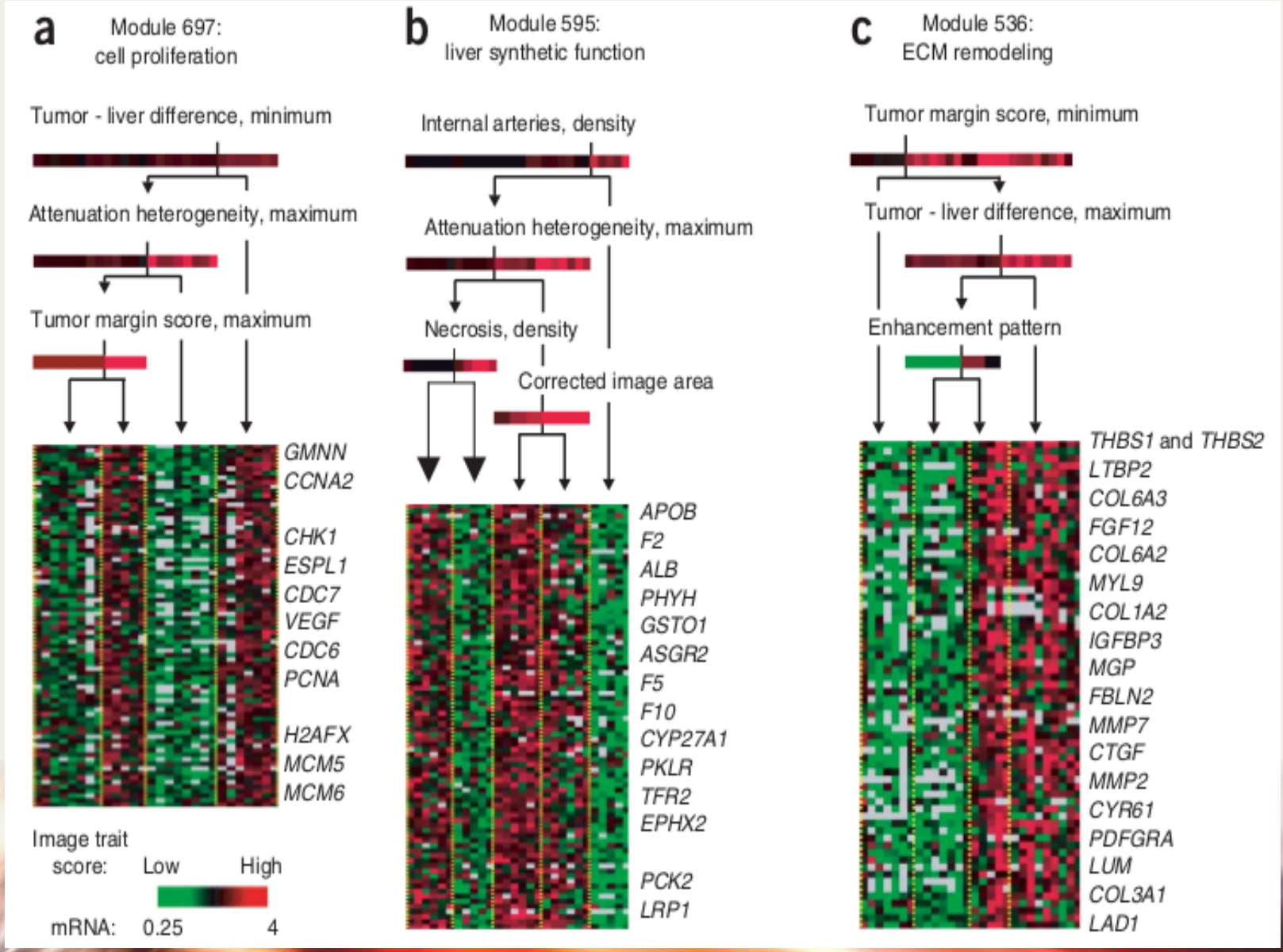
Decoding global gene expression programs in liver cancer by noninvasive imaging

- 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
- Once identified, such 'coding' of imaging traits can be used to translate visual images into global gene expression programs

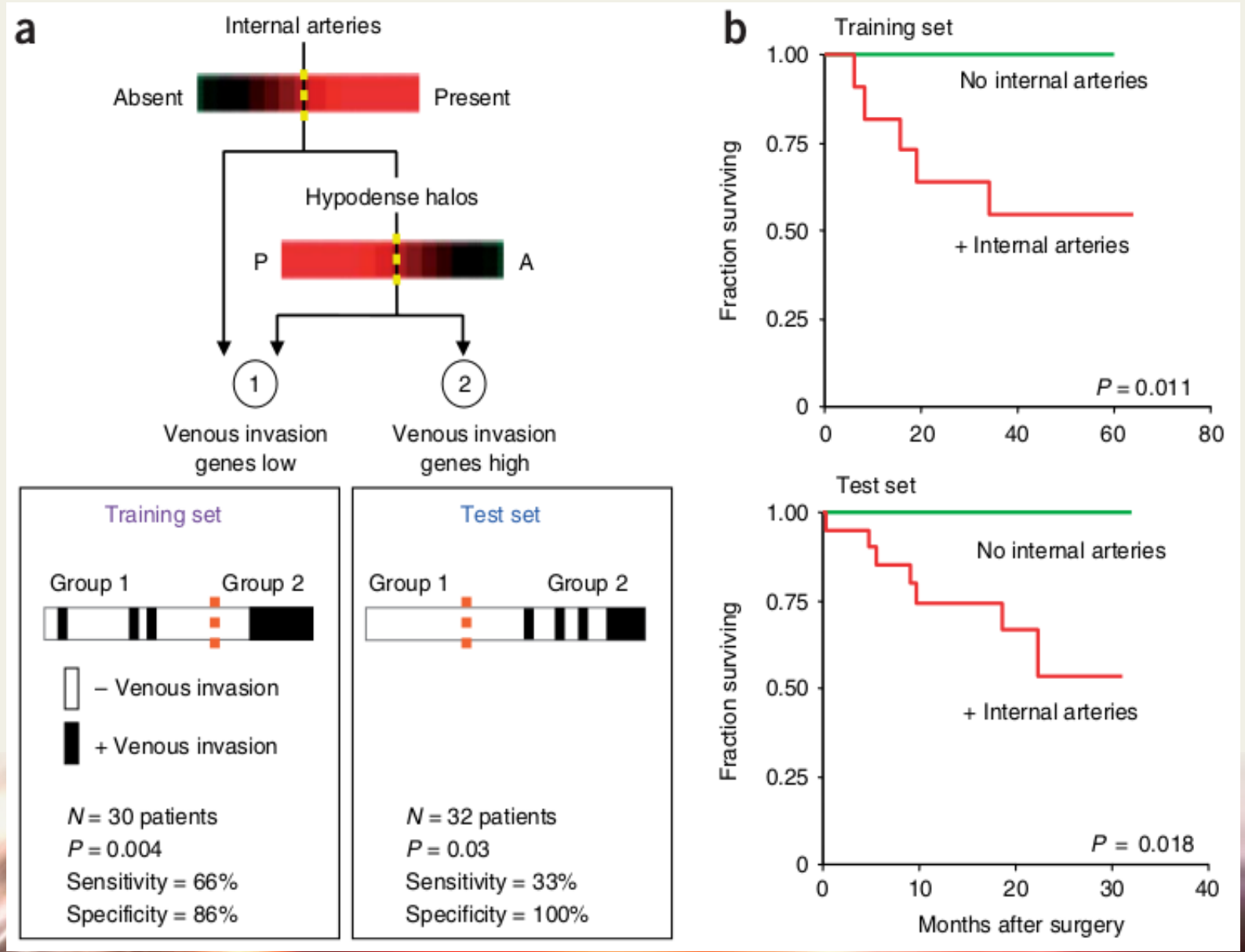
Decoding global gene expression programs in liver cancer by noninvasive imaging

- To further validate the association map, the predictive power of the map was tested in an independent group of 19 prospectively collected patients with HCC
- It was found that 71 out of 116 gene modules, comprising 4996 out of 6732 genes of the transcriptome under consideration, were significantly predicted by their cognate imaging traits
- 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

Decoding global gene expression programs in liver cancer by noninvasive imaging



Decoding global gene expression programs in liver cancer by noninvasive imaging



Decoding global gene expression programs in liver cancer by noninvasive imaging

- It was also 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’
- In 30 patients with HCC, tumors with this combination of imaging traits had a 12 fold increase risk of microscopic venous invasion

Decoding global gene expression programs in liver cancer by noninvasive imaging

- The predicted 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
- The presence of the trait ’internal arteries’ in the preoperative CT scan of HCCs was a significant univariate predictor of overall survival in both groups of patients
- So it can be said that the association map can identify novel imaging traits corresponding to gene expression signatures and provide useful information to guide clinical decision making

Conclusion

- The 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 way.
- The algorithm for associating imaging features and gene expression is generalizable, and in principle may be applied to any disease state and imaging modality.
- This method potentially provides gene expression profiling that is noninvasive, fast, repeatable and in the native anatomic context of the patient.

Obrigado pela atenção.