Inferring transcriptional and microRNA-mediated regulatory programs in glioblastma

Setty, M., *et al*

Goal

- Integrate multiple layers of data for tumor – DNA copy number, promoter methylation, mRNA expression, and miRNA expression.

- Understand the role of miRNA-mediated and transcription factors (TFs) regulation.

- Characterize the pattern of dysregulation in tumors in terms of TFs and miRNAs

Glioblastoma muliforme (GBM)

- Four expression-based subtypes –

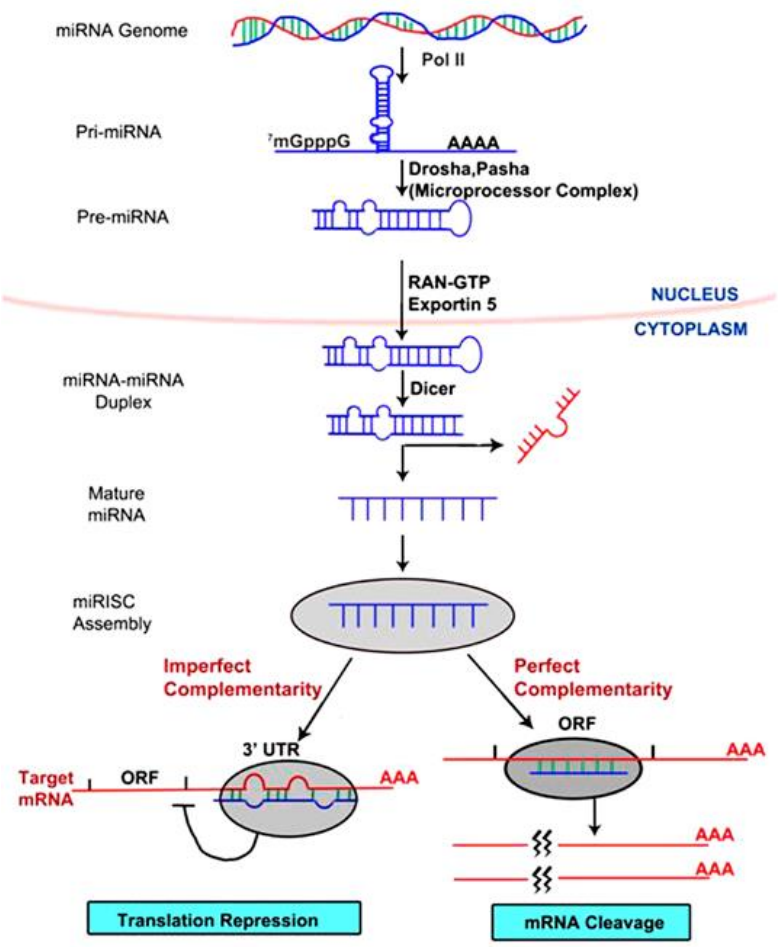
Proneural

Classical

Mesenchymal

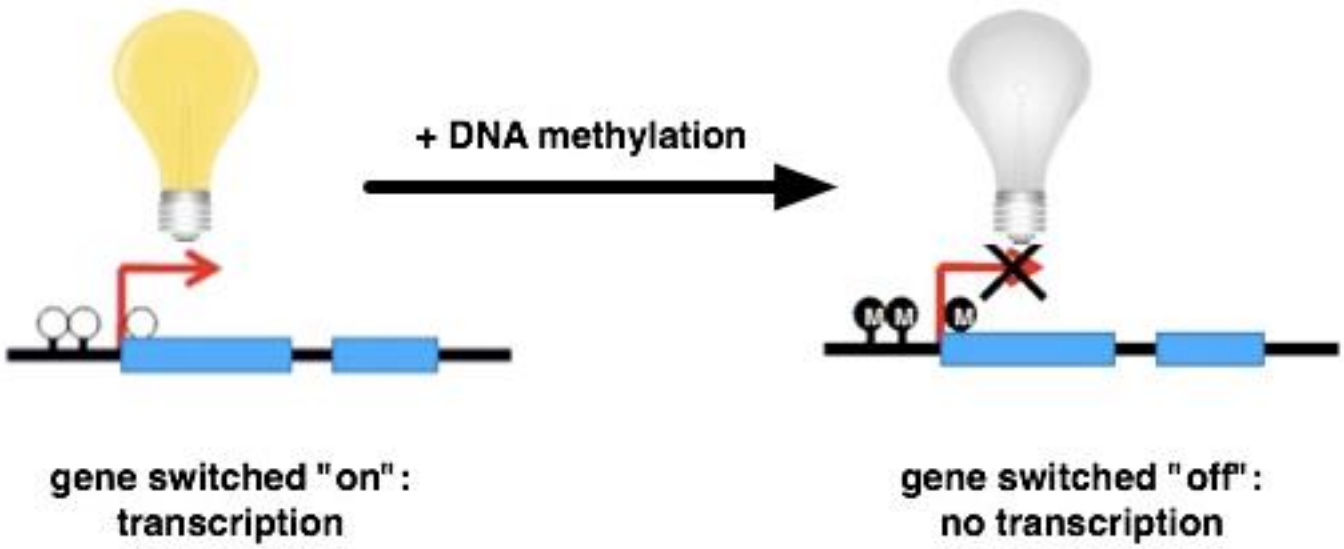
Neural

miRNA Regulation



DNA methylation

DNA methylation is a biochemical process where a methyl group is added to the cytosine or adenine DNA nucleotides.



Why Important to Study miRNA Regulation?

- Impairment of the miRNA regulatory network is viewed as a key mechanism of glioblastma pathogenesis.

- miRNA expression signatures have been used to classify GBM into subtypes related to lineages in the nervous system

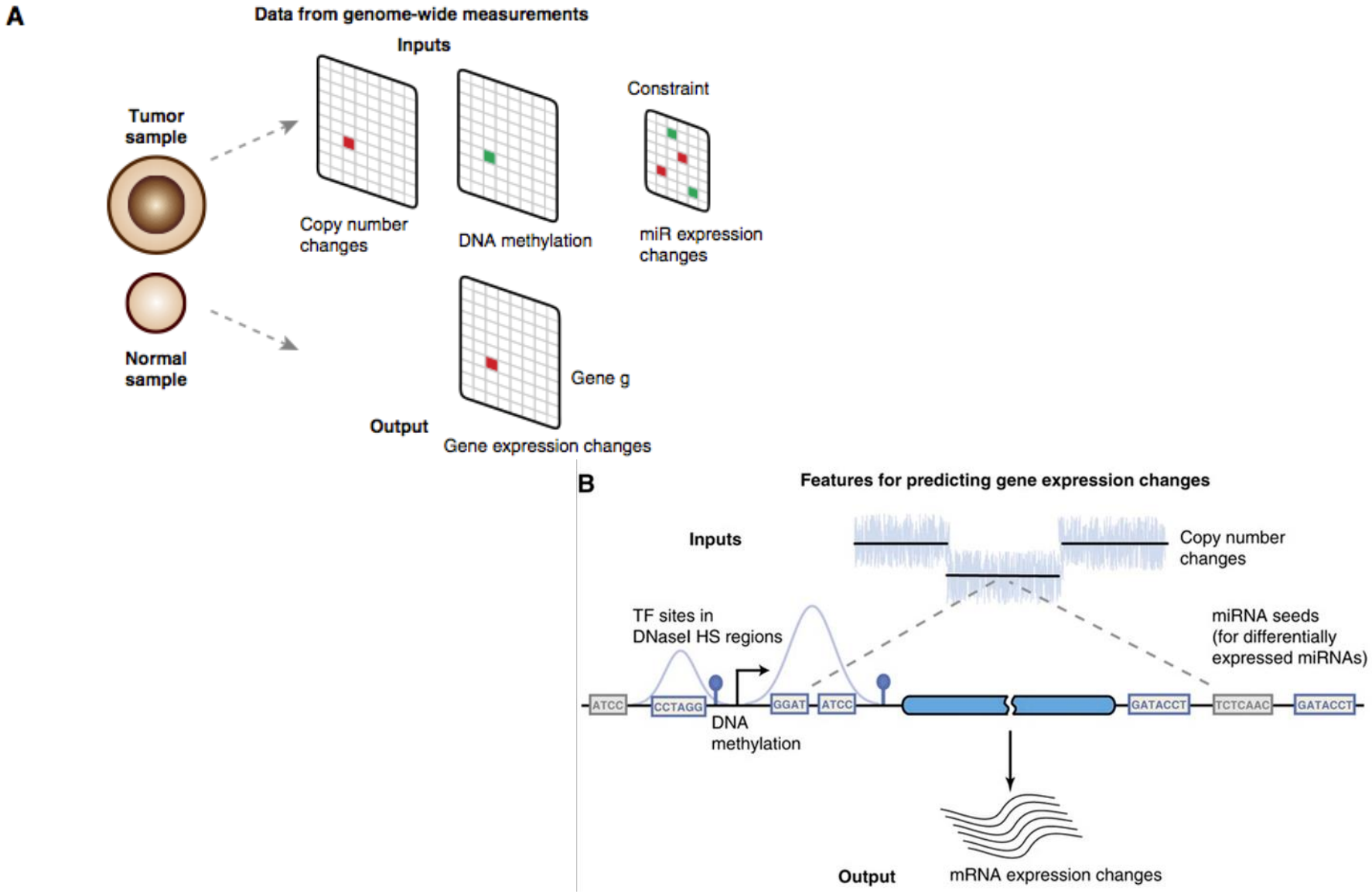
- miR-26a has been shown to promote gliomagenesis in vivo by repression of the tumor suppressor PTEN.

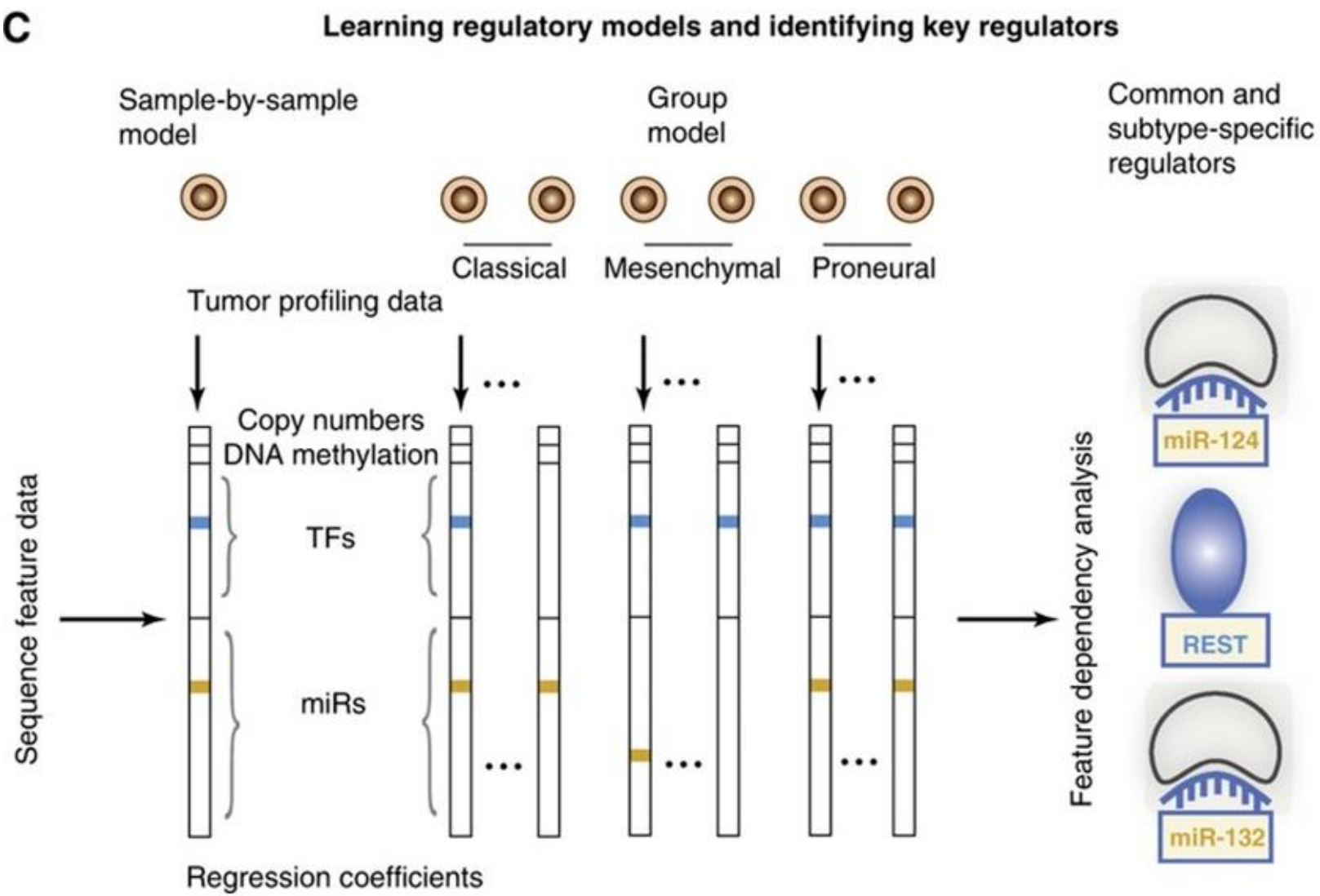
Scheme

- Combine mRNA, copy number and miRNA profiles with regulatory sequence information

- Learn the key direct regulators – TFs and miRNAs using promoter and 3’UTR motif features with sparse regression

Method-outline




C
Learning regulatory models and identifying key regulators
Sample-by-sample model
Group model
Common and subtype-specific regulators
Classical
Mesenchymal
Proneural
Tumor profiling data
Copy numbers
DNA methylation
TFs
miRs
Sequence feature data
Regression coefficients
Feature dependency analysis
miR-124
REST
miR-132

Target prediction for TFs and miRNAs

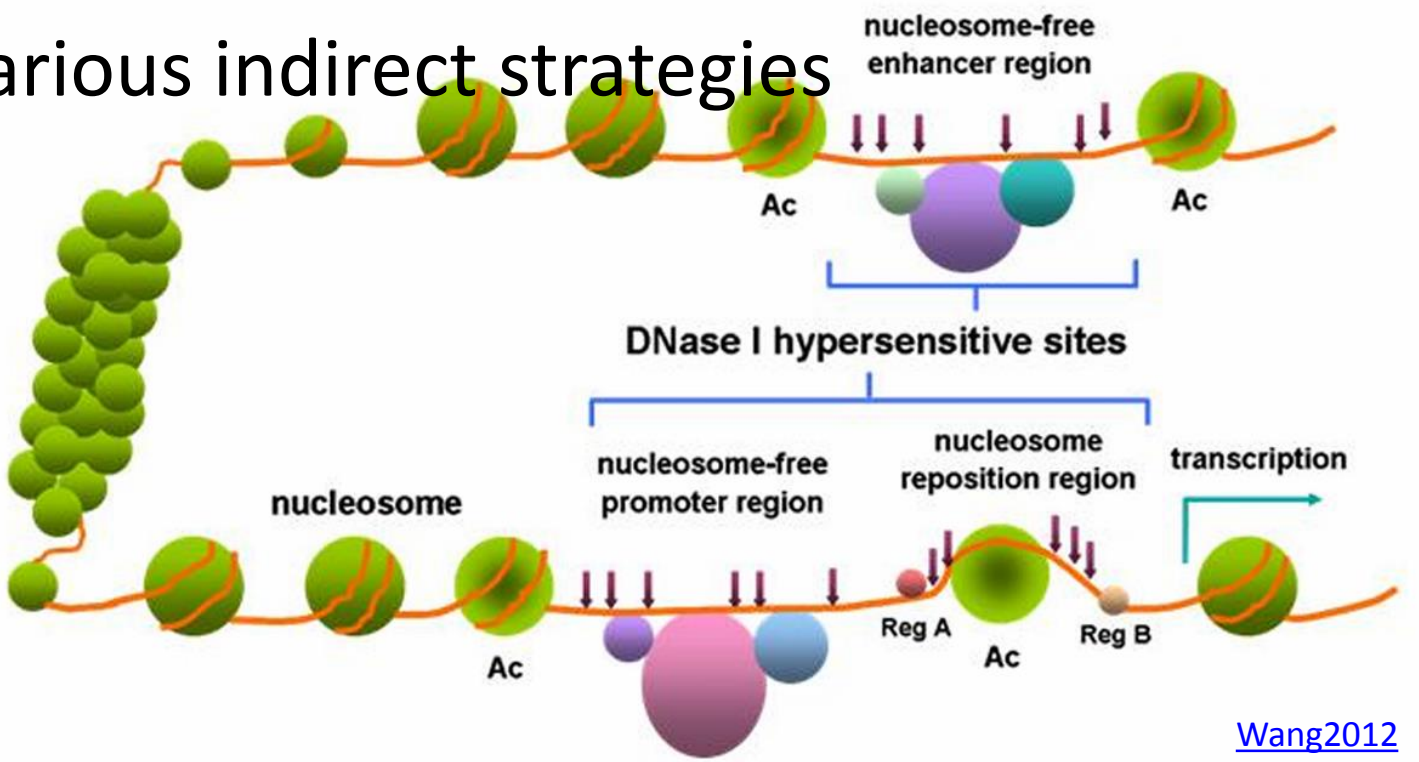
- Determine TFs binding site using DnaseI HS Sequencing

- Determine miRNA binding sites using 7-mer seed matches in the 3’UTR of the Refseq genes.

From Lecture of Jan 22nd by Prof. Gitter

Transcriptional regulation

- ChIP-seq directly measures transcription factor (TF) binding but requires a matching antibody
- Various indirect strategies

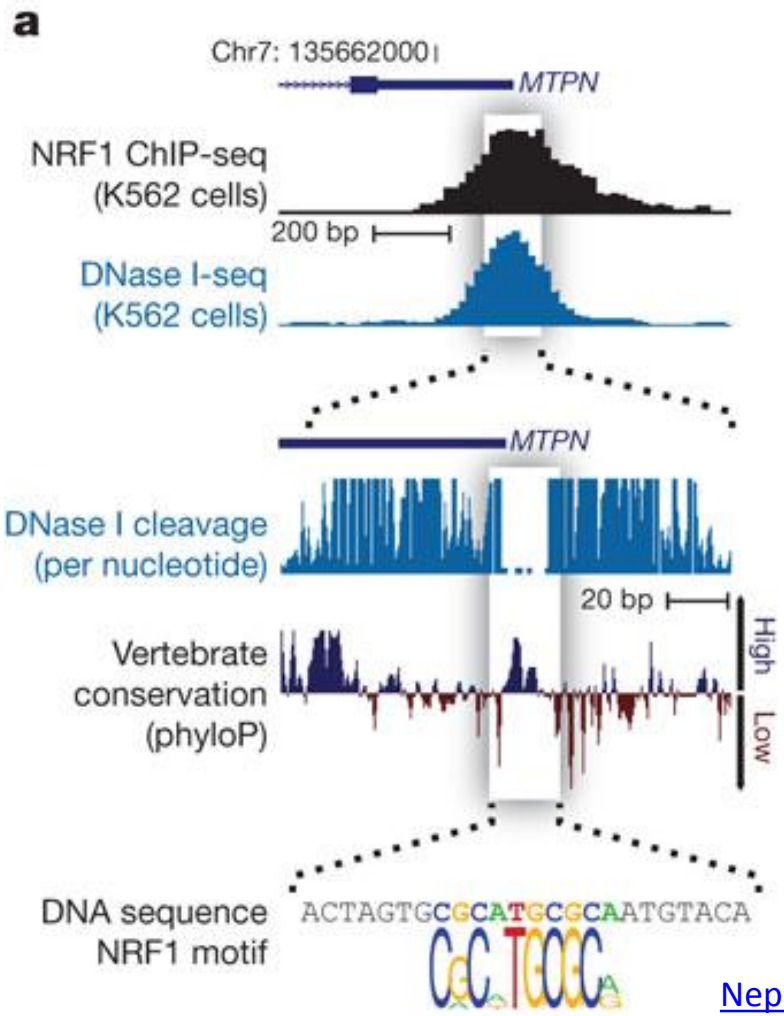


Wang2012

From Lecture of Jan 22nd by Prof. Gitter

Predicting regulator binding sites

- Motifs are signatures of the DNA sequence recognized by a TF
- TFs block DNA cleavage
- Combining accessible DNA and DNA motifs produces binding predictions for hundreds of TFs



Neph2012

Regression model to predict log gene expression changes

- Counts of TF and miRNA binding sites
- An estimate of gene's average copy number
- Promoter DNA methylation

$$y_g \approx w^{\mathrm{CN}} C_g + \sum_{\mathrm{miR}} w^{\mathrm{miR}} N_{g,\mathrm{miR}} + \sum_{\mathrm{TF}} w^{\mathrm{TF}} N_{g,\mathrm{TF}}$$

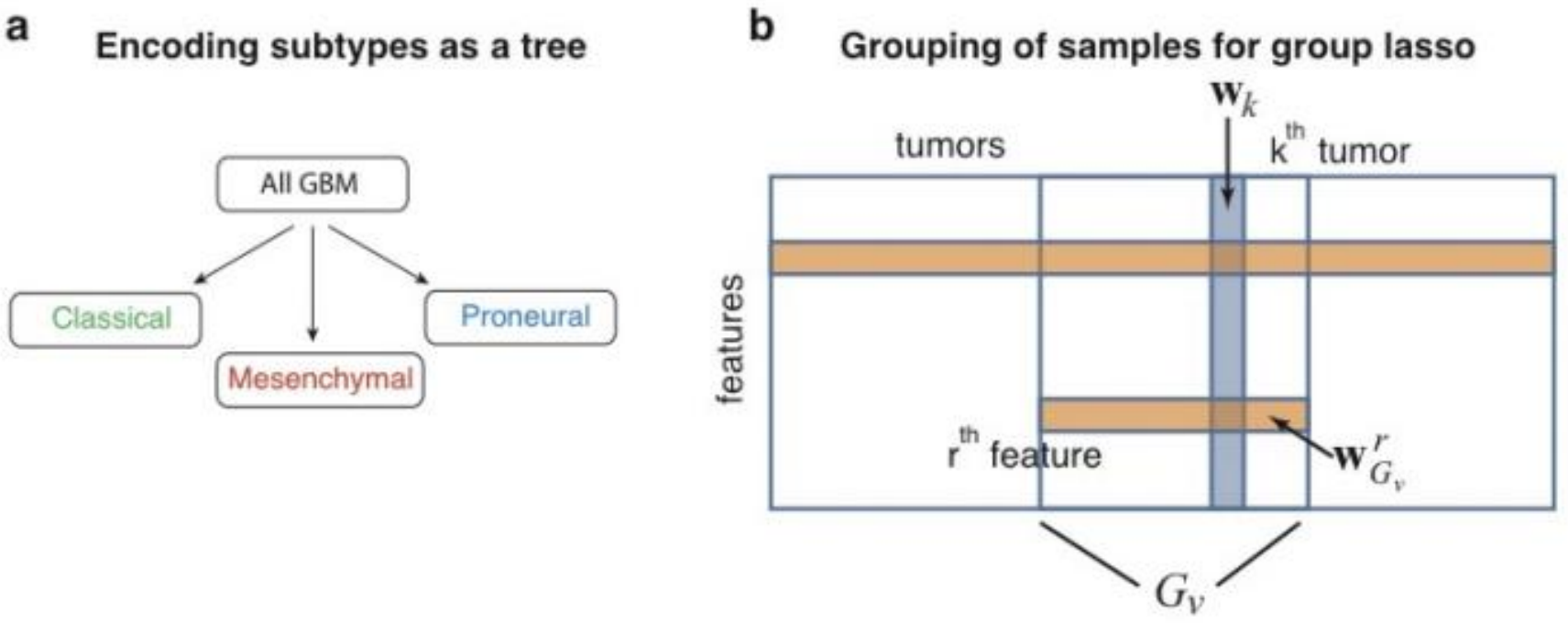
Lasso regression models

- To avoid overfitting
- Use lasso constraint to identify a small number of TFs and miRNA

$$\min_{w} \sum_{g} (y_g - w \cdot x_g)^2 + \lambda \sum_{r \in \{CN,miR,TF\}} | w^r |$$

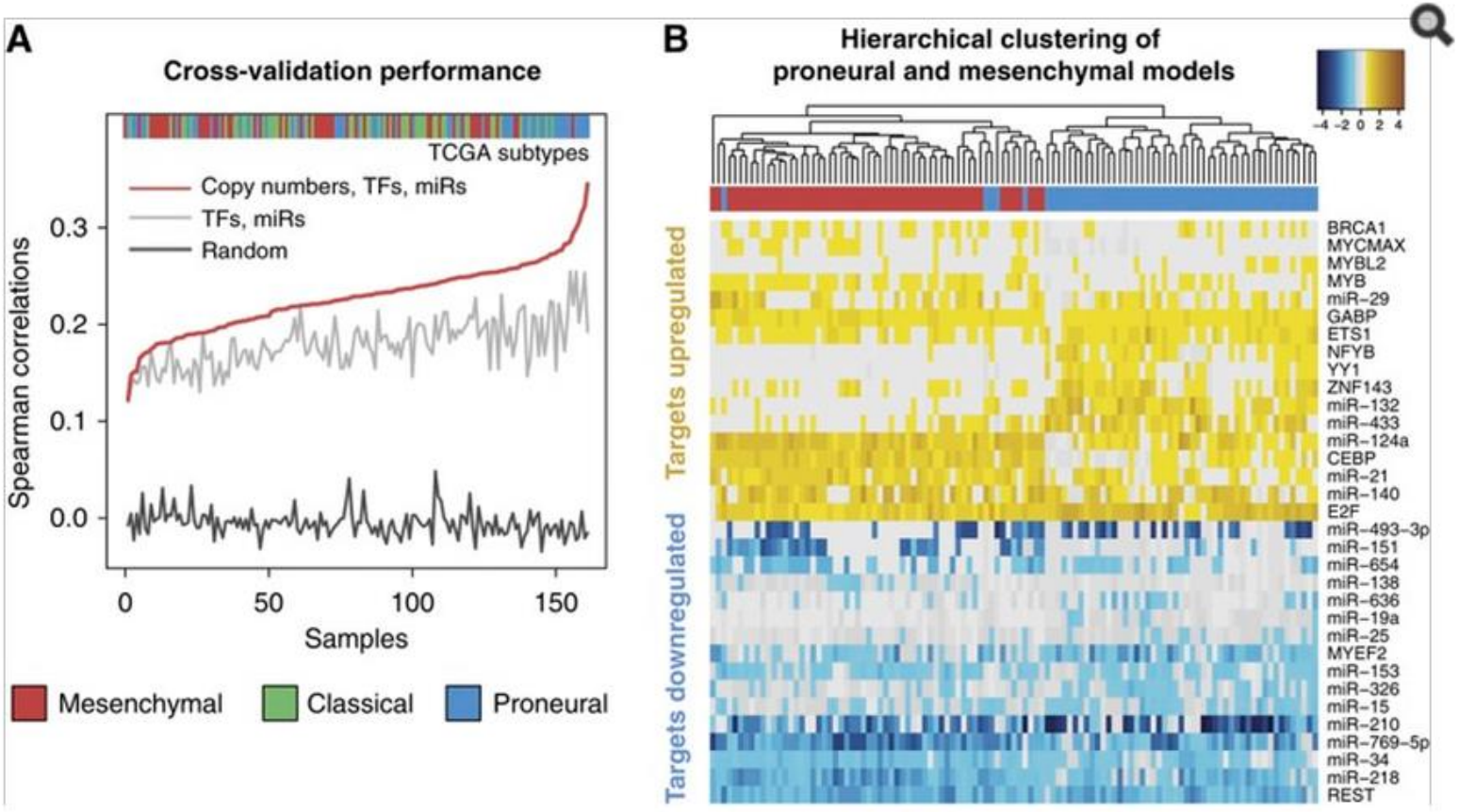
Joint Learning with Group Lasso

$$\min_{w} \sum_{g,k} (y_{g,k} - w_k \cdot x_{g,k})^2 + \lambda \sum_{r \in \{CN,miR,TF\}} \sum_{v} a_v \| w^r_{G_v} \|_2$$



Sparse Regression Models Predict Differential of Subtypes of Tumor Samples

Figure 2



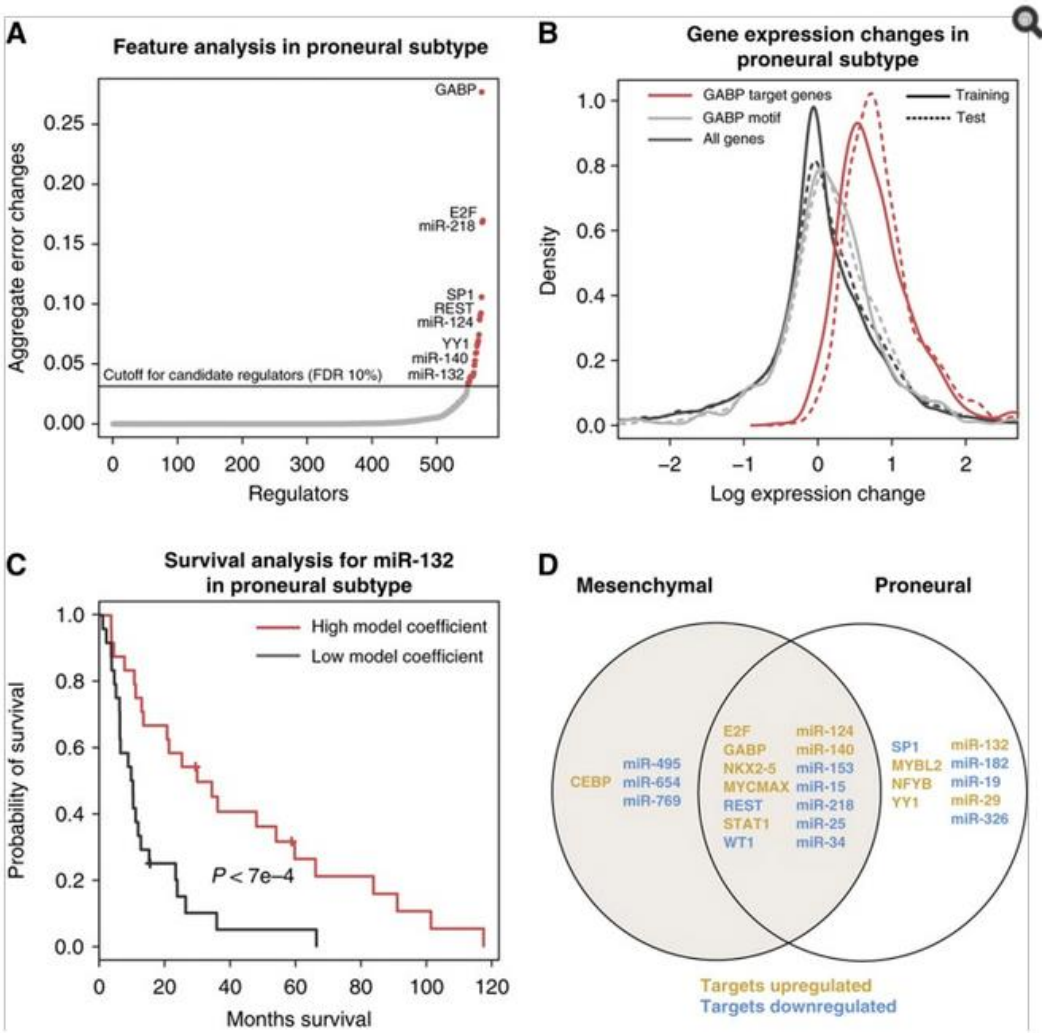
Dependency analysis

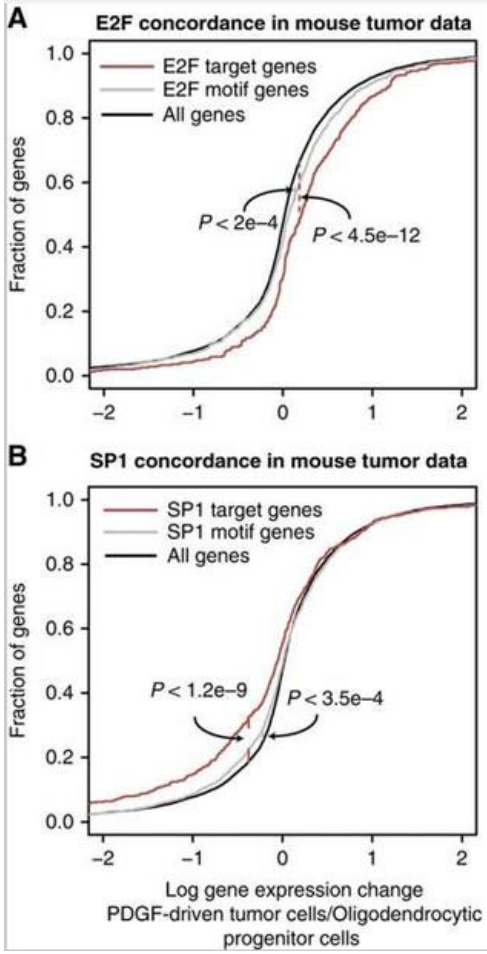
- To determine regulators (TFs and miRNAs) that significantly account for common and subtype-specific gene expression changes.

$$\text{score}(r,v) = \sum_{g} \text{score}(r,v,g) = \sum_{g} \sum_{k \in G_v} [L(y_{g,k}, w_k^{r \to 0} \cdot x_{g,k}) - L(y_{g,k}, w_k \cdot x_{g,k})]$$

Results - Feature Analysis of Group Models Identifies Common and Subtype Specific Regulators

Figure 3

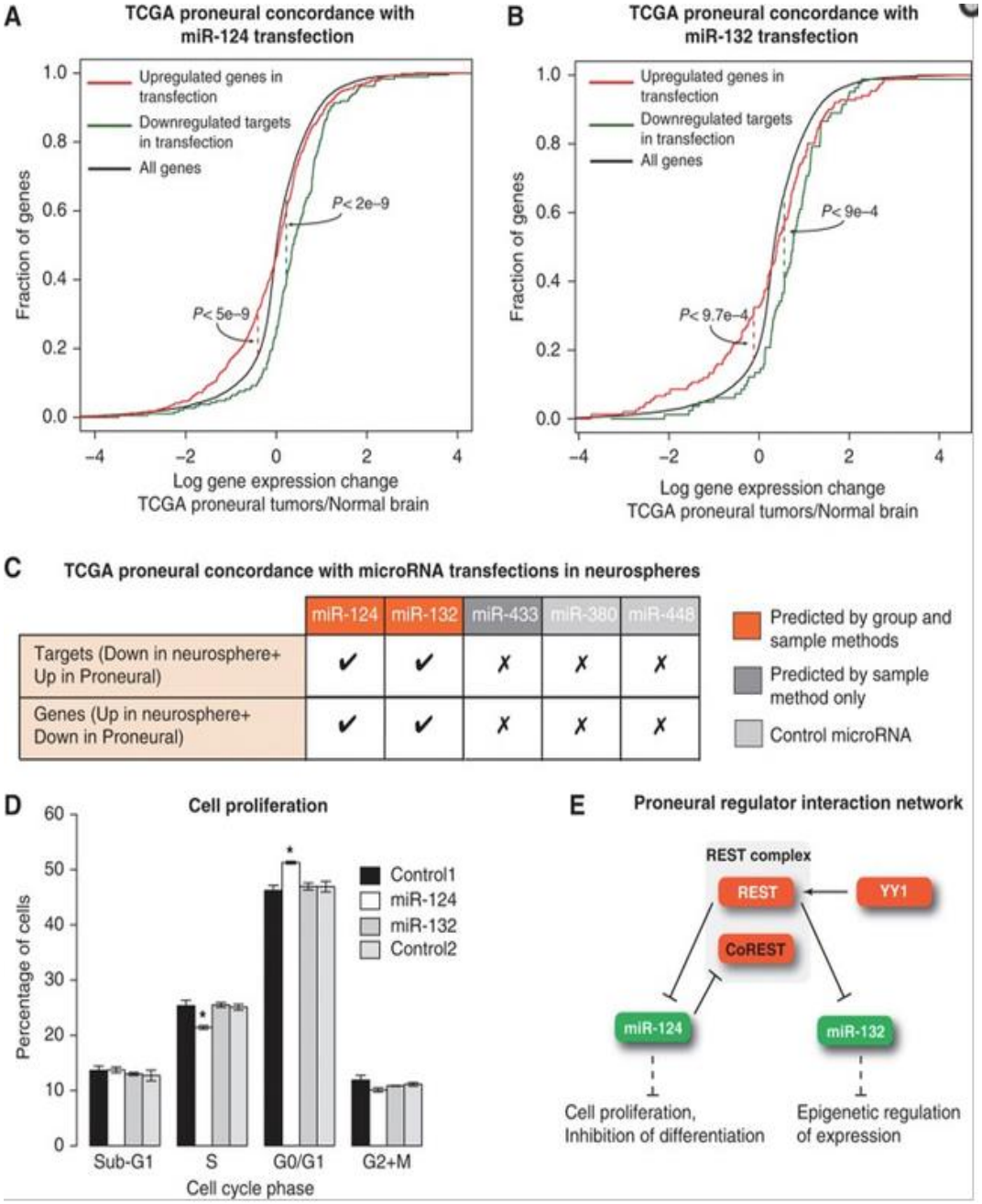




C

Candidate regulators in proneural subtype

Candidate regulator	Ontologies associate with geneset	Target regulation	Regulator significance *P*-value
GABP	DNA metabolism, DNA replication	Up	2e–6
E2F	Cell-cycle process	Up	2e–6
miR-218	Synaptic transmission	Down	2e–6
SP1		Down	2e–6
miR-15	Cell morphogenesis, microtubule process	Up	9e–6
REST		Down	2e–5
MYBL2	Mitotic cell cycle, chromosome organization	Up	2e–5
WT1	Neuron development, axonogenesis	Down	8e–5
miR-153	Cell–cell signaling, synaptic transmission	Down	9e–5
miR-124	Cell differentiation	Up	9.5e–5
miR-182		Down	2e–4
miR-34	Neuron development	Down	2e–4
YY1	RNA metabolism, chromosome organization	Up	5e–4
miR-29	Matrix organization	Up	5.5e–4
miR-140	Cell development and migration	Up	1e–3
NKX2-5	Regulation of apoptosis	Up	1.2e–3
STAT1		Up	1.3e–3
miR-25		Down	1.5e–3
miR-132	Epigenetic regulation of gene expression	Up	1.5e–3
MYCMAX	RNA metabolic process	Up	1.5e–3
NFYB	Cell-cycle process	Up	2e–3
miR-326		Down	2.5e–3
miR-19	Signal transmission	Down	3e–3


A
TCGA proneural concordance with miR-124 transfection
Upregulated genes in transfection
Downregulated targets in transfection
All genes
P< 2e–9
P< 5e–9
Fraction of genes
Log gene expression change TCGA proneural tumors/Normal brain
B
TCGA proneural concordance with miR-132 transfection
Upregulated genes in transfection
Downregulated targets in transfection
All genes
P< 9e–4
P< 9.7e–4
Fraction of genes
Log gene expression change TCGA proneural tumors/Normal brain
C
TCGA proneural concordance with microRNA transfections in neurospheres
miR-124
miR-132
miR-433
miR-380
miR-448
Targets (Down in neurosphere+ Up in Proneural)
Genes (Up in neurosphere+ Down in Proneural)
Predicted by group and sample methods
Predicted by sample method only
Control microRNA
D
Cell proliferation
Percentage of cells
Control1
miR-124
miR-132
Control2
Sub-G1
S
G0/G1
G2+M
Cell cycle phase
E
Proneural regulator interaction network
REST complex
REST
YY1
CoREST
miR-124
miR-132
Cell proliferation, Inhibition of differentiation
Epigenetic regulation of expression

- Thanks for your attention!