Construction of cancer pathways for personalized medicine

Predictive, Preventive and Personalized Medicine & Molecular Diagnostics

Presented By Dr. Anton Yuryev

Date Nov 4, 2014

Curse of Dimensionality of OMICs data:
We will never have enough patient samples to calculate robust signatures from large scale molecular profiling data

Hua et al. *Optimal number of features as a function of sample size for various classification rules*. Bioinformatics. 2005

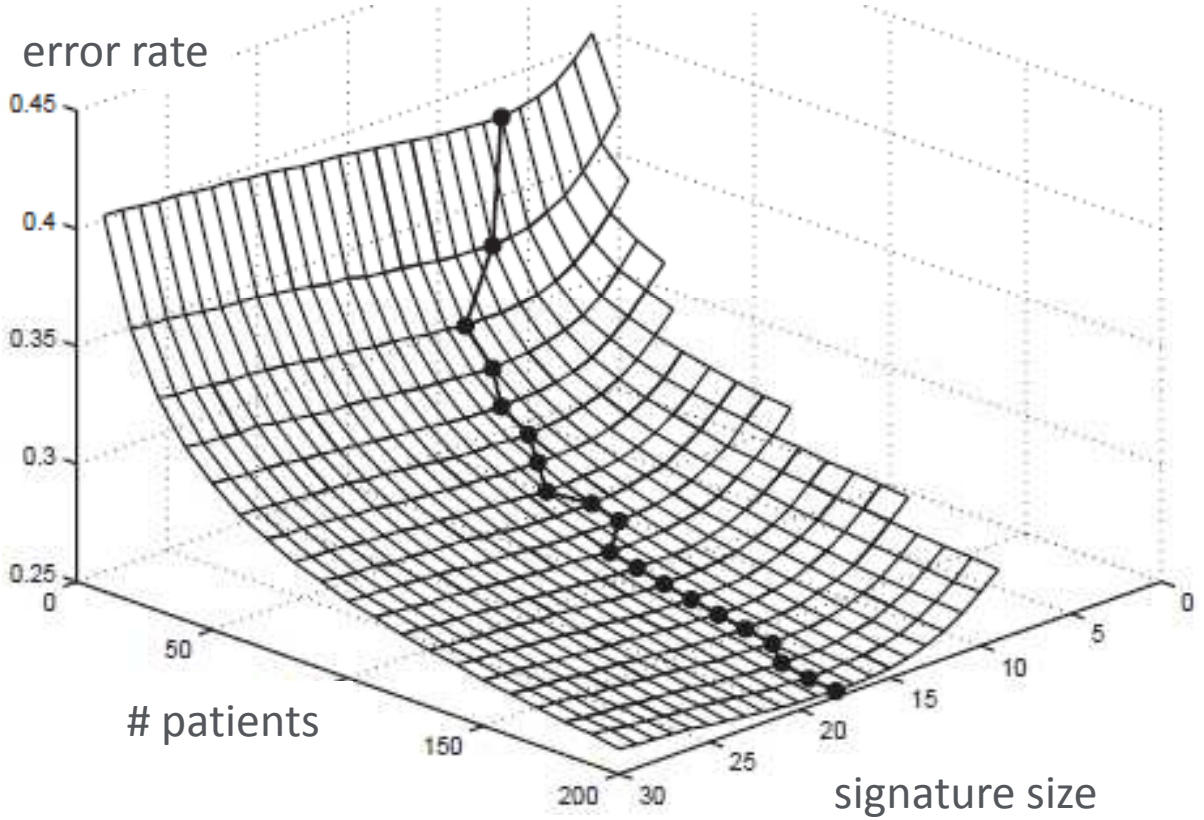


Fig.3 Optimal feature size versus sample size for **Polynomial SVM** classifier.
nonlinear model, correlated feature, G=1, ρ= 0.25. σ^2 is set to let Bayers error be 0.05

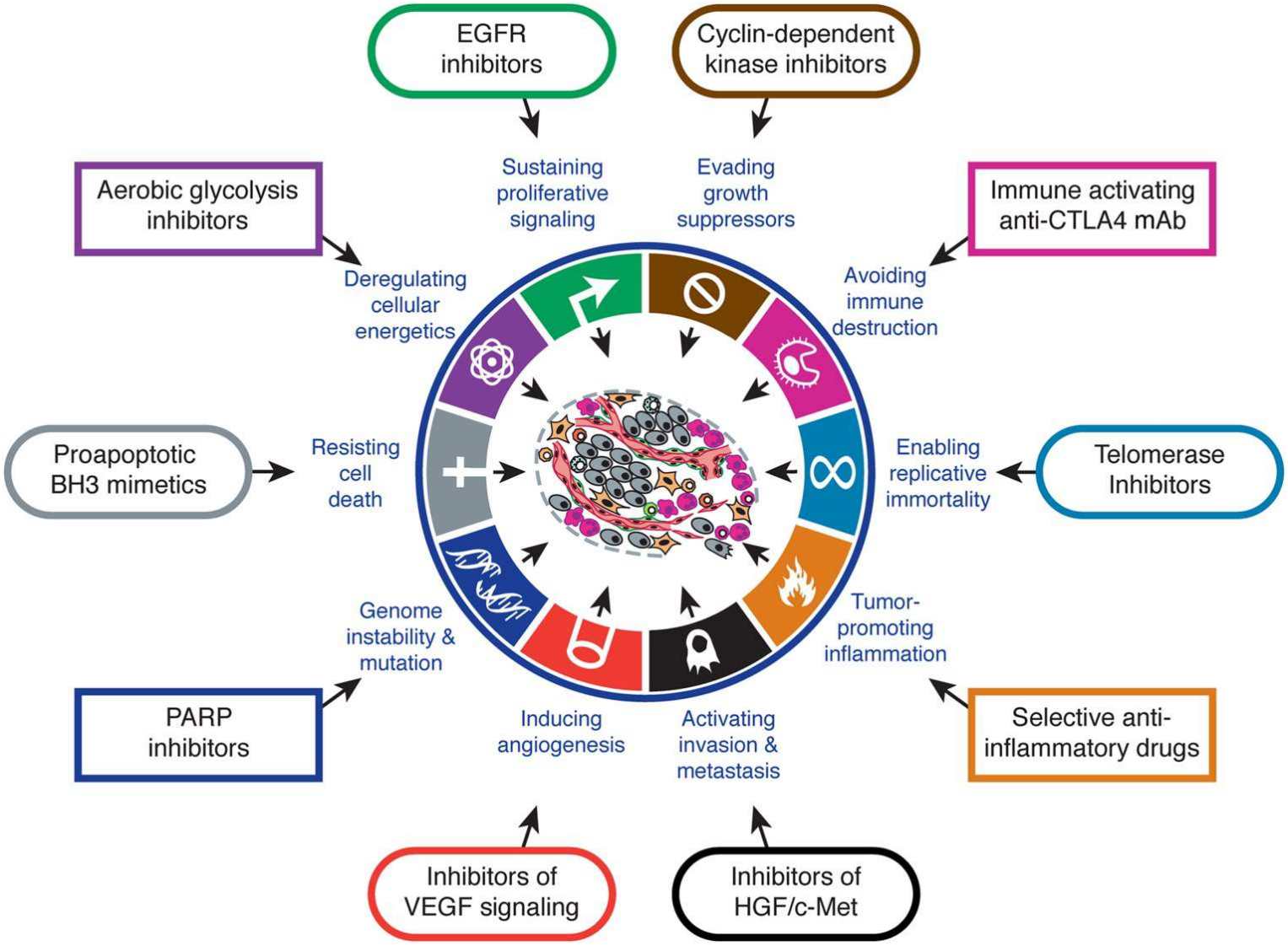
Mathematical requirements for short signature size vs. Biological reality

Mathematical requirement	Biological reality
Signature size must be **20-30** genes	Typical cancer transcriptomics profile has **500-2000** differentially expressed genes with p-value < 0.005
Increasing number of samples above **200** does not change optimal signature size	Typical cancer dataset has not more than **100** patients.
	Increasing number of patients results in finding different **cancer sub-types** each having small number of samples
Error rate and robustness of signatures from **uncorrelated** feature is better than from correlated features	Most DE genes are **correlated** due to transcriptional linkage and different TFs regulated by only few biological pathways
	We can use **prior knowledge** about transcriptional regulation to select most uncorrelated features, e.g. genes controlled by different TFs in different pathways

Our solution: Pathway Activity signature

SNEA (sub-network enrichment analysis) -> pathway analysis

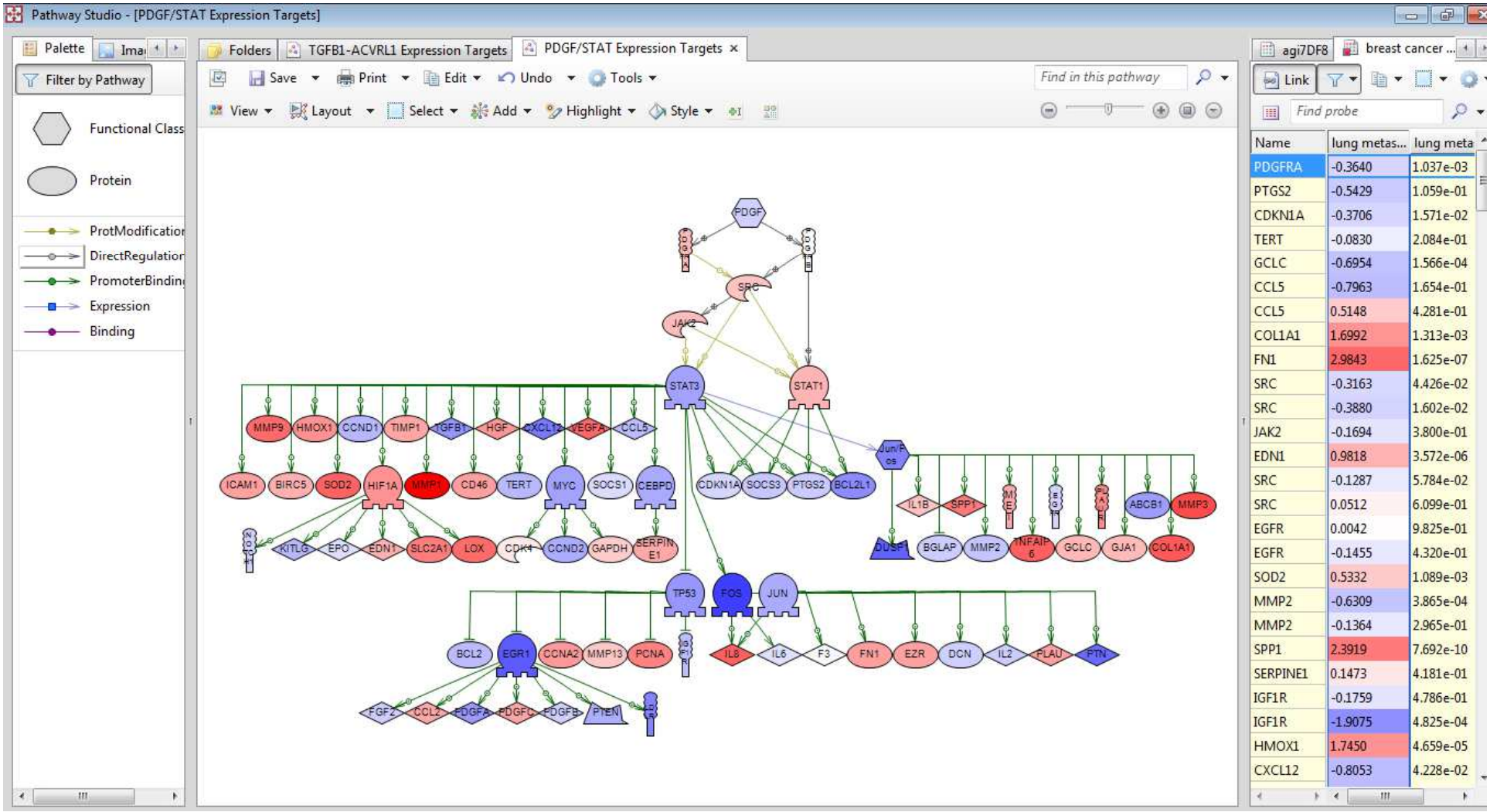
Hanahan & Weinberg. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74



Common misconception

Pathway activity = Differential Expression of its expression targets

Pathway activity ≠ Differential Expression of its components



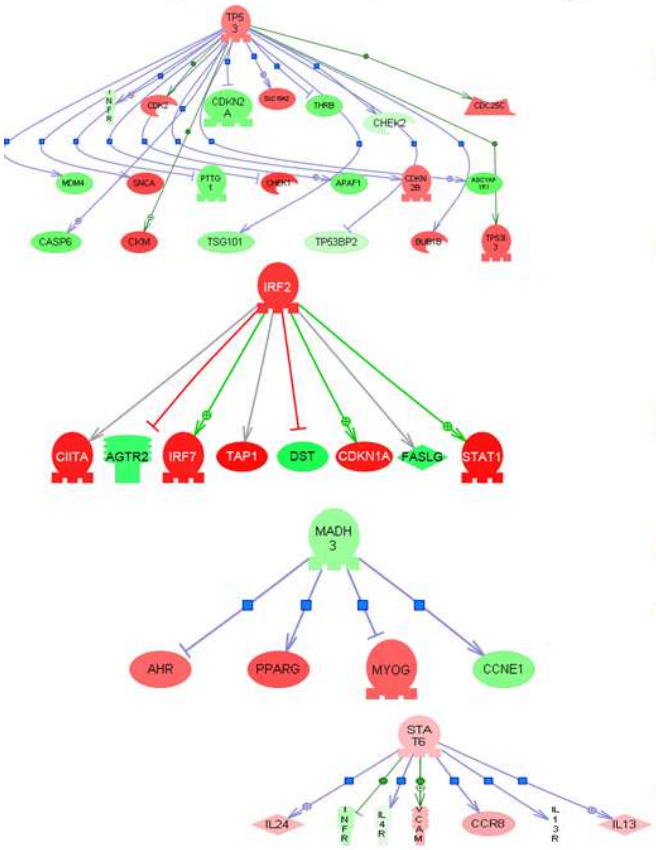
STEP1: SNEA

calculating activity transcriptional activity of upstream regulators

Input: DE fold changes + prior knowledgebase of known expression regulation events

SNEA	Reverse Causal Reasoning
Mann-Whitney enrichment test	Fisher's overlap test

Lower p-value (more significant)



Higher p-value (less significant)

- SNEA builds networks from all genes/proteins measured in the experiment using all relations in the database.
- SNEA can include indirect regulation i.e. expression regulatory cascades consisting of 2-3 steps
- Significant network centers may be found that are not measured in the primary dataset
- No prior curation of gene sets is required.
- Can work with partial information about TF targets. Does not require knowledge about all targets for TF
- P-value is sensitive to the size of the chip

Molecular networks in microarray analysis.
Sivachenko A, Yuryev A, Daraselia N, Mazo I. J Bioinform Comp. Biol. 2007

Pathway Studio Knowledgebase for SNEA powered by Elsevier NLP

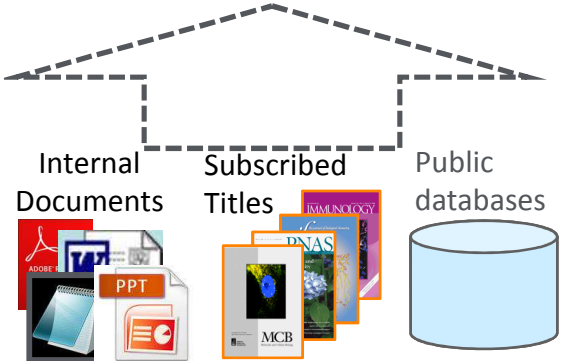
23,641,270 Pubmed abstracts from >9,500 journals

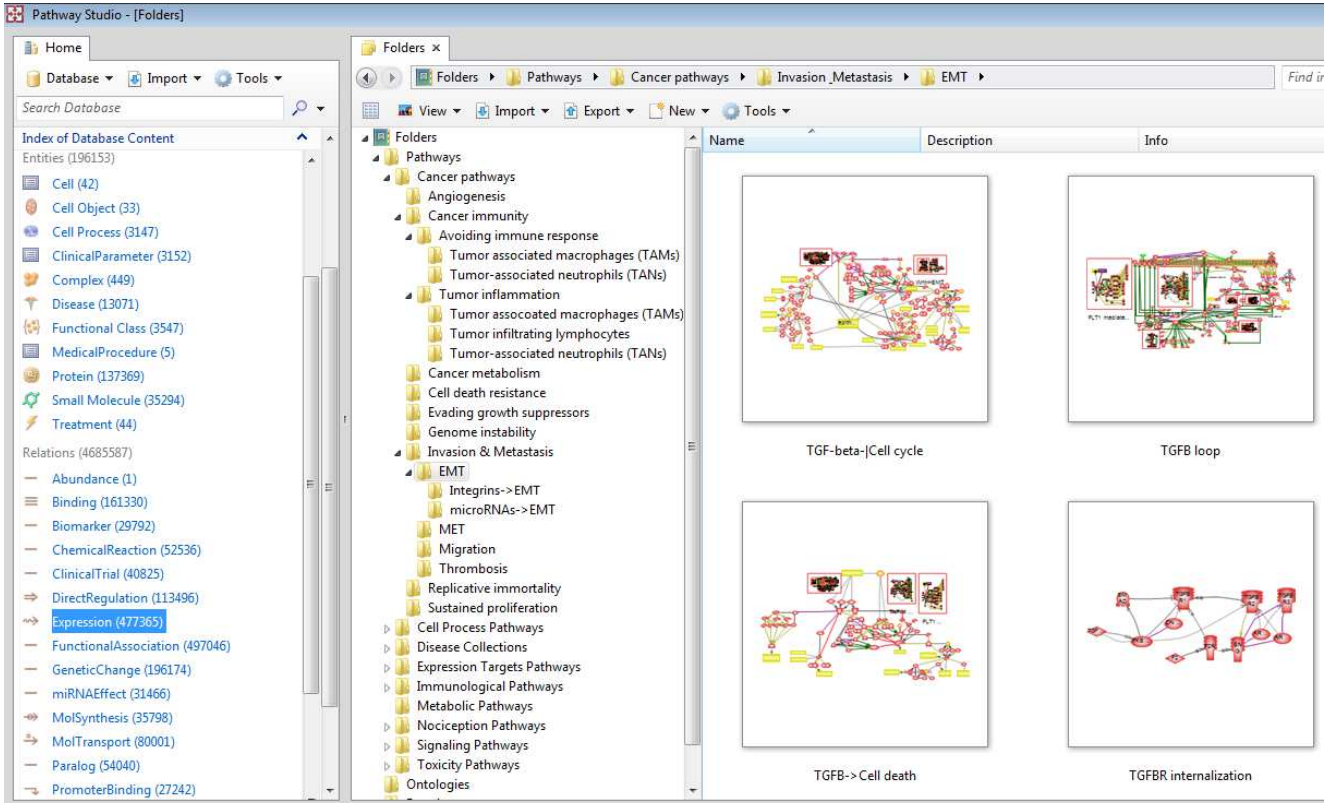
613 Elsevier journals

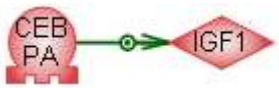
884 non-Elsevier journals

>3,500,000 full-text articles

Custom data can be imported into dedicated PS instance

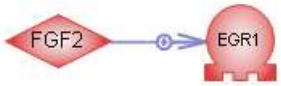






27,243

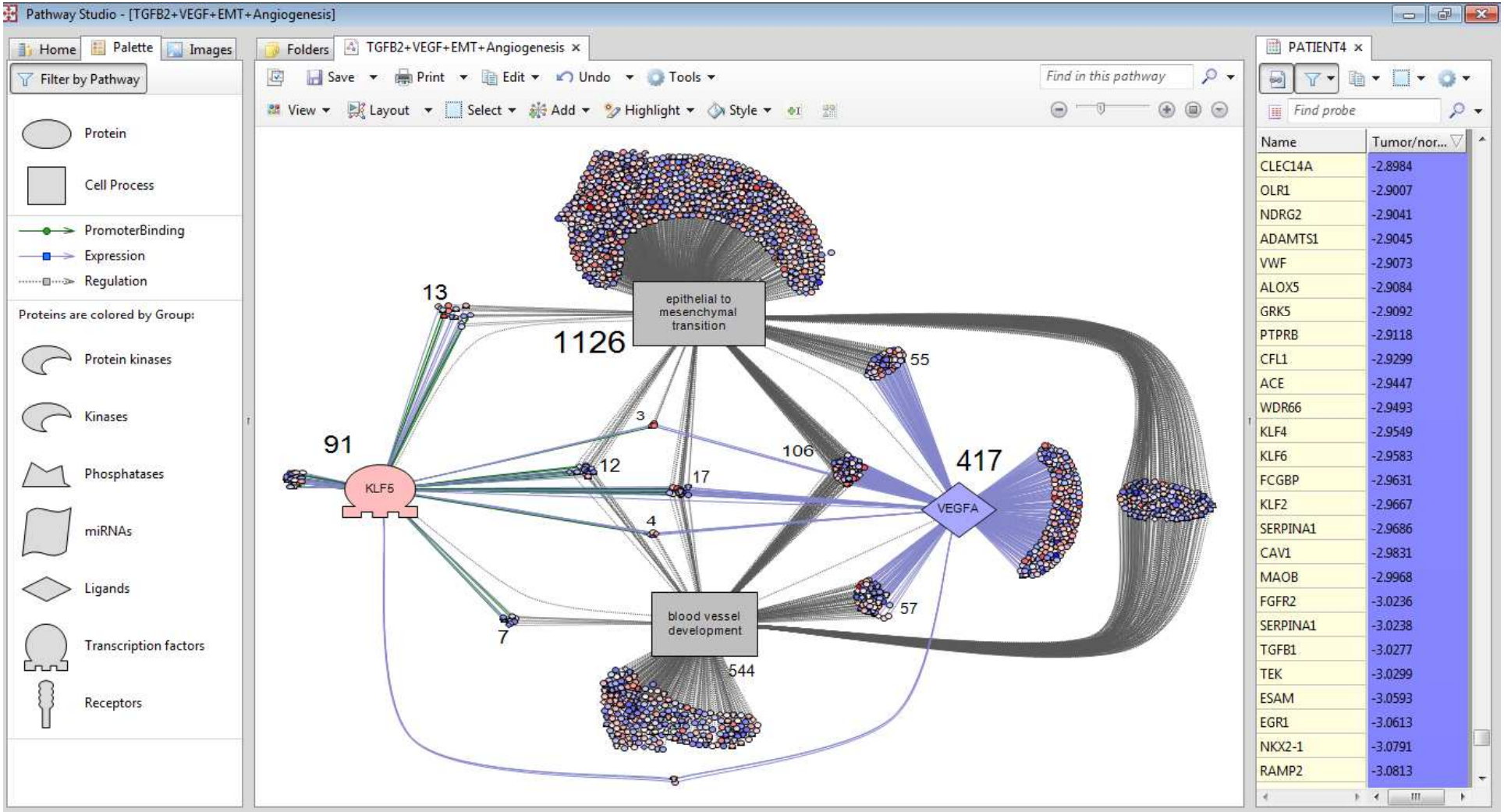
Promoter Binding: Protein TF->Protein



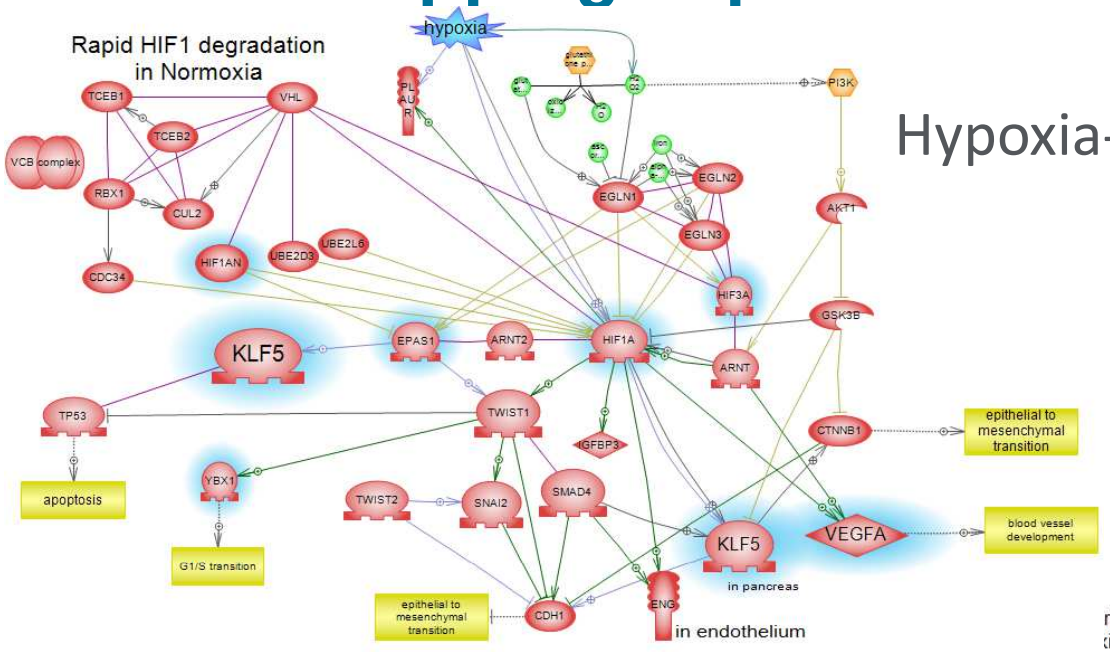
477,365

Expression: Protein->Protein

Example of expression regulators and Cell processes identified by SNEA in lung cancer patient

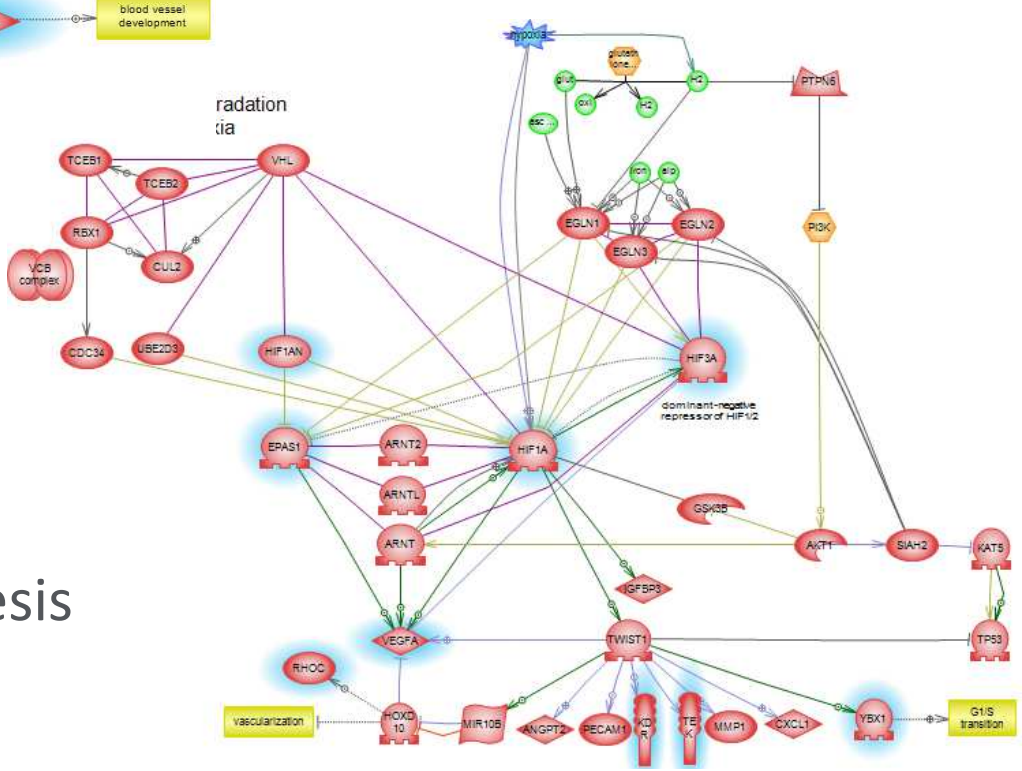


STEP2: Mapping expression regulators on pathways



Hypoxia->EMT

Blue highlight – expression regulators in Lung cancer patient identified by SNEA



Hypoxia->Angiogenesis

Personalized Hematology-Oncology of Wake Forest

5 cancer patients

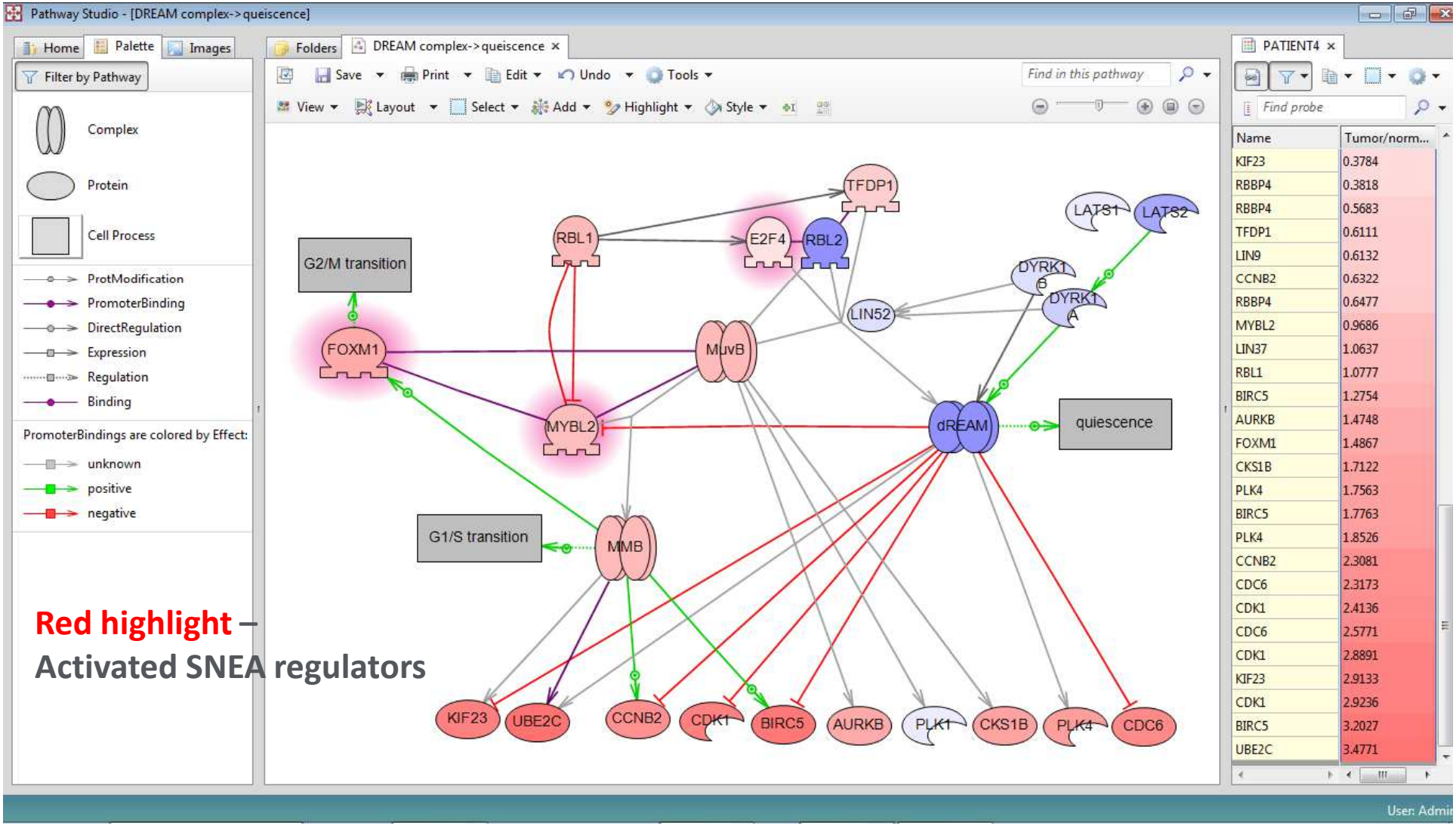
analyzed with SNEA to build cancer pathways

- Gallbladder/Liver cancer
- Lung cancer #1
- Lung cancer #2
- Breast cancer metastasis in lung
- Colon cancer metastasis in liver

How many cancer pathways must be built?

Upper estimate: 10 hallmarks X 250 tissues = 2,500

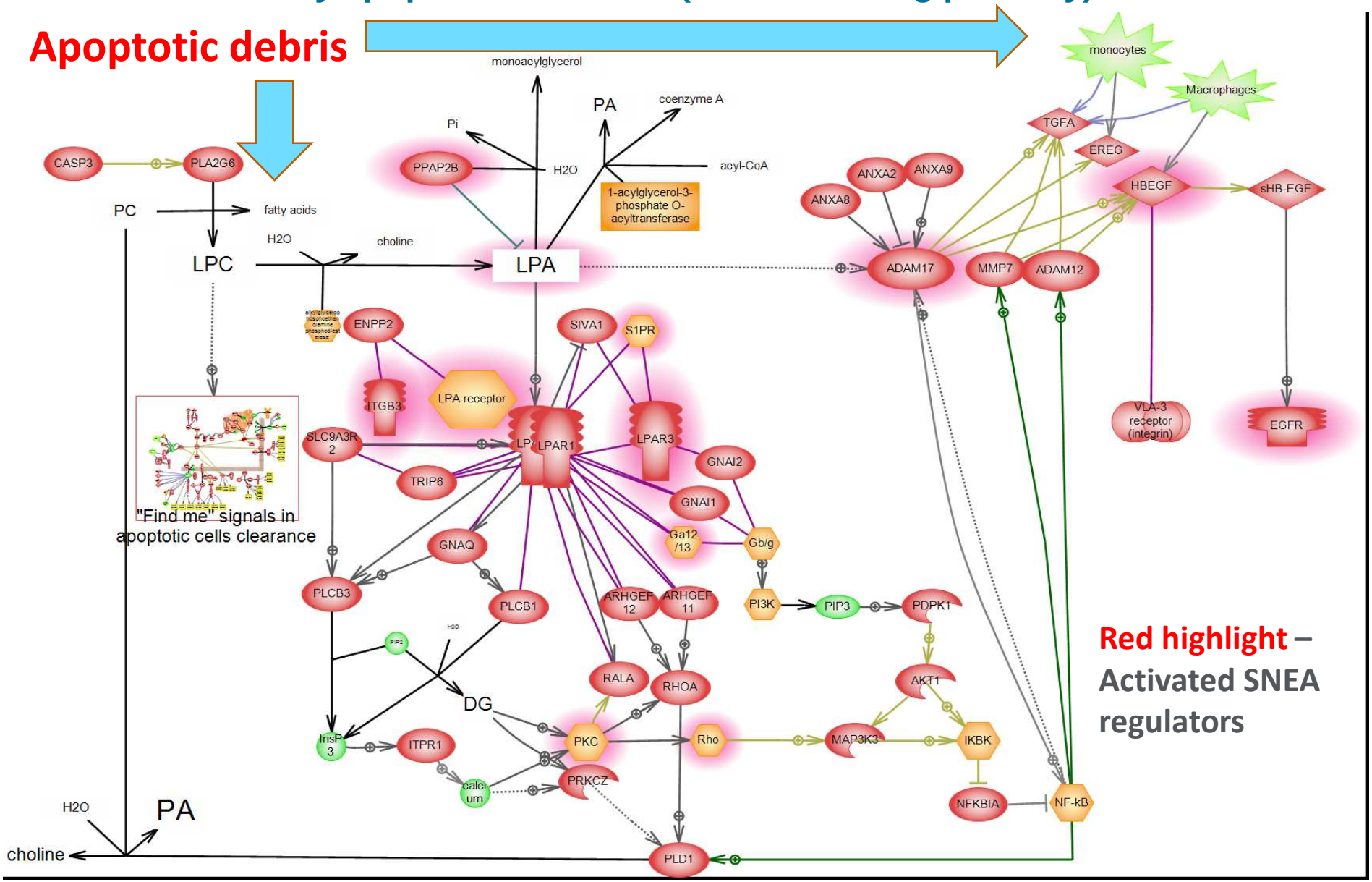
In practice some pathways may be common for all tissues. Example: Cell cycle pathways



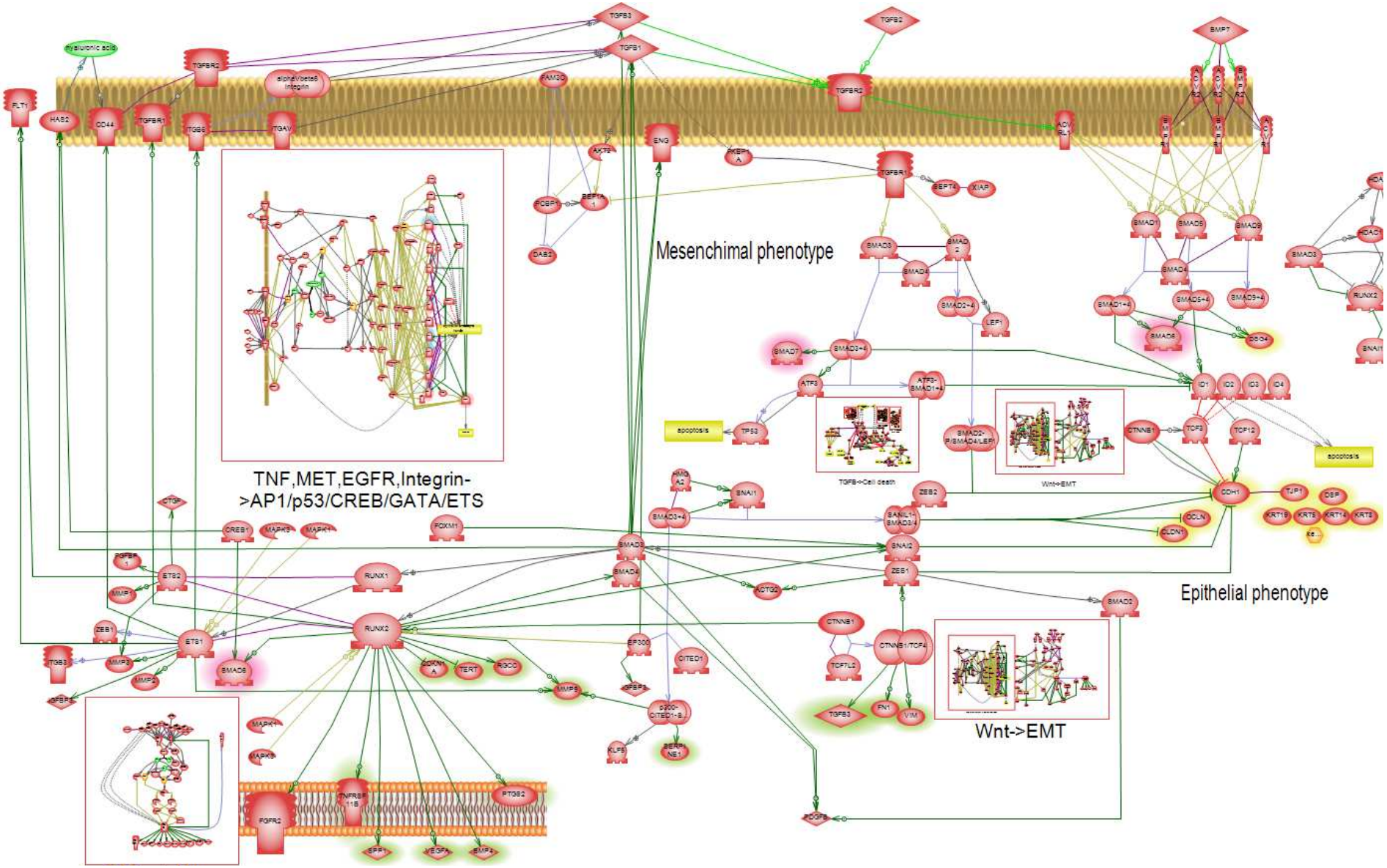
Red highlight –
Activated SNEA regulators

Cancer pathways: Insights to cancer biology

EGFR activation by apoptotic clearance (wound healing pathway)

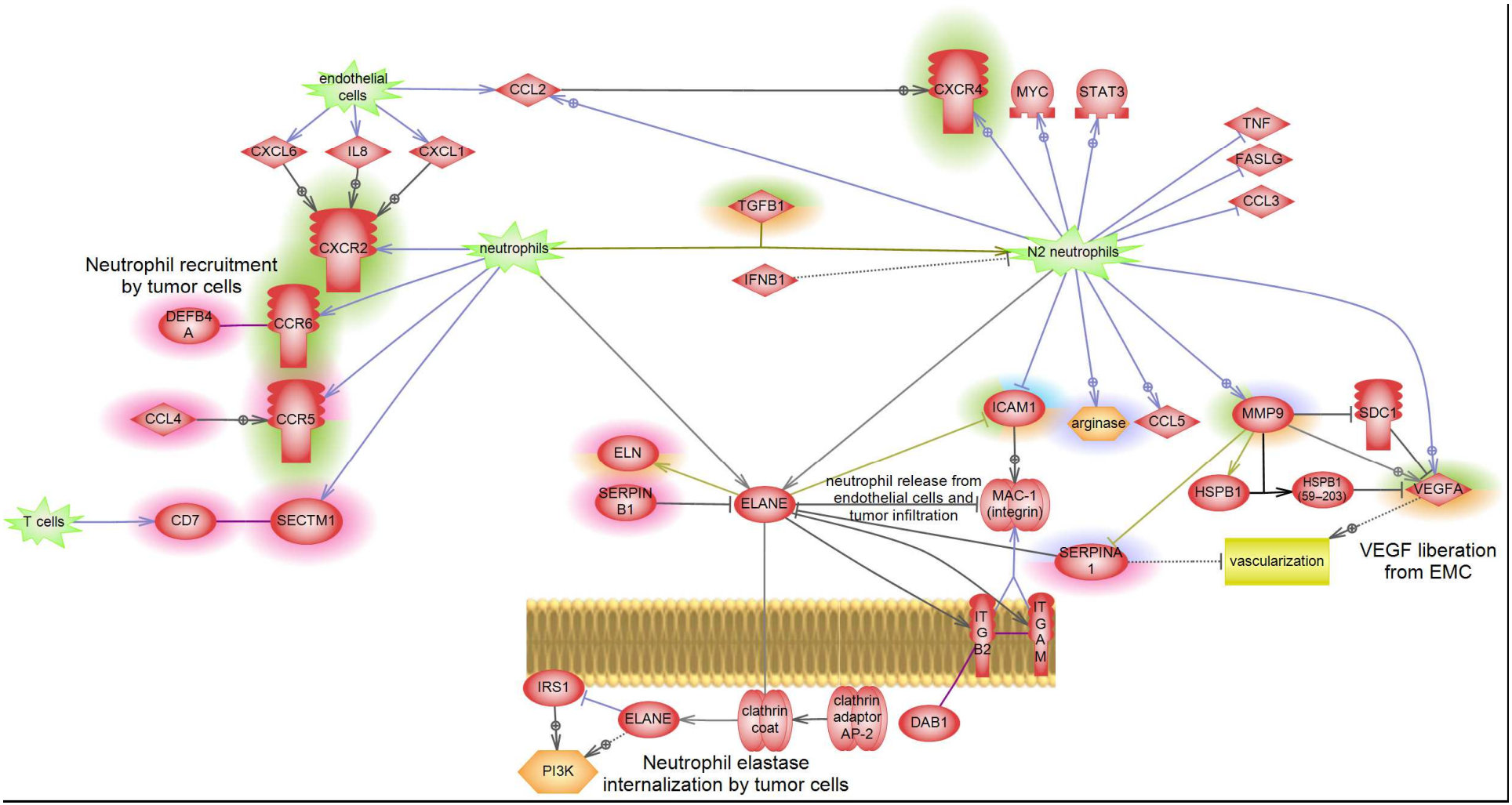


TGF-β autocrine loop sustains EMT

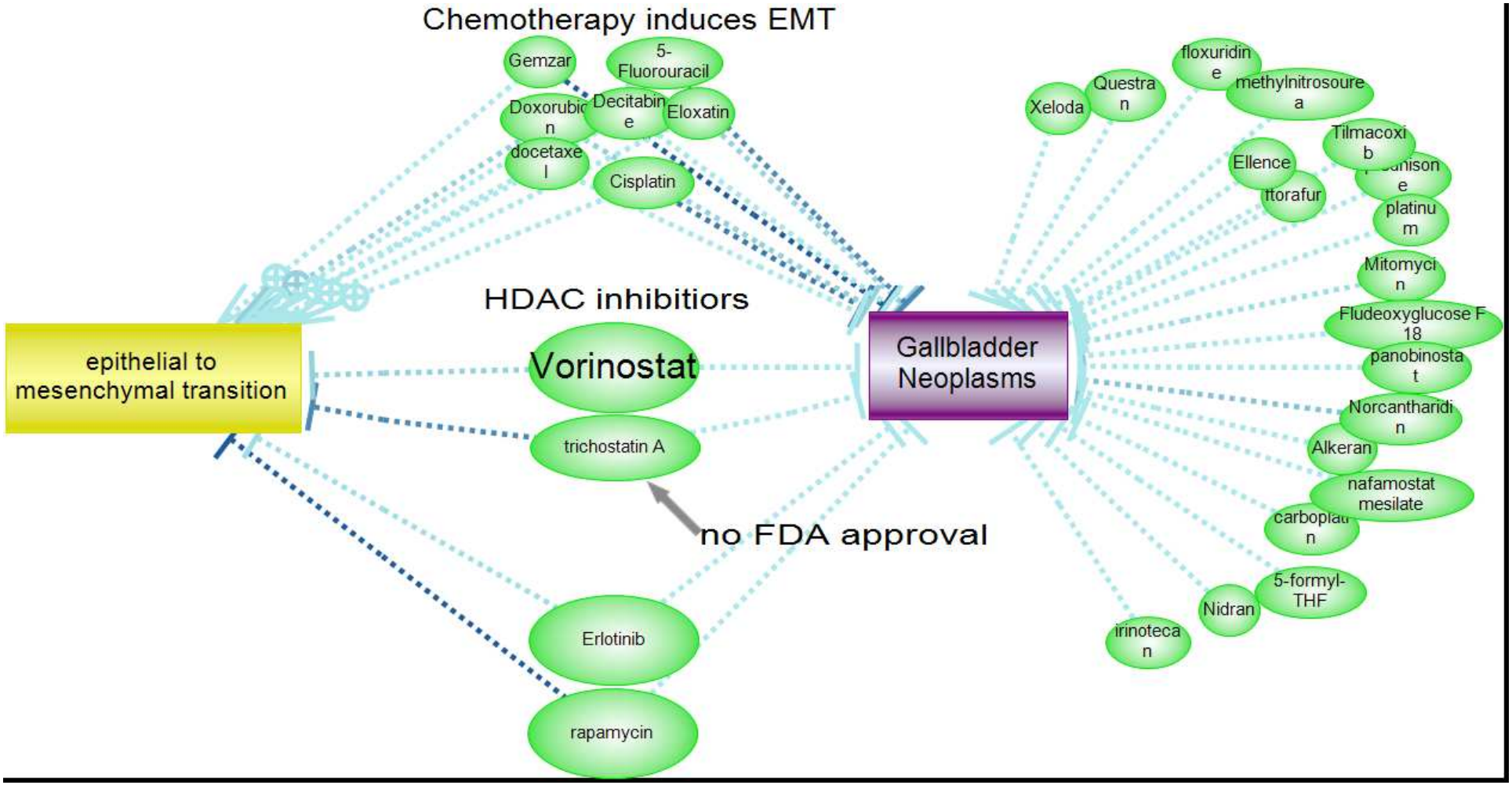


Avoiding immune response: N1->N2 polarization

Highlights – SNEA regulators from different patients



How to select anti-cancer drugs in Pathway Studio



ELSEVIER

Conclusions:

- *48 pathways containing 2,796 proteins provide mechanism for advanced cancer in 5 patients*
- *Pathways explain about 50% (378) of all top 100 SNEA regulators indentified in five patients*
- *Pathways are validated by:*
  - *Scientific literature*
  - *Patient microarray data*
  - *Efficacy of personalized therapy*