V12: gene-regulatory networks related to cancerogenesis

TCGA breast cancer study

Towards a breast cancer GRN ...

Motifs in GRNs ... TMmiR

The Cancer Genome Atlas (TCGA): breast cancer

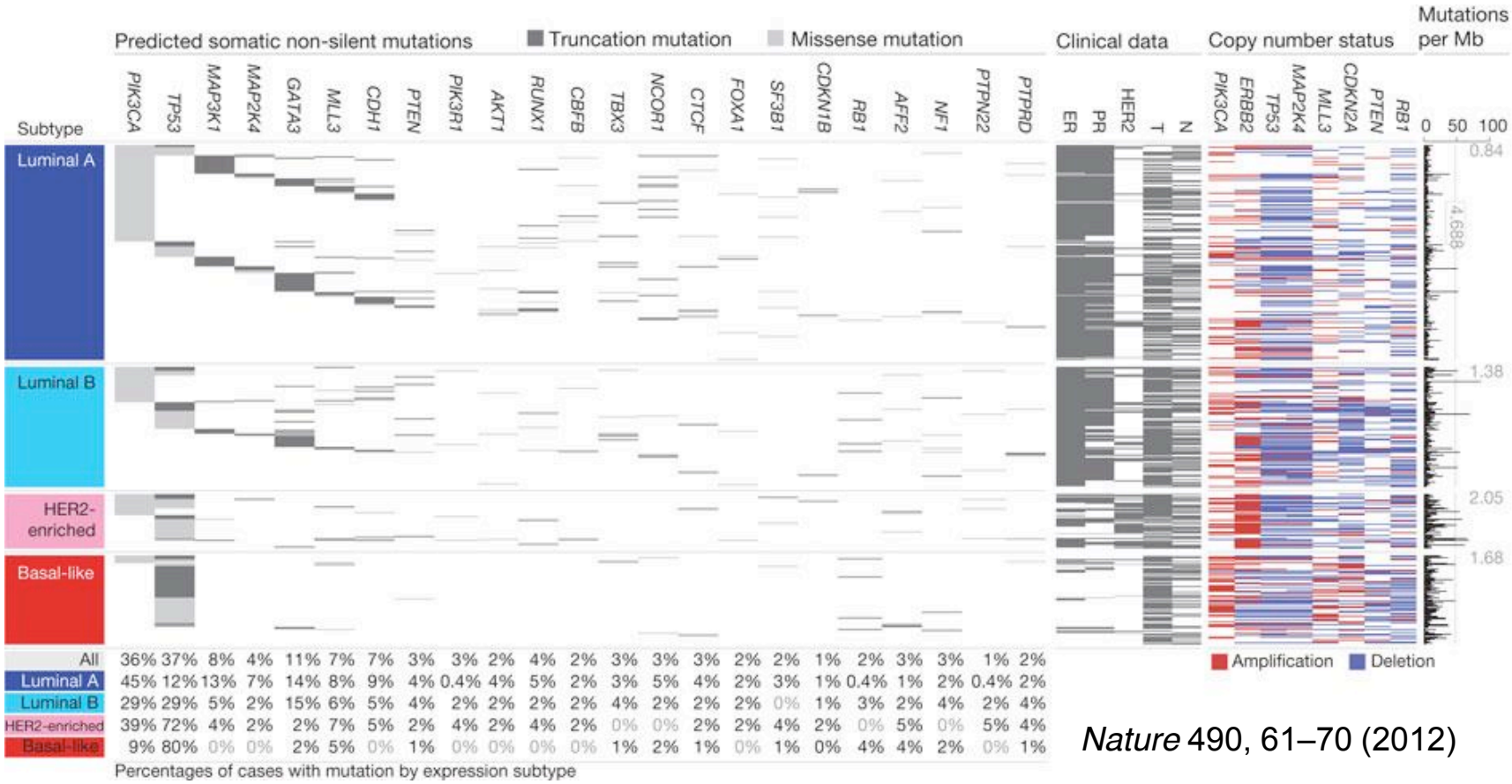
TCGA consortium analysed primary breast cancers by
- genomic DNA copy number arrays,
- DNA methylation,
- exome sequencing,
- messenger RNA arrays,
- microRNA sequencing and
- reverse-phase protein arrays.

Combining data from 5 platforms showed the existence of 4 main breast cancer classes. Each of them shows significant molecular heterogeneity.

Somatic mutations in only 3 genes (*TP53, PIK3CA and GATA3)* occurred at >10% incidence across all breast cancers.

Breast cancer genomes in TCGA

Tumour samples are grouped by mRNA subtype: luminal A (*n = 225), luminal B (n = 126), HER2E (n = 57)* and basal-like (n = 93). Clinical features: dark grey, positive or T2–4; white, negative or T1; light grey, N/A or equivocal. N, node status; T, tumour size.
Right: significantly mutated genes with frequent copy number amplifications (red) or deletions (blue).
Far-right: non-silent mutation rate per tumour (mutations per megabase, adjusted for coverage).

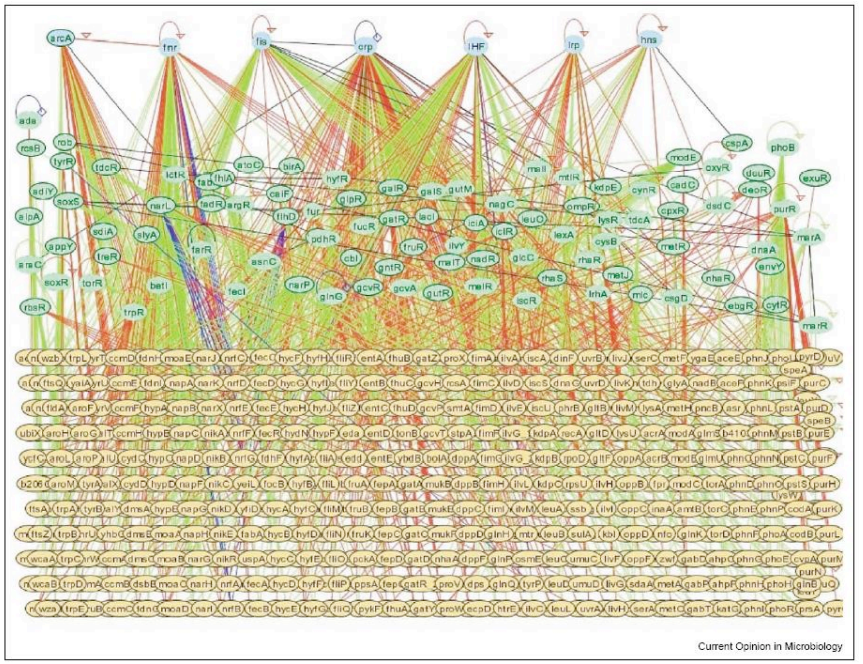


Nature 490, 61–70 (2012)

Review (bioinformatics III) – GRN of E. coli

`RegulonDB`: database with information on transcriptional regulation and operon organization in *E.coli*; 105 regulators affecting 749 genes

→ 7 regulatory proteins (CRP, FNR, IHF, FIS, ArcA, NarL and Lrp) are sufficient to directly modulate the expression of more than half of all *E.coli* genes.



Martinez-Antonio, Collado-Vides, Curr Opin Microbiol 6, 482 (2003)

Review (bioinformatics III) – Regulatory cascades in *E.coli*

When more than 1TF regulates a gene, the order of their binding sites is as given in the figure.

Arrowheads indicate positive regulation when the position of the binding site is known.

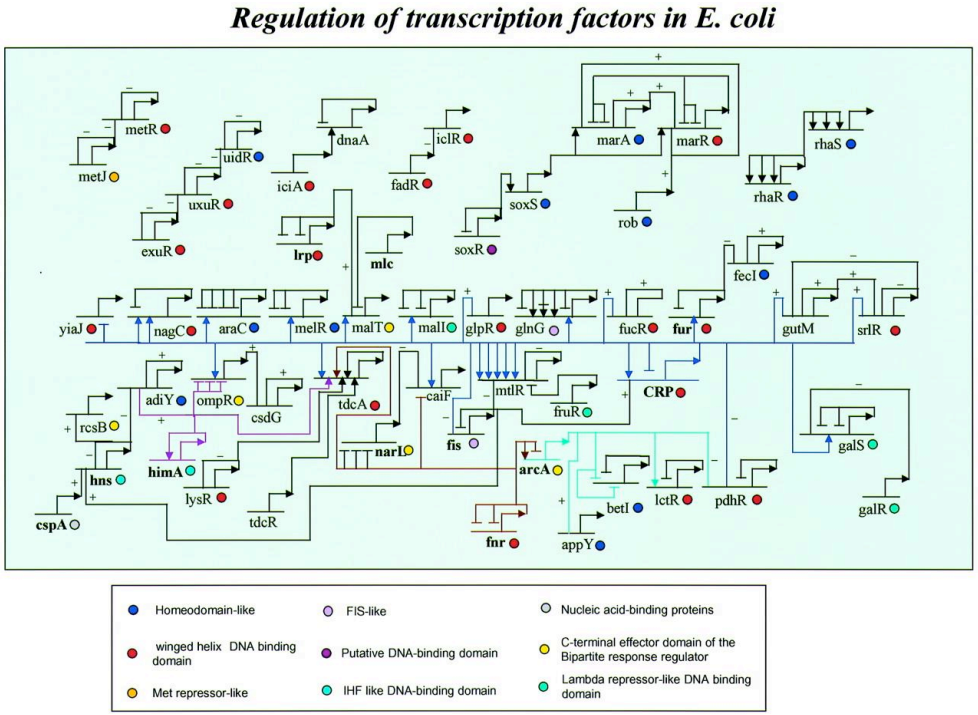
Horizontal bars indicates negative regulation when the position of the binding site is known.

In cases where only the nature of regulation is known, without binding site information, + and – are used to indicate positive and negative regulation.

The DNA binding domain families are indicated by circles of different colours

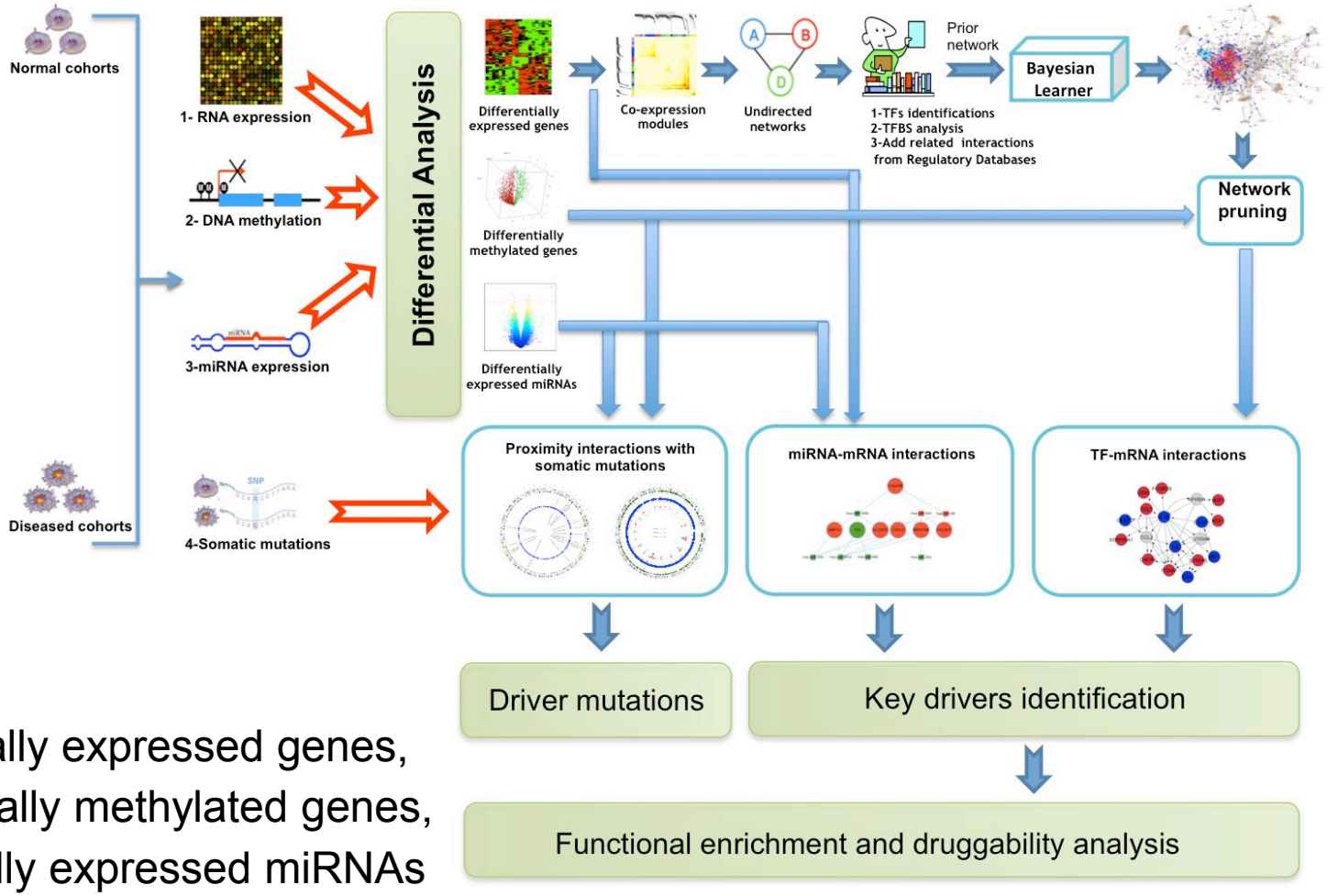
.

The names of global regulators are in bold.



Babu, Teichmann, Nucl. Acid Res. 31, 1234 (2003)

Aim: construct GRN for breast cancerogenesis



We found
-1317 differentially expressed genes,
- 2623 differentially methylated genes,
- 121 differentially expressed miRNAs between 131 tumor and 20 normal tissues.

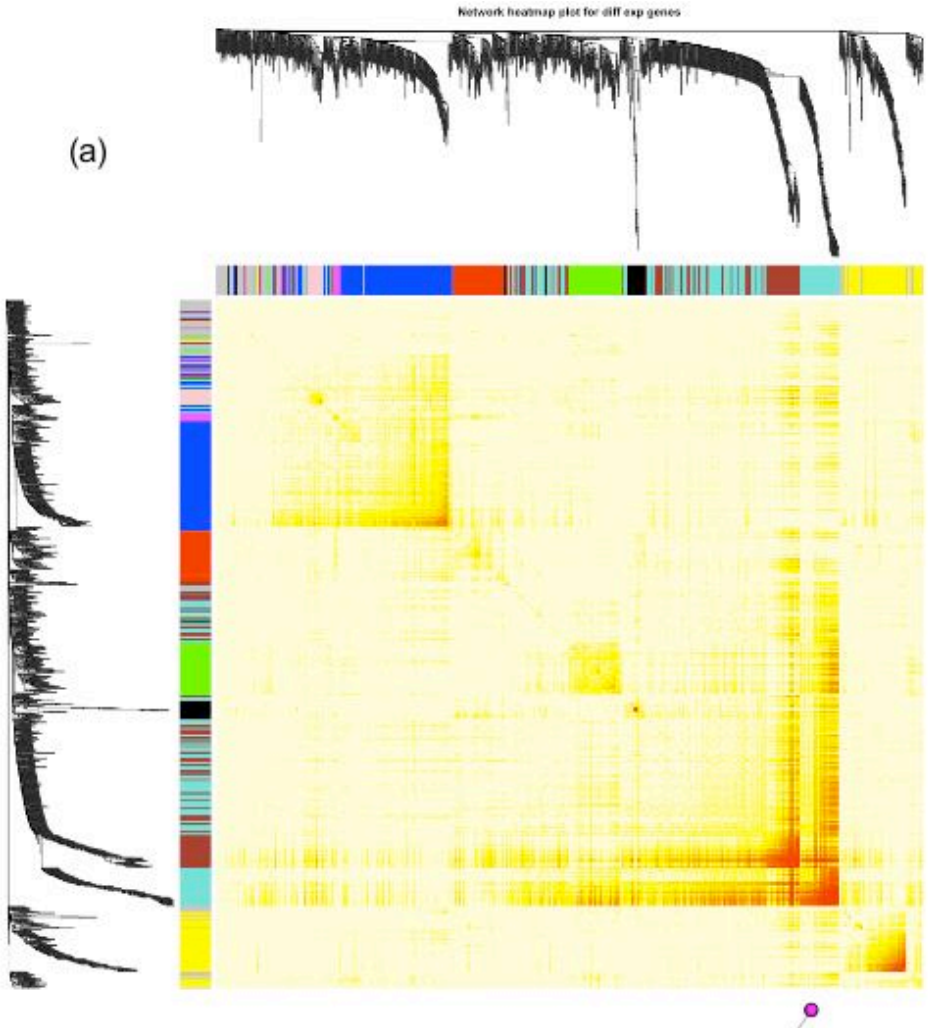
Hamed et al. *BMC Genomics* 16, S2 (2015)

Organize genes into modules

The expression profiles of the 1317 identified differentially expressed genes were used to compute the co-regulation strength between genes.

An undirected co-expression network was obtained by hierarchical clustering (HCL).

HCL yielded 10 segregated network modules that contain between 26 and 295 gene members (different colors).



Organize genes into modules

Table 1. The key driver elements identified TF-gene interactions and miRNA-mRNA interactions

	Module	Gene count	Top GO category	Top KEGG categories	Key driver count	Key drivers
TF-mRNA interactions	black	41	Regulation of transcription	Pathways in cancer, Renal cell carcinoma	5	SORBS3, ZNF43, ZNF681, RBMX, POU2F1
	blue	247	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	Cell cycle, Prostate cancer, Melanoma	9	**AR, BRCA1, ESR1, JUN, MYB**, RPN1, E2F1, E2F2, PPARD
	brown	195	Anatomical structure morphogenesis	Leukocyte transendothelial migration	5	TMOD3, CREB1, POU5F1, SP3, TERT
	green	110	Cellular macromolecule metabolic process	Endometrial cancer, Insulin signaling pathway	15	**B4GALT7, OS9, CDC34**, MAN2C1, MYO1C, SH3GLB2, INPP5E, PLXNB1, USF2, PPP1R12C, CDK9, DAP, E4F1, E2F4, USF1
	grey	148	Anatomical structure development	Sulfur metabolism	18	**AHCTF1, NQO2, FGFR2, CCDC130, ABCG4, BIRC6, CA6**, SP4, RNF2, SPRR1B, C16orf65, DNAJC5G, SNCAIP, GRIK5, SLC6A4, SMAD1, DAD1, POU4F2
	magenta	26	Regulation of metabolic process	p53 signaling pathway, Alzheimer's disease	3	**ATF6**, NGEF, POGK
	pink	30	Transcription initiation from RNA polymerase II promoter	Basal transcription factors	4	**CCDC92**, TMEM70, RNF139, E2F5
	red	93	Regulation of cellular process	Endometrial cancer, Neurotrophin signaling pathway	14	**ATP1B1, STAT3, ABCB8, MYC, TGFB1, SP1, TP53**, PCGF1, SUMF2, GTF3A, IPO13, GMPPA, HTR6, TGIF1
	turquoise	295	Regulation of cellular metabolic process	p53 signaling pathway, Pancreatic cancer, Apoptosis	2	UBL5, RNF111
	yellow	132	Immune system process	Chemokine signaling pathway, Natural killer cell mediated cytotoxicity	19	**APOC1, CD2, CD79B, LRRC28, DAPK1**, FAM124B, EML2, LAP3, TSPAN2, FCRL3, ELMO1, SLC7A7, RASSF5, SLC31A2, TRAF3IP3, GALNT12, ITGA4, SPI1, TFAP2A
	Total	1317				

Enriched GO categories and KEGG pathways determined among genes in module by hypergeometric test ($p < 0.05$).

Hamed et al. *BMC Genomics* 16, S2 (2015)

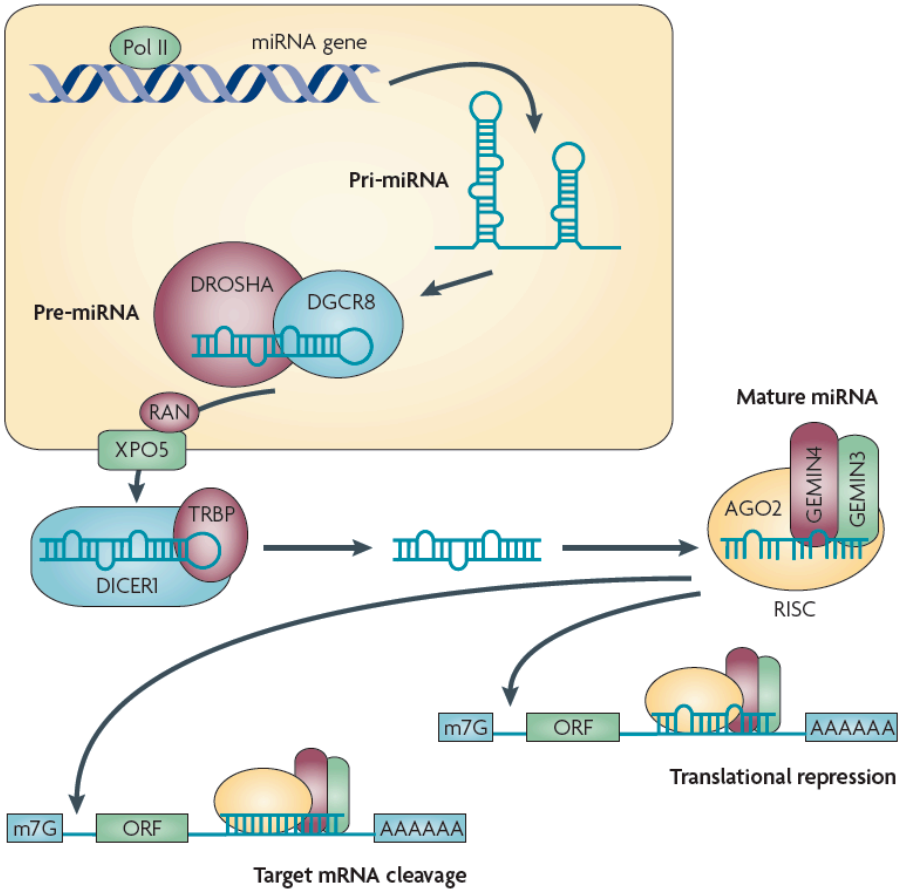
Key driver genes

Key regulators (**drivers**) in the constructed modules were identified by determining the **minimal set of nodes** that regulate the entire module.

These nodes typically include the nodes with **highest degree** plus some further ones that are required to „reach“ the remaining target genes.

Hamed et al. *BMC Genomics* 16, S2 (2015)

Functions of miRNAs



Single stranded miRNAs are incorporated into the RISC complex.

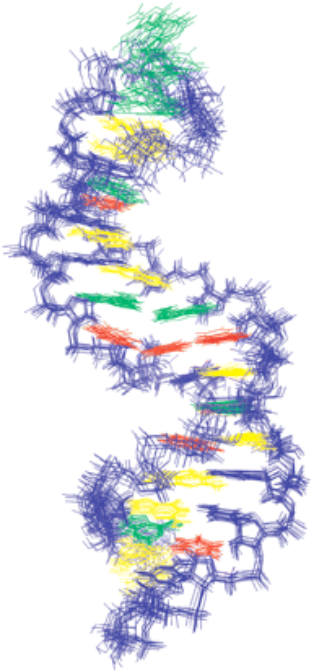
This complex then targets the miRNA e.g. to the target 3′ untranslated region of a mRNA sequence to facilitate repression and cleavage.

AA, poly A tail;
m7G, 7-methylguanosine cap;
ORF, open reading frame.

Ryan et al. Nature Rev. Cancer (2010) 10, 389

Binding partners of miRNAs

Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules.



solution NMR-structure of *let-7* miRNA:*lin-41* mRNA complex from *C. elegans*
Cevec et al. *Nucl. Acids Res. (2008) 36: 2330.*

The main function of miRNAs is to down-regulate translation of their target mRNAs.

miRNAs typically have incomplete base pairing to a target and inhibit the translation of many different mRNAs with similar sequences.

In contrast, siRNAs typically base-pair perfectly and induce mRNA cleavage only in a single, specific target.

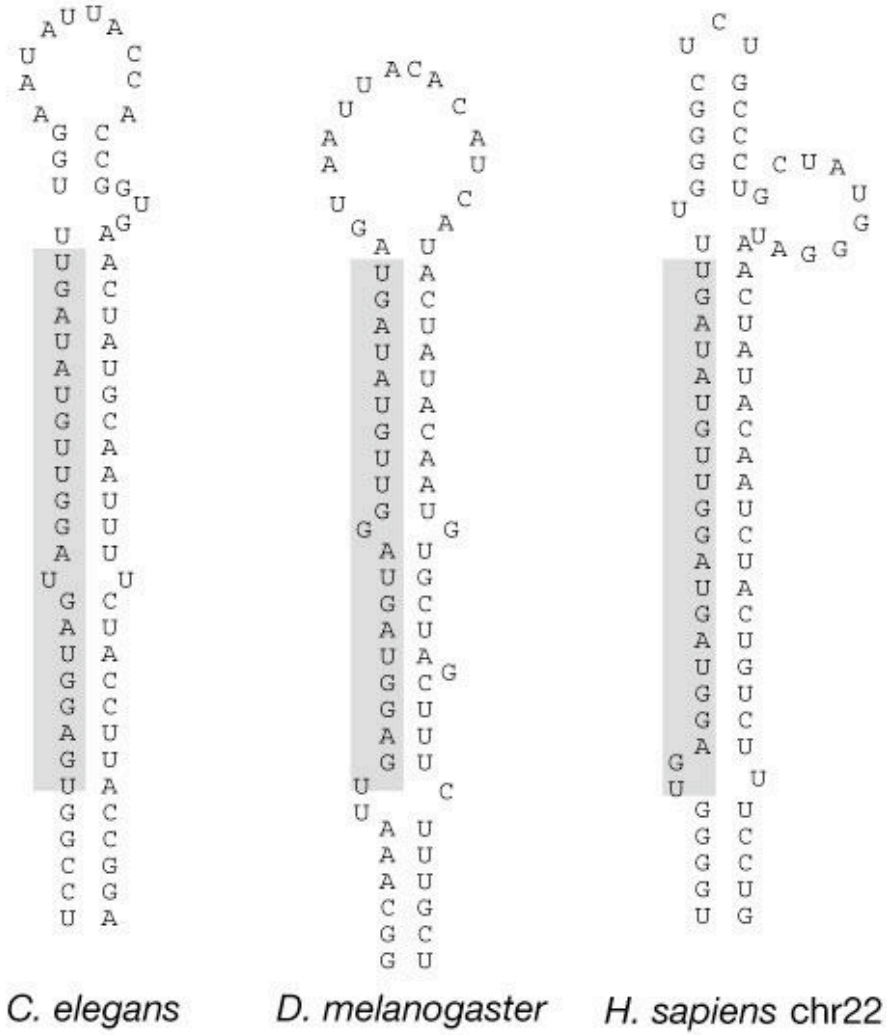
www.wikipedia.org

discovery of let7

The first two known microRNAs, lin-4 and let-7, were originally discovered in the nematode *C. elegans.*

They control the timing of stem-cell division and differentiation.
let-7 was subsequently found as the first known human miRNA.

let-7 and its family members are **highly conserved** across species in sequence and function.
Misregulation of let-7 leads to a less differentiated cellular state and the development of cell-based diseases such as cancer.



Pasquinelli et al. Nature (2000) 408, 86
www.wikipedia.org

Construct regulatory interactions

* For the 7 smallest modules, we collected the related directed **regulatory interactions** available in 3 online regulatory databases (JASPAR, TRED, MsigDB).

These were used as a prior for a Bayesian learner to learn the causal probabilistic regulatory interactions and to generate a directed network topology.

* We removed 89 inferred interactions whose target genes are downregulated and their expression profiles showed absolute **anti-correlation** measure > 0.65 with their **methylation profiles**.

In those cases we reasoned that downregulation of these target genes was most likely due to their **promoter methylation** and not due to TF binding

Hamed et al. *BMC Genomics* 16, S2 (2015)

Construct regulatory interactions involving miRNAs

* For the set of differentially expressed miRNAs, which were either up- or down-regulated between the tumor and normal samples, we used **miRTrail** via MicroCosm Targets V5 to extract their target mRNAs (regulated genes) and overlapped them with the identified differentially expressed mRNAs.

* We used the experimentally validated database **TransmiR** to retrieve the regulatory genes (**TFs**) that potentially regulate the differentially expressed **miRNAs**.

Hamed et al. *BMC Genomics* 16, S2 (2015)

miRNA-mRNA interactions in breast cancer network

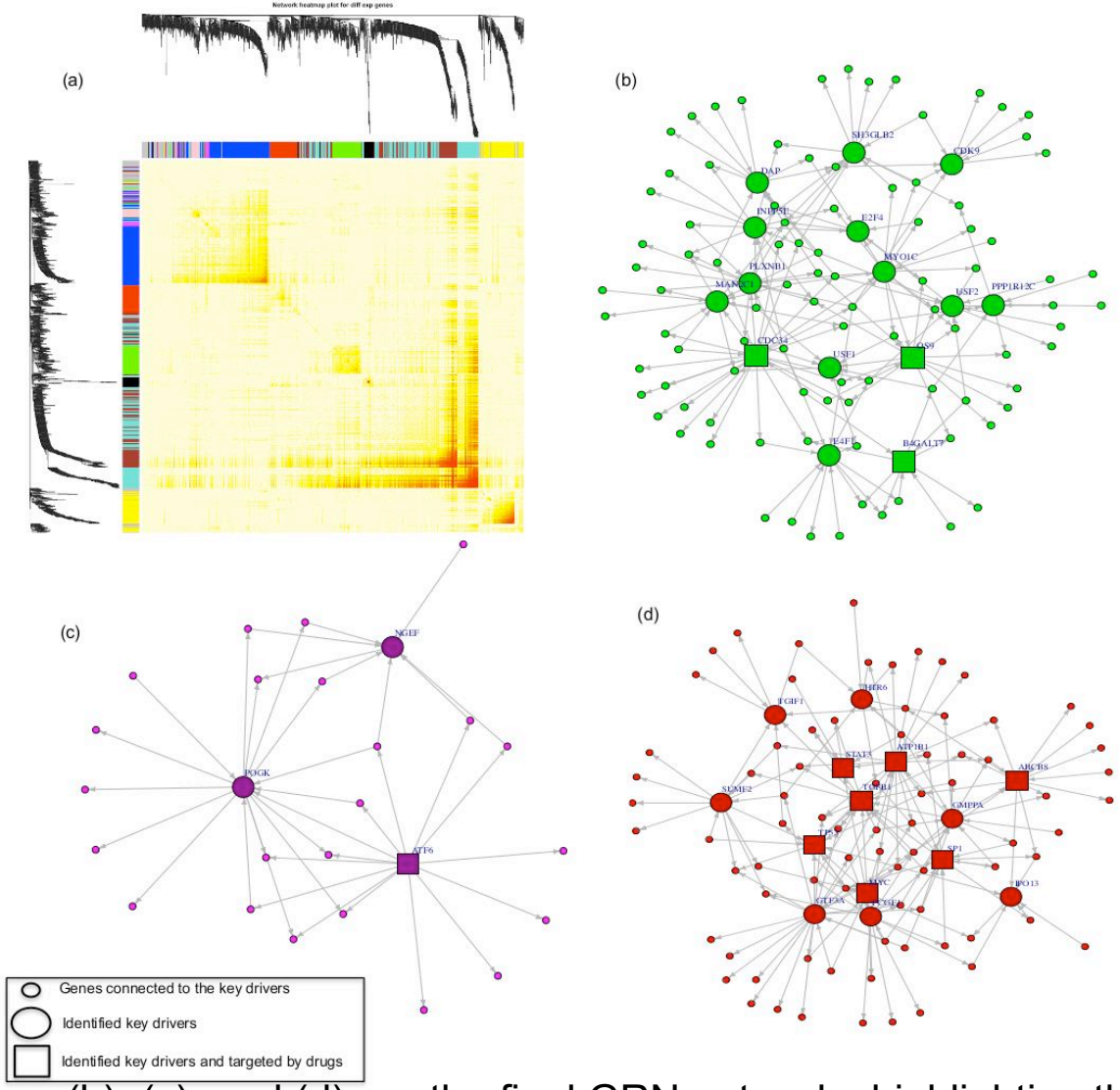
miRNA-mRNA interactions	Genes	Gene count	Top GO category	Top KEGG categories	Key driver count	Key drivers
		869	Regulation of macromolecule metabolic process	Pathways in cancer, Pancreatic cancer, Prostate cancer	17	**MYC, ATG4C, TGFB1, NFKB1, AKT1, EGR1, TP53**, SOX10, SPI1, MECP2, E2F3, CREB1, TCF3, TPP1, FLICE, LPS, PACS1
	miRNAs	**miRNA count**	**Top functional categories**	**Top HMDD categories**	**Key driver count**	**Key drivers**
		120	miRNA tumor suppressors, immune response, Onco-miRNA, cell death, human embryonic stem cells regulation	Breast cancer (65), Neoplasms (58), Melanoma (56), Ovarian Neoplasms (51), Pancreatic Neoplasms (38), Prostatic Neoplasms (38)	68	mir-126, mir-609, mir-488, mir-191, mir-200c, mir-200a, mir-30a, mir-30d, mir-335, mir-190b, mir-223, mir-106b, mir-519e, mir-210, mir-379, mir-203, mir-205, mir-708, mir-29c, mir-29a, mir-182, mir-183, mir-127, mir-187, mir-425, let-7g, let-7d, mir-152, mir-155, mir-21, mir-22, mir-758, mir-921, mir-922, mir-375, mir-377, mir-181a-2, mir-657, mir-302d, mir-100, mir-10b, mir-10a, mir-625, mir-629, mir-92a-2, mir-26b, mir-25, mir-145, mir-143, mir-141, mir-221, mir-193b, mir-193a, mir-374a, mir-134, mir-146a, mir-31, let-7a-2, mir-27a, mir-27b, mir-133a-1, let-7i, mir-93, mir-23a, mir-148a, mir-196a-2, mir-487b, mir-149

For the 10 gene modules identified in TF-mRNA interactions, we list counts of the involved genes, the most significant GO and KEGG terms, and the identified key driver genes from each module. Similarly for the miRNA-mRNA interactions, we list the key driver molecules of both genes and miRNAs. The driver genes, whose protein products are known to be targeted by drugs, are in bold.

Hamed et al. *BMC Genomics* 16, S2 (2015)

GRN modules

Gene network modules of TF-gene interactions.
(a) Topological overlap matrix (TOM) heatmap corresponding to the 10 co-expression modules.
Each row and column of the heatmap represent a single gene. Spots with bright colors denote weak interaction whereas darker colors denote strong interaction. Dendrograms on the upper and left sides show the hierarchical clustering tree of genes.



(b), (c), and (d) are the final GRN networks highlighting the identified key drivers genes for the green, magenta, and red modules, respectively. Square nodes : identified driver genes that are targeted by drugs.

Hamed et al. *BMC Genomics* 16, S2 (2015)

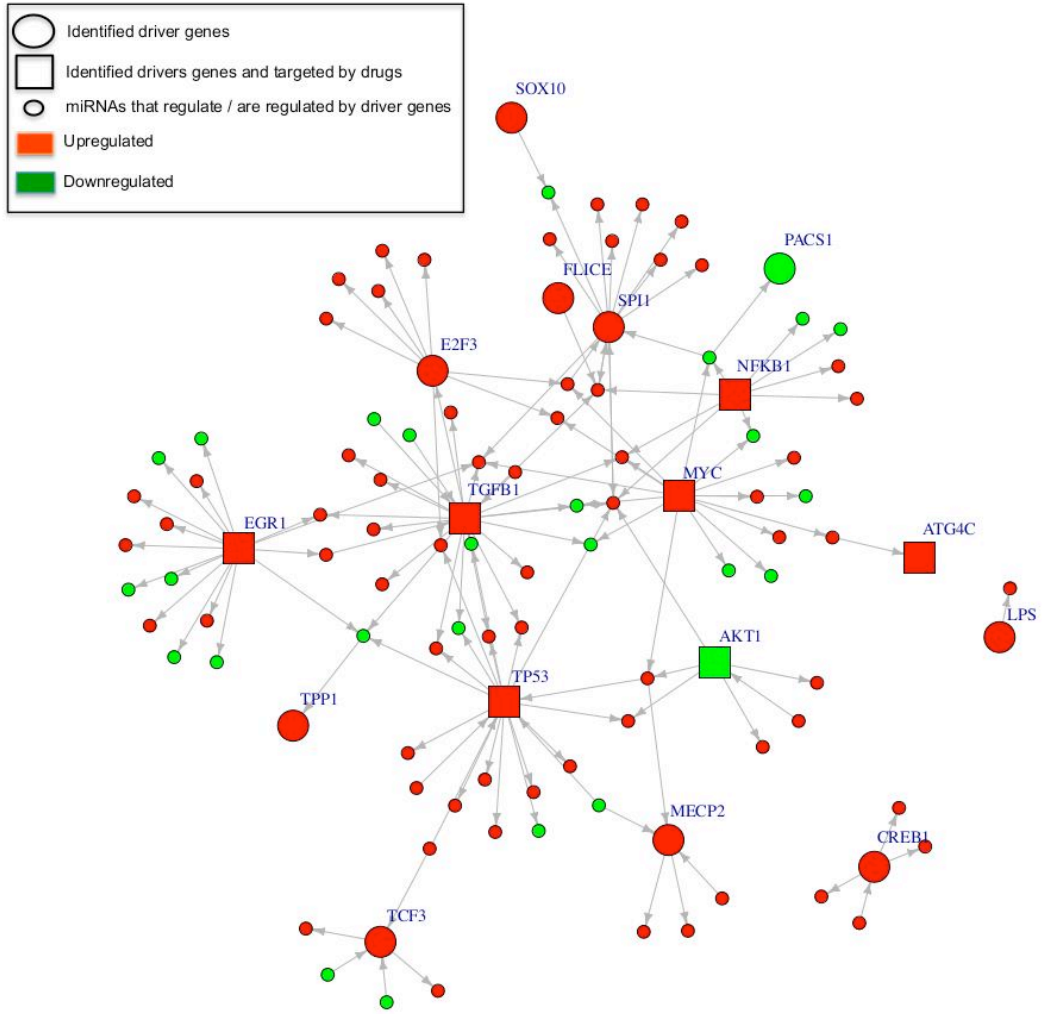
Regulatory interactions of driver genes

Regulatory interactions of the 17 key driver genes identified from miRNA-mRNA interactions.

Large nodes: key driver genes

small nodes : miRNAs, which regulate or are regulated by these driver genes.

Square nodes : driver genes that are targeted by available drug molecules.



Hamed et al. *BMC Genomics* 16, S2 (2015)

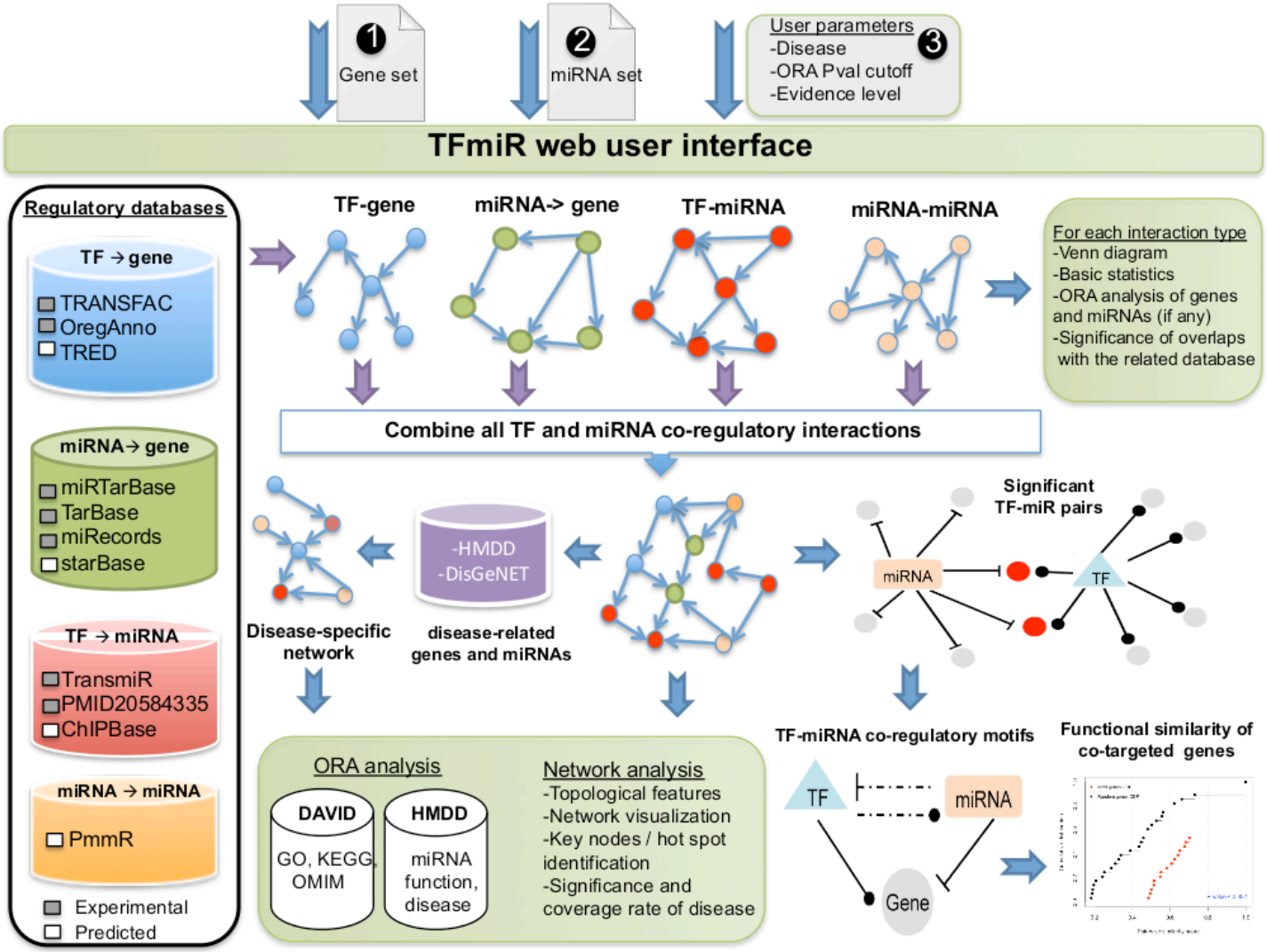
Target gene	Drug and antineoplastic agents	CTD	PharmGKB	Cancer Resource
ABCB8	docetaxel; Cyclosporine; Progesterone	1	0	0
ABCG4	indole-3 carbinol; Methotrexate; exemestane; Vincristine	1	0	0
AHCTF1	Methotrexate; bisphenol A	1	0	0
AKT1	U 0126;tyrphostin AG 1478; Ursodeoxycholic Acid;Valproic Acid;tyrphostin AG 1024; trametinib; Tretinoin	1	0	1
APOC1	tanshinone; Quercetin; Fluorouracil; bexarotene; Cisplatin; Tamoxifen	1	0	1
AR	Dihydrotestosterone; Acetylcysteine; celecoxib	1	0	0
ATF6	Nelfinavir; Tretinoin;bisphenol A; Cyclosporine; Curcumin	1	0	0
ATG4C	epigallocatechin gallate	1	0	0
ATP1B1	resveratrol; Ranitidine; vorinostat; Genistein; Progesterone; epigallocatechin gallate	1	0	0
B4GALT7	Cytarabine; Cyclosporine	1	0	0
BIRC6	Dieldrin; Cyclosporine; Cisplatin; Fluorouracil; Doxorubicin; Epirubicin;Estradiol; zoledronic acid; bisphenol A	1	0	0
BRCA1	Tretinoin; trichostatin A; Estradiol; transplatin; troglitazone; Tunicamycin; fulvestrant	1	0	1
CA6	Tretinoin;Carmustine	1	0	0
CCDC130	Quercetin;Tamoxifen;Cyclosporine;bisphenol A	1	0	0
CCDC92	Quercetin; Folic Acid	1	0	0
CD2	Dexamethasone; Methotrexate; Cyclophosphamide	1	0	0
CD79B	Cyclophosphamide	1	0	0
CDC34	Estradiol; bortezomib; Fluorouracil; Tamoxifen	1	0	0
DAPK1	paclitaxel;gemcitabine	0	1	0
EGR1	Fluorouracil; gemcitabine	0	0	1
ESR1	exemestane;tamoxifen	0	1	1
JUN	andrographolide; cinnamic aldehyde; Daunorubicin; decitabine; Cisplatin;Doxorubicin	0	0	1
LRRC28	gemcitabine	0	0	1
MYB	Fluorouracil;gemcitabine;Quercetin	0	0	1
MYC	alitretionoin; Amsarcine; bicalutamide; Camtothecin; decitabine; Cisplatin; Doxorubicin	0	0	1
NFKB1	Curcumin; decitabine; Doorubicin; Echinomycin; Fluorouracil; gefitinib; indole-3-carbinol; parthenolide	0	0	1
NQO2	doxorubicin; cyclophosphamide	0	1	0
OS9	alitretionoin	0	0	1
SP1	Etoposide; indole-3-carbinol; Ionidamine; Quercetin; Adaphostin	0	0	1
STAT3	azaspirane; bisphenol A; Capsaicin; Fluorouracil; interferon alfacon-1; resveratrol;sulindac sulfide; Tamoxifen	0	0	1
TGFB1	Doxorubicin; Fluorouracil; Thalidomide; Entinostat; Hyaluronidase	0	0	1
TP53	4-biphenylmine; alliin; Apigenin; Atropine;bicalutamide;butylidenephthalide	0	0	1

We identified 94 driver genes from the TF-mRNA interactions and 17 driver genes from the miRNA-mRNA interactions. 5 breast cancer associated genes *CREB1, MYC, TGFB1, TP53*, and *SPI1* were common in both sets -> in total 106 driver genes.

31% (33 proteins) of the proteins belonging to the identified driver genes are binding targets of at least one anti-breast cancer drug

-> validates our approach

TFmiR: identify regulatory motifs in GRNs



Hamed et al. *Nucl Acids Res* 43: W283-W288 (2015)

Data sources used for TFmiR

Table S1. The integrated databases and interaction types in TFmiR.

Interaction	Databases (P/E) *	Genes	miRNAs	Regulatory links	Version /frozen date
TF→gene	TRANSFAC (E) (1)	1279	--	2943	V11.4
	OregAnno (E)(2)	1132	--	1083	Nov 2010
	TRED (P) (3)	3038	--	6462	2007
TF→miRNA	TransmiR (E) (4)	158	175	567	V1.2, Jan 2013
	PMID20584335 (E) (5)	58	56	102	Apr 2009
	ChipBase (P) (6)	119	1380	33087	V1.1, Nov 2012
miRNA → gene	miRTarBase (E)(7)	2244	551	5640	V4.5, Nov 2013
	TarBase (E) (8)	422	79	492	V7.0
	miRecords (E)(9)	543	157	780	Mar 2009
	starBase (P)(10)	5720	249	56051	V2.0, Sept 2013
miRNA→miRNA	PmmR (P) (11)	--	312	3846	Mar 2011

* (P) means predicted interactions and (E) means experimentally validated interactions.

Hamed et al. *Nucl Acids Res* 43: W283-W288 (2015)

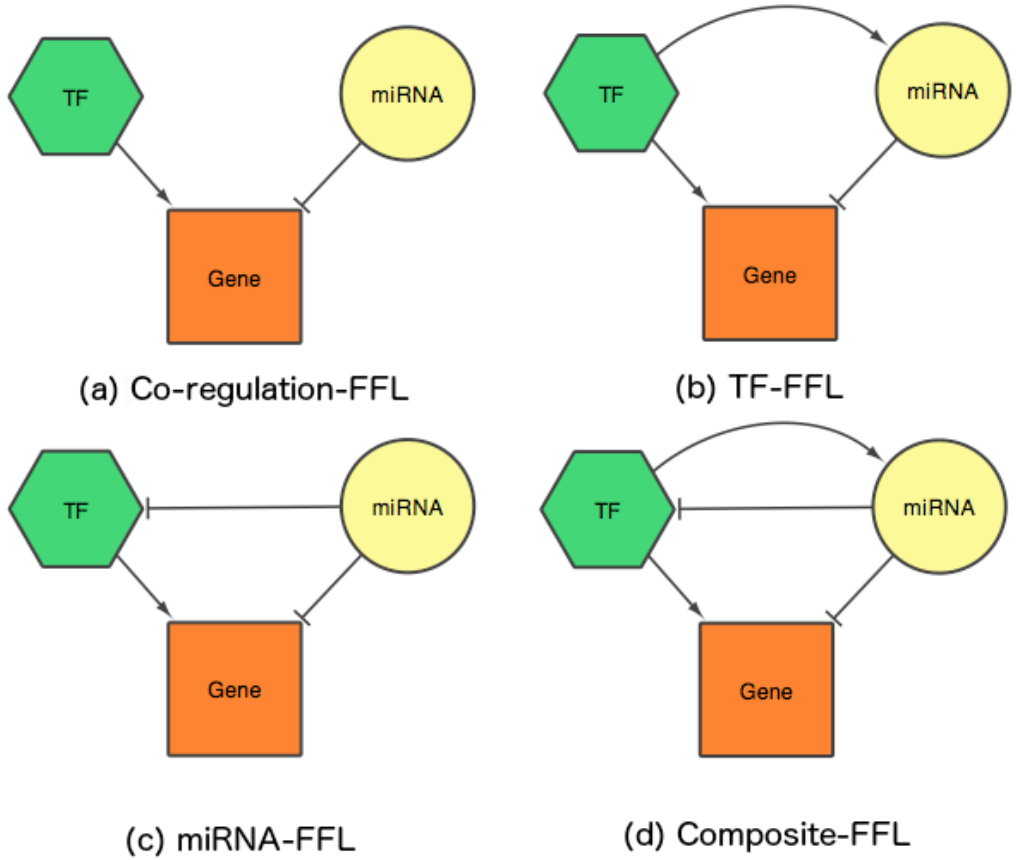
Possible co-regulatory motifs

We postulated that a TF and a miRNA may act occasionally on the same gene (or its transcribed mRNA).

In motif (a), a TF could first stimulate gene expression.

Later, the miRNA would degrade the transcript.

-> search generated networks for such co-regulatory motifs



Hamed et al. *Nucl Acids Res* 43: W283-W288 (2015)

Statistical significance of motifs

To evaluate the significance of each FFL motif type, we compare how often they appear in the real network to the number of times they appear in randomized ensembles preserving the same node degrees.

We applied a degree preserving randomization algorithm.

Each random network was generated by 2 * number of edges steps, where in each step we choose 2 edges $e_1 = (v_1, v_2)$ and $e_2 = (v_3, v_4)$ randomly from the network and swap their start and end nodes, i.e. $e_3 = (v_1, v_4)$ $e_4 = (v_3, v_2)$.

We construct 100 such random networks and count the motifs in them.

N_{motif_random} is the number of random networks that contain more than or equal numbers of a certain motif than the real network

The we compute the p-value for the motif significance as $p = N_{motif\_random} / 100$

Identified composite FFL motif

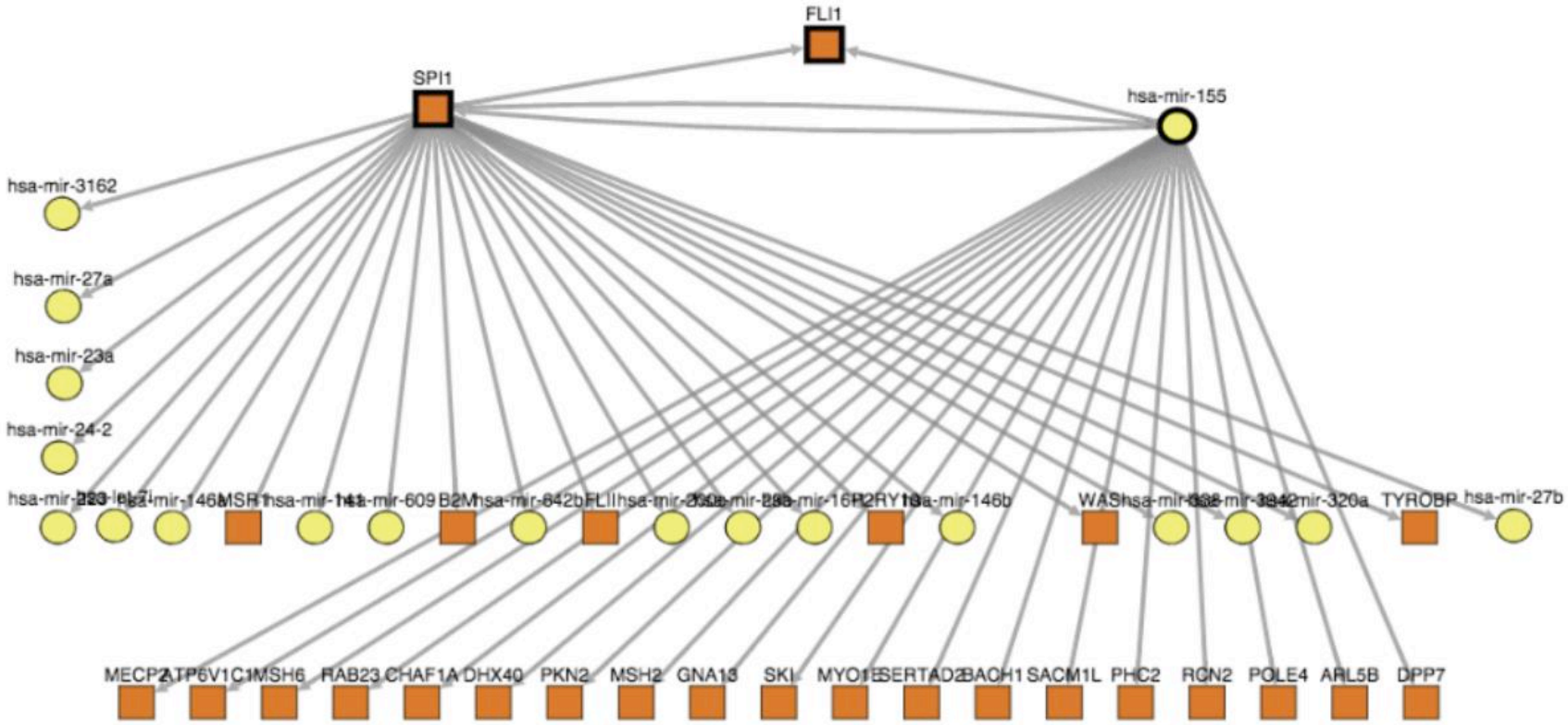


Figure S4. A composite FFL motif involves the TF *SPI1*, the miRNA *has-mir-155*, and the target gene *FLI1*. The co-regulated nodes are also visualized and are further tested whether they compose a cooperative functional module in breast cancerogenesis (see Fig S5).

Hamed et al. *Nucl Acids Res* 43: W283-W288 (2015)

Enriched biological functions in breast cancer network

Table S2. The most significant functions and diseases enriched in the miRNA nodes of the breast cancer disease network (12).

Category	Term	miRNAs Count	P-value
Function	Epithelial-mesenchymal transition	17	0.022
Function	glucose metabolism	4	0.048
Disease	Breast Neoplasms	67	1.43E-25
Disease	Lung Neoplasms	50	4.33E-17
Disease	Neoplasms	44	3.15E-15
Disease	Ovarian Neoplasms	43	1.30E-14
Disease	Adenocarcinoma	27	2.59E-13
Disease	Pancreatic Neoplasms	39	7.30E-13
Disease	Prostatic Neoplasms	41	3.49E-12
Disease	Melanoma	45	1.25E-11
Disease	Colonic Neoplasms	32	4.67E-11
Disease	Colorectal Neoplasms	45	5.69E-11

Hamed et al. *Nucl Acids Res* 43: W283-W288 (2015)

Drug Targets in breast cancer network

Table S4. The identified key gene nodes in the breast cancer network (12) whose protein products are targeted by anti-cancer drugs. (1) means that at least one drug that targets this gene product is reported in this database, and (0) means no drugs are reported for the respective gene in this database. Not included are substances that are known to be cancerogenous or mutagenic.

Target gene	Drug and antineoplastic agents	CTD	PharmGKB	Cancer Resource
AKT1	U 0126;tyrphostin AG 1478; Ursodeoxycholic Acid;Valproic Acid;tyrphostin AG 1024; trametinib; Tretinoin	1	0	1
BRCA2	Tretinoin; trichostatin A; Estradiol; transplatin; troglitazone; Tunicamycin; fulvestrant	1	0	1
ESR1	exemestane;tamoxifen	0	1	1
TGFB1	Doxorubicin; Fluorouracil; Thalidomide; Entinostat; Hyaluronidase	0	0	1
TP53	4-biphenylmine; alliin; Apigenin; Atropine;bicalutamide;butylidenephthalide	0	0	1

Hamed et al. *Nucl Acids Res* 43: W283-W288 (2015)

End of lecture

This concludes our small tour on modeling cell fate.

We looked at:

- Circadian clocks
- Cell cycle
- Cell differentiation
- Cancerogenesis

We showed you all the necessary techniques to **generate GRNs by yourself**.

Next week: mini-test 3, no lecture

Also open: assignment #6

If you like to conduct a Master thesis in these areas please contact me.