

A bioinformatics driven cancer study

Chao Li and Yixue Li

yxli@sibs.ac.cn

***Shanghai Institutes for Bio-Medicine
Shanghai Academy of Sciences
Shanghai Institutes for Biological Sciences
Chinese Academy of Sciences***



上海生物信息技术研究中心

SHANGHAI CENTER FOR BIOINFORMATION TECHNOLOGY



Outline

➤ Background

➤ Study design

➤ Results

➤ Summary

Background

Information **cancer network model** can

Provide insights into:

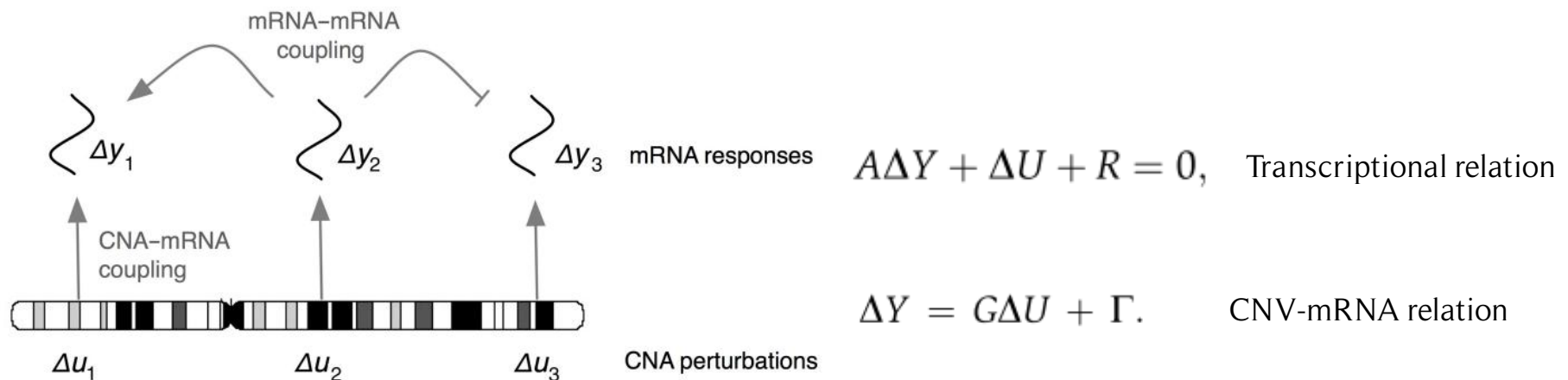
- a. Identify functionally important genes whose perturbations have a significant impact on cancer progression
- b. Identification of features that are predictive of patient prognosis or therapeutic responses

Background

Network modeling of the transcriptional effects of copy number aberrations in glioblastoma

Rebecka Jörnsten¹, Tobias Abenius^{1,2}, Teresia Kling², Linnéa Schmidt², Erik Johansson^{2,3}, Torbjörn EM Nordling⁴, Bodil Nordlander², Chris Sander⁵, Peter Gennemark^{1,6}, Keiko Funa^{2,3}, Björn Nilsson⁷, Linda Lindahl² and Sven Nelander^{2,*}

Rebecka Jörnsten et al, *Mol Syst Biol*, 2011



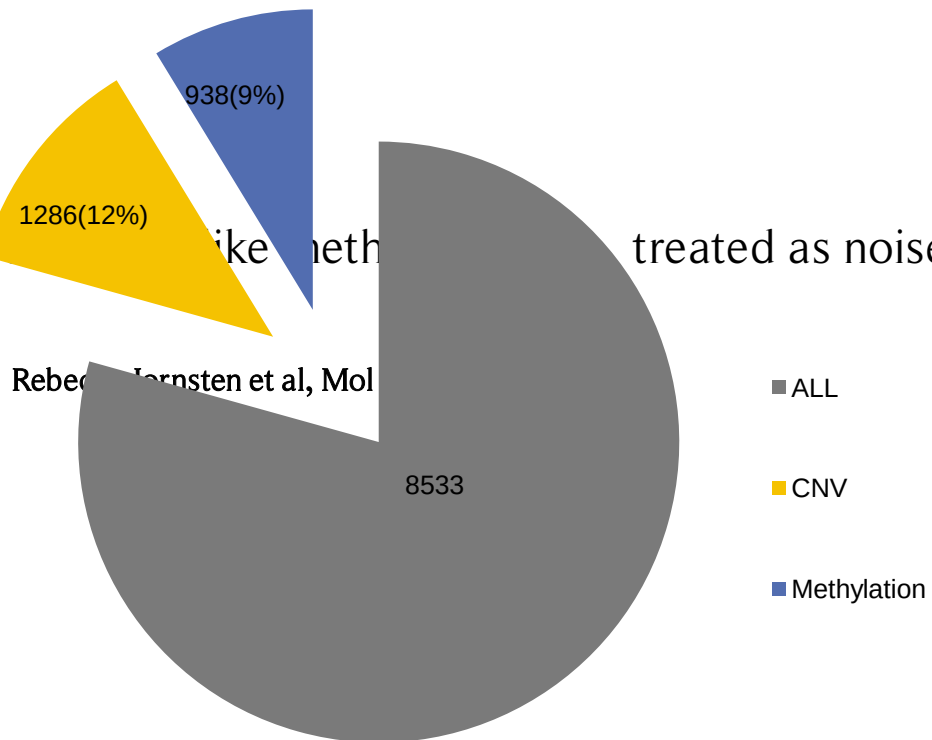
In the article, using CNV and expression data, Lasso regression was used to obtain the regulatory matrix.

Limitations

$$A\Delta Y + \Delta U + R = 0,$$

$$\Delta Y = G\Delta U + \Gamma.$$

Besides CNV, all other perturbations like methylation are treated as noise



In our analysis in case of some real data, statistically, it showed that methylation contribute to the regulations of gene expression architecture almost equal to CNV

Background: Melanoma cancer study

An Integrated Approach to Uncover Drivers of Cancer

Uri David Akavia,^{1,2,5} Oren Litvin,^{1,2,5} Jessica Kim,^{3,4} Felix Sanchez-Garcia,¹ Dylan Kotliar,¹ Helen C. Causton,¹ Panisa Pochanard,^{3,4} Eyal Mozes,¹ Levi A. Garraway,^{3,4} and Dana Pe'er^{1,2,*}

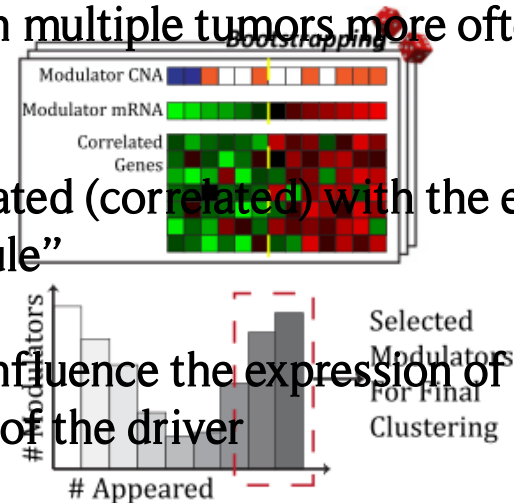
Uri David et al, *Cell*, 2010

Three assumptions of the paper:

- (1) A driver mutation should occur in multiple tumors more often than would be expected by chance
- (2) A driver mutation may be associated (correlated) with the expression of a group of genes that form a “module”
- (3) Copy number aberrations often influence the expression of genes in the module via changes in expression of the driver

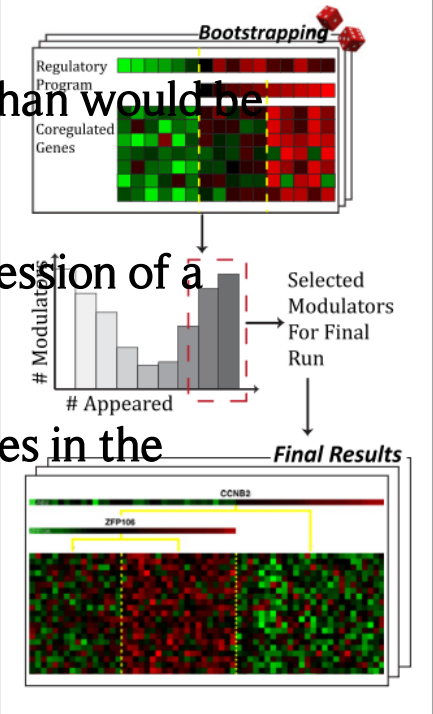
Candidate gene selection

2. Single Modulator



Single modulator

3. Network Learning



Network learning

This model may be useful in integrating non-linear perturbations

Outline

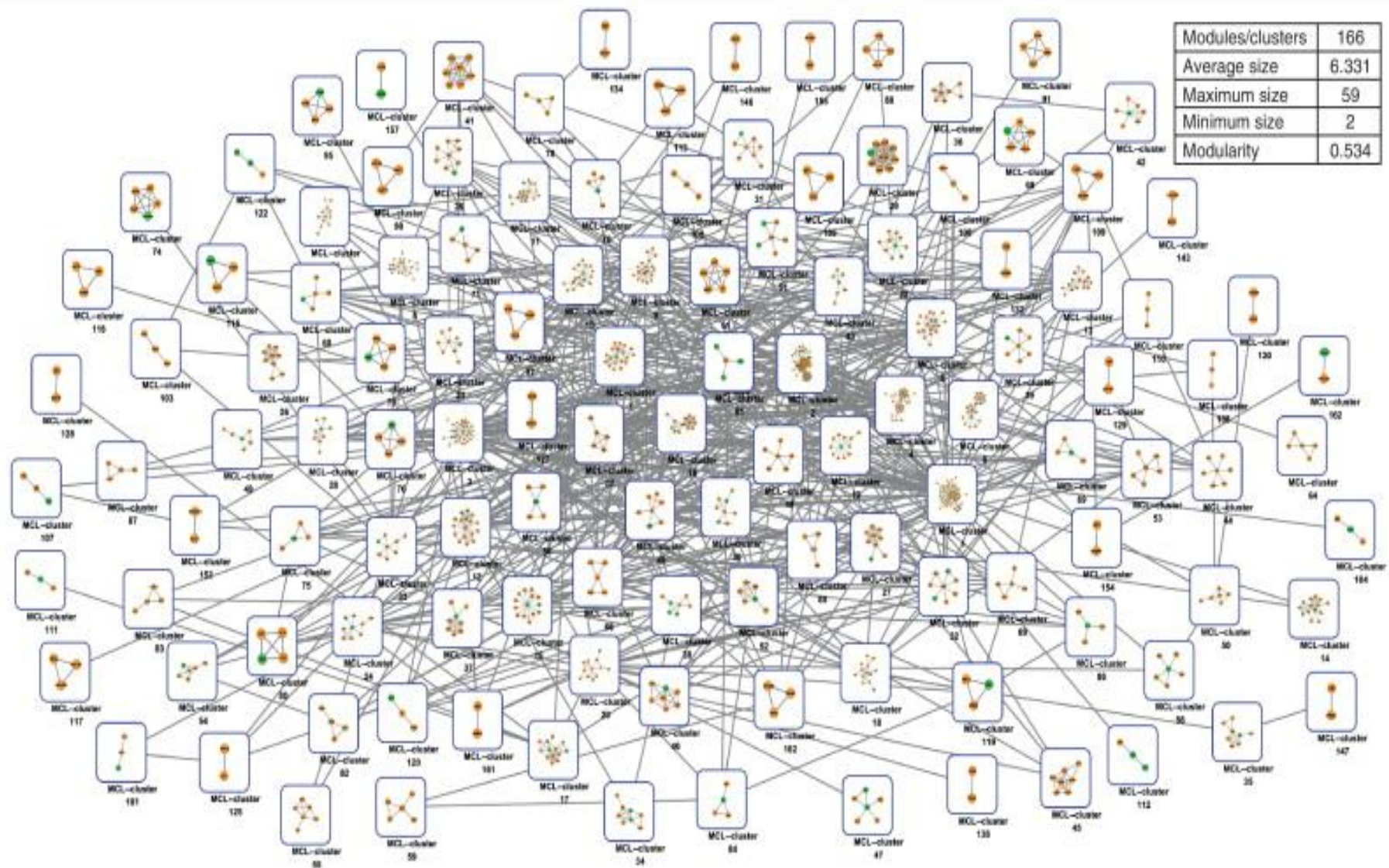
- Background
- Study design
- Results
- Summary

Study design

- **Motivation:**

Using module network model to build an integrated network covering CNV and methylation driver

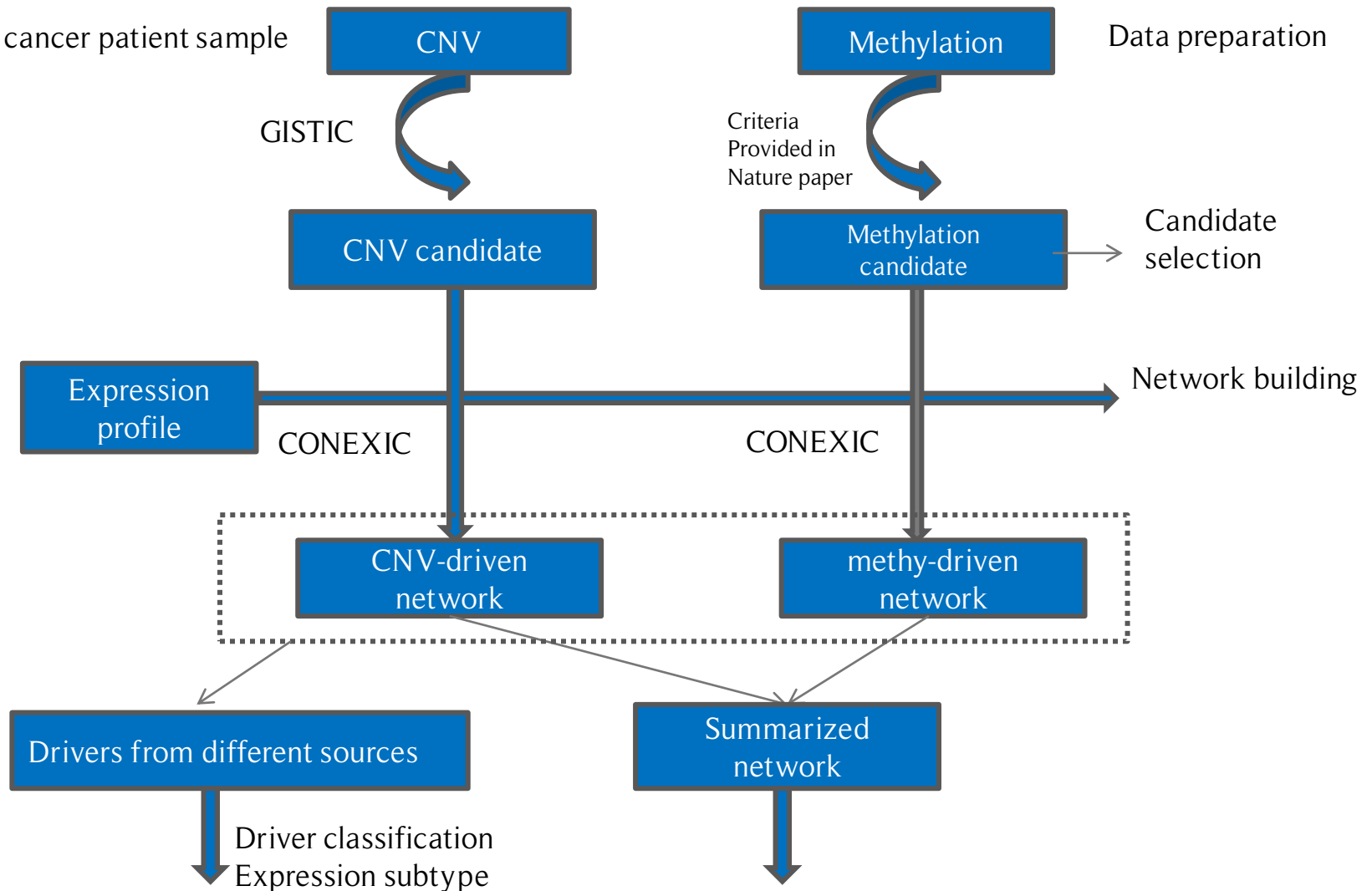
- a. More complete network can be achieved
- b. Find drivers from different sources



Drug Discovery Today

Gene Module Network: based on a global gene regulatory network, a set of genes can be explicitly put into certain module. Each module represents a set of genes that have the same statistical behavior.

Work flow



Outline



➤ Background

➤ Study design

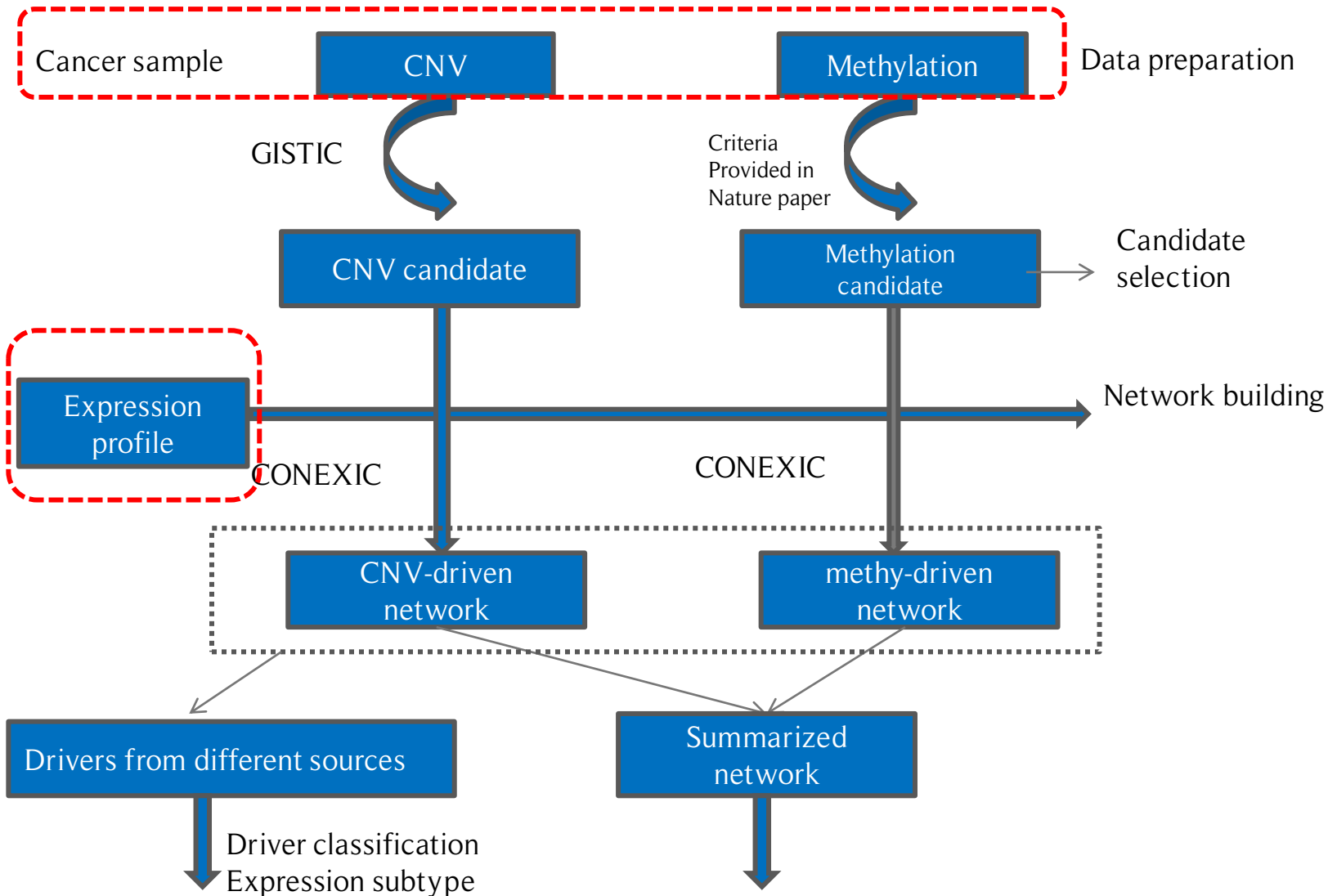
➤ **Results**

➤ Summary

Results

- Data preparation
- Candidate driver selection and network building
- Drivers' contribution to Expression subtype
- Drivers' function annotation and pathway analysis

Data preparation



Data preparation

Integrated genomic analyses of ovarian carcinoma

The Cancer Genome Atlas Research Network*

The Cancer Genome Atlas Research Network, *Nature*, 2011

489 samples of ovarian cancer patients

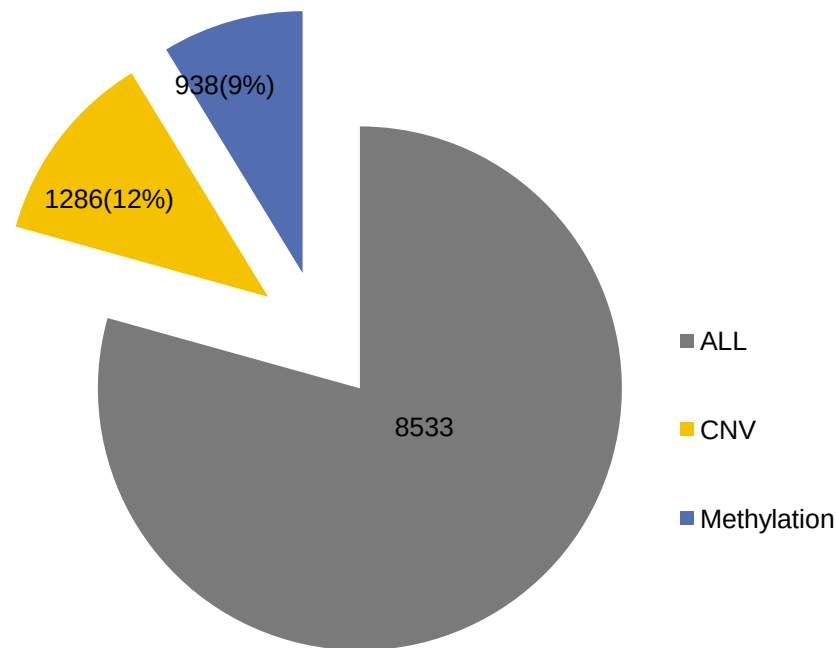
CNV profile: Agilent 1M
Methylation: Illumina 27K
miRNA profile: Agilent
mRNA profile: Affymetrix U133A
Affymetrix Exon
Agilent 244K

Clinical data

Contribution of variation

We use ANOVA to test the contribution of CNV and methylation

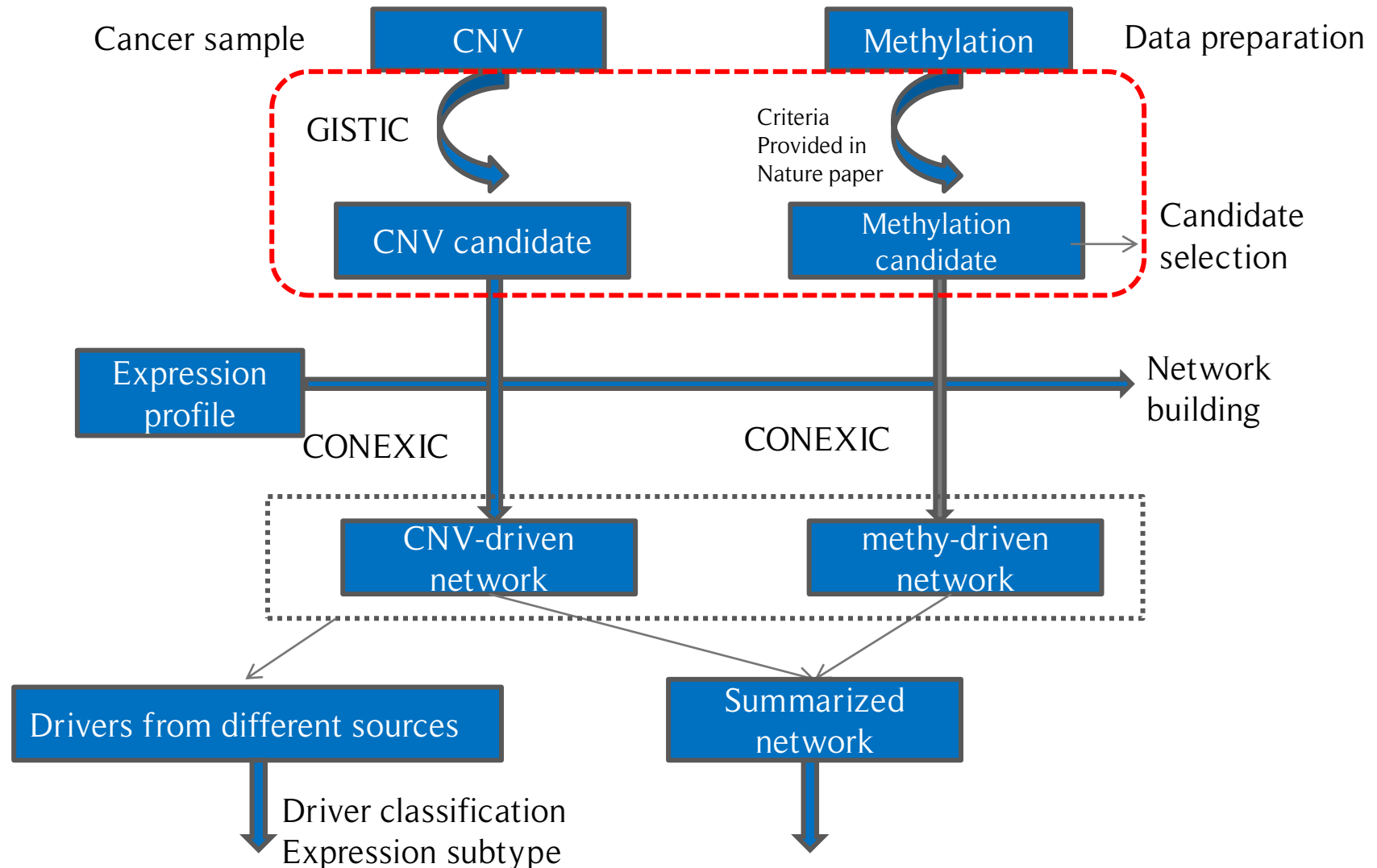
Of all 8533 genes with both types of data, CNV contribute to the regulation of 1286(12%) genes' expression and methylation contribute to the regulation of 938(9%) genes' expression as *cis/trans* regulators.



Results

- Data preparation
- Candidate driver selection and network building
- Drivers' contribution to Expression subtype
- Drivers' function annotation and pathway analysis

Candidate driver selection



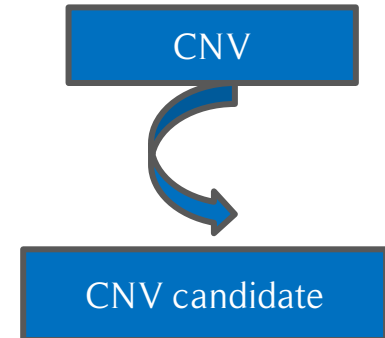
CNV candidate

Using GISTIC 2.0, q-value threshold 0.2

63 amplified, 50 deleted significant region found,
contain 3910 gene

Constantly expressed gene excluded, expression sd value > 0.25

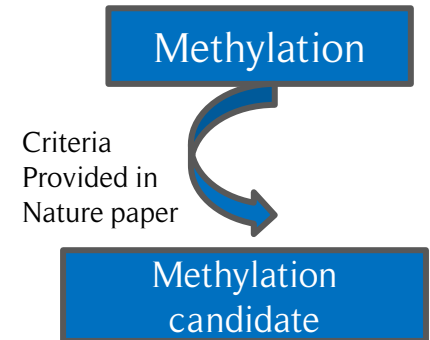
Candidate gene 2638



Methylation candidate

Criteria Provided in previous work

The Cancer Genome Atlas Research Network, *Nature*, 2011



1. Low normal tissue methylation : beta value < 0.5
2. Large difference between normal and cancer tissues : mean beta value difference > 0.3
3. Correlation between expression and methylation

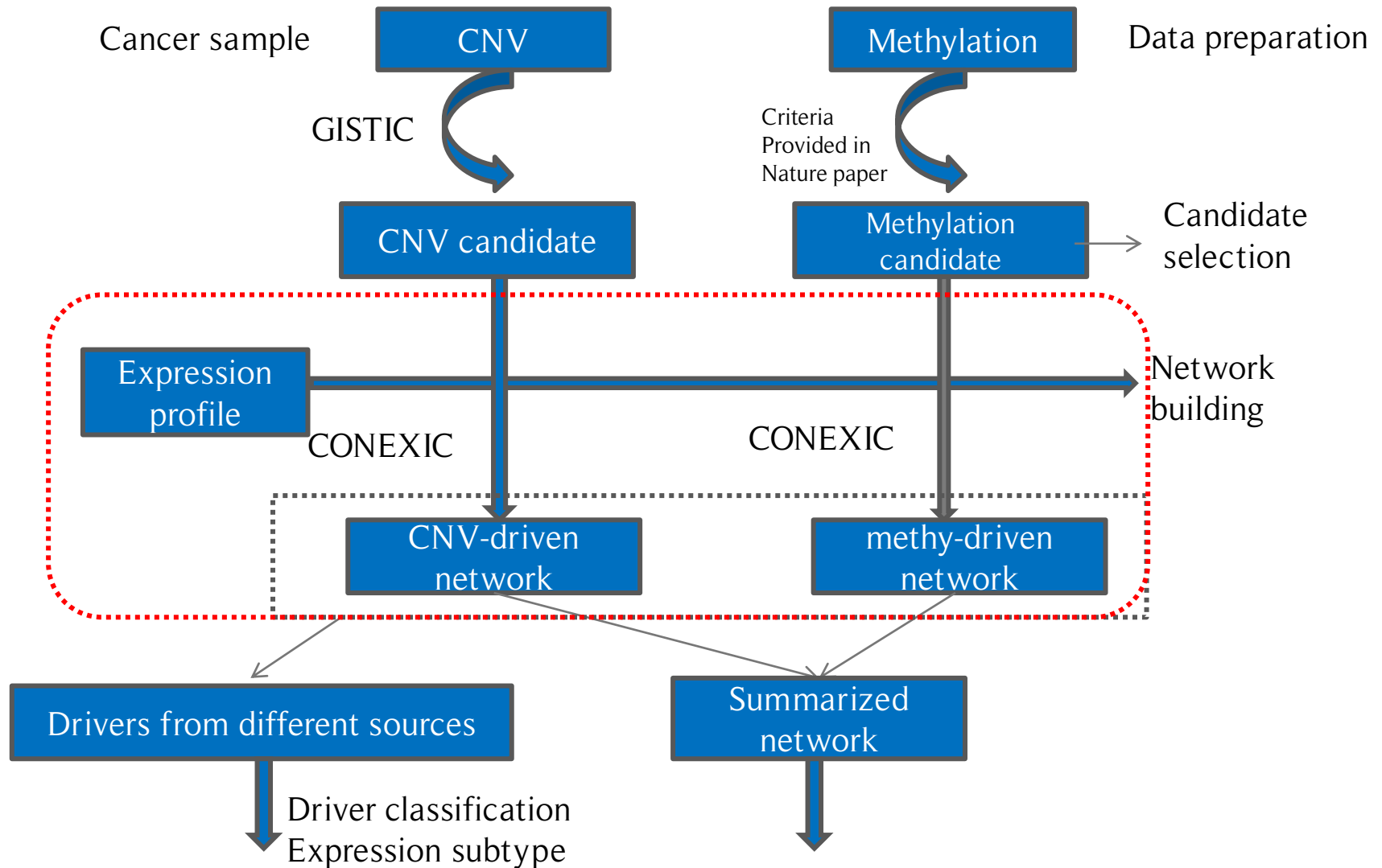
168 candidate gene in *Nature* paper

928 candidate gene selected in our study

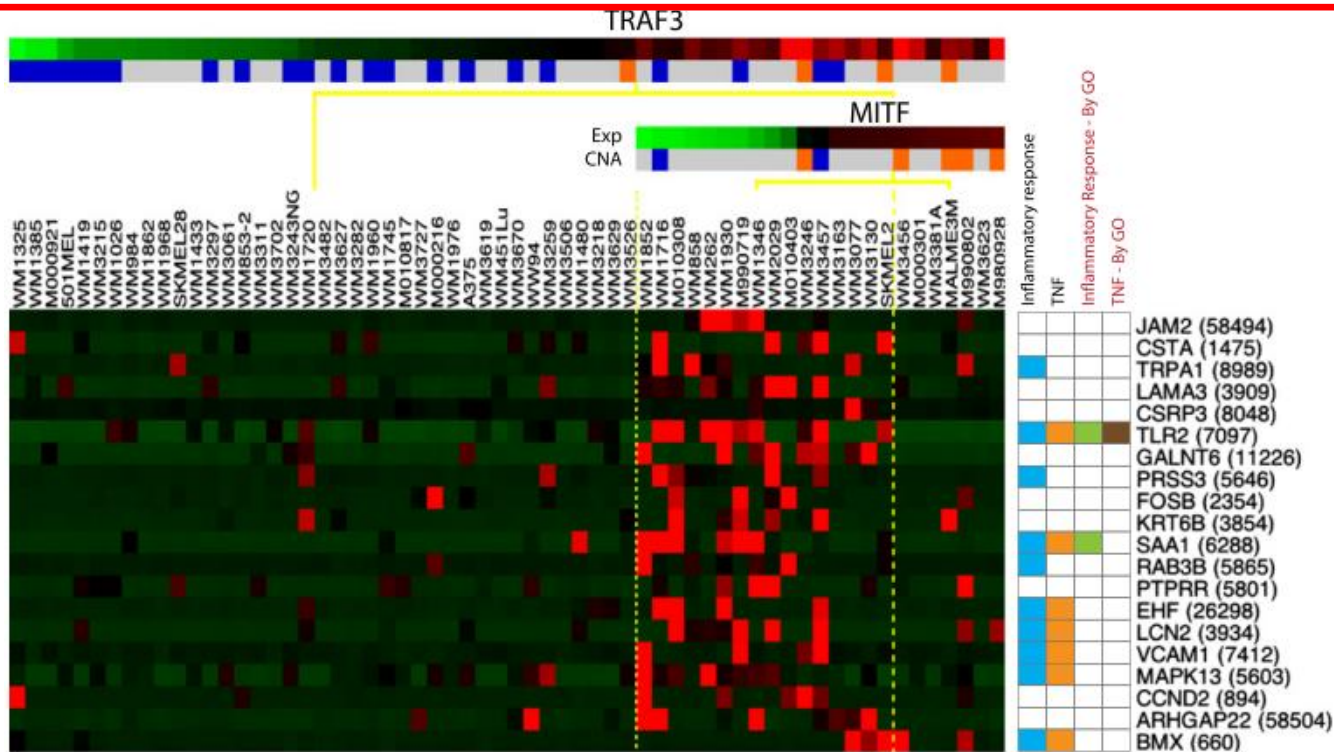
After excluding constant expressed gene : sd > 0.25

831 candidate gene selected

Module network building



Module network built



117 driver genes were found, 10 are miRNAs, 107 are genes

Among the drivers, 59 are methylation-drivers, 58 are CNV-drivers

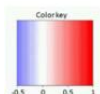
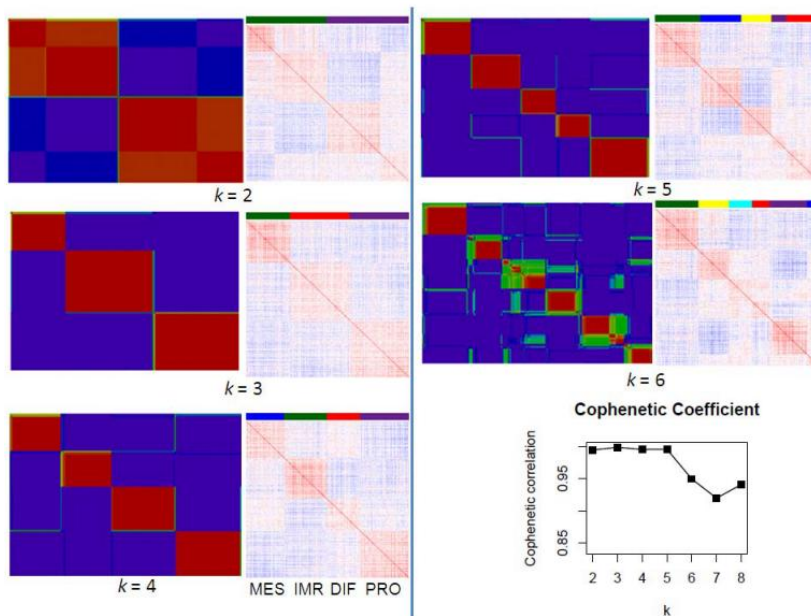
Results

- Data preparation
- Candidate driver selection and network building
- Drivers' contribution to Expression subtype
- Drivers' function annotation and pathway analysis

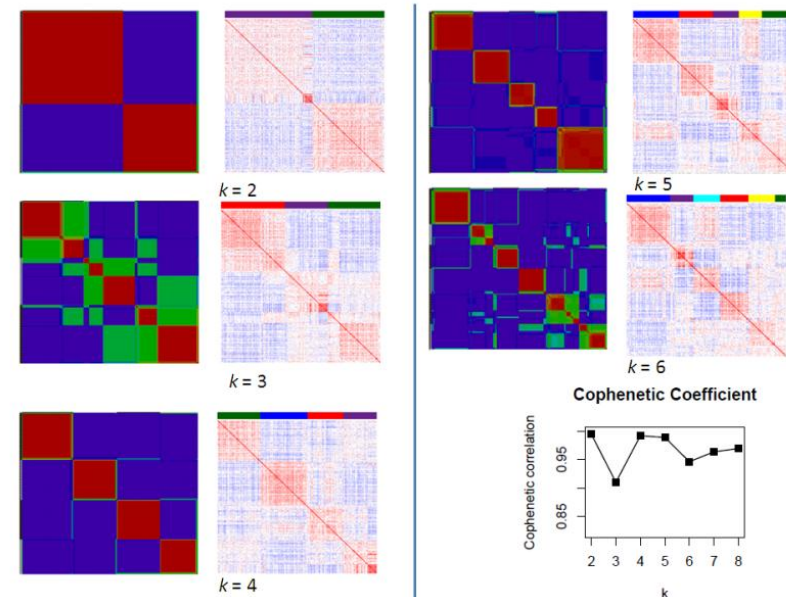
Drivers' contribution to subtyping

Ovarian cancer can be divided into 4 expression subtype

Using NMF(Non-negative matrix factorization) clustering method and 1500 highly variably expressed genes, previous paper report 4 expression subtypes of ovarian cancer can be identified.

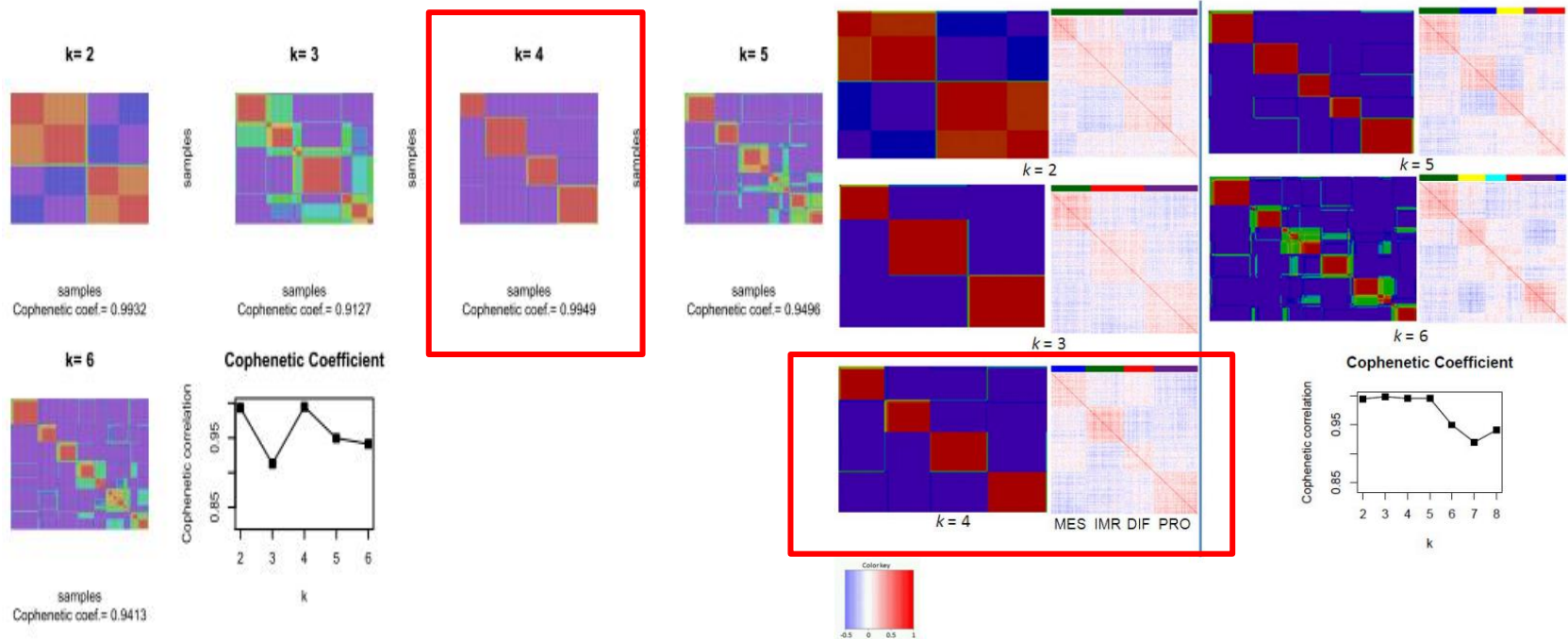


489 ovarian cancer sample, 1500 genes



245 independent ovarian cancer sample, 1500 genes

Drivers' contribution to subtyping



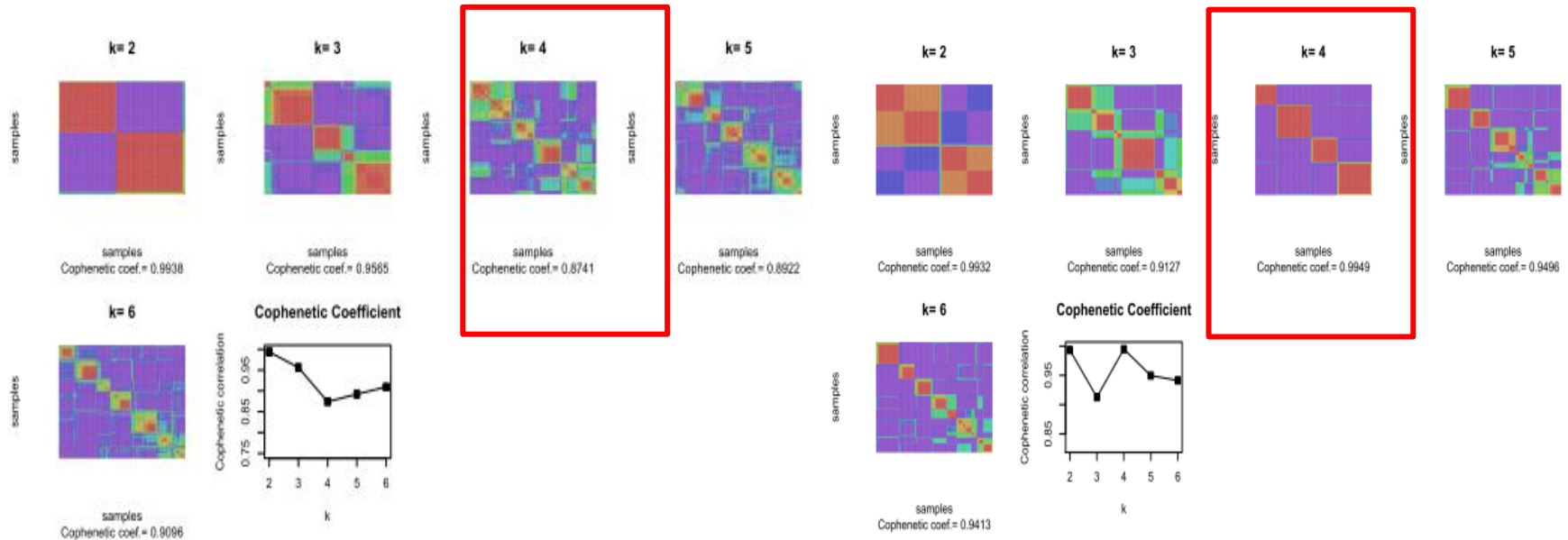
489 ovarian cancer sample, 107 drivers

489 ovarian cancer sample, 1500 genes

Among 489 samples, 332 samples are labeled the same group as using 1500 gene sets

Among the 107 drivers, 30 drivers are in the 1500 gene sets

Drivers' contribution to subtyping

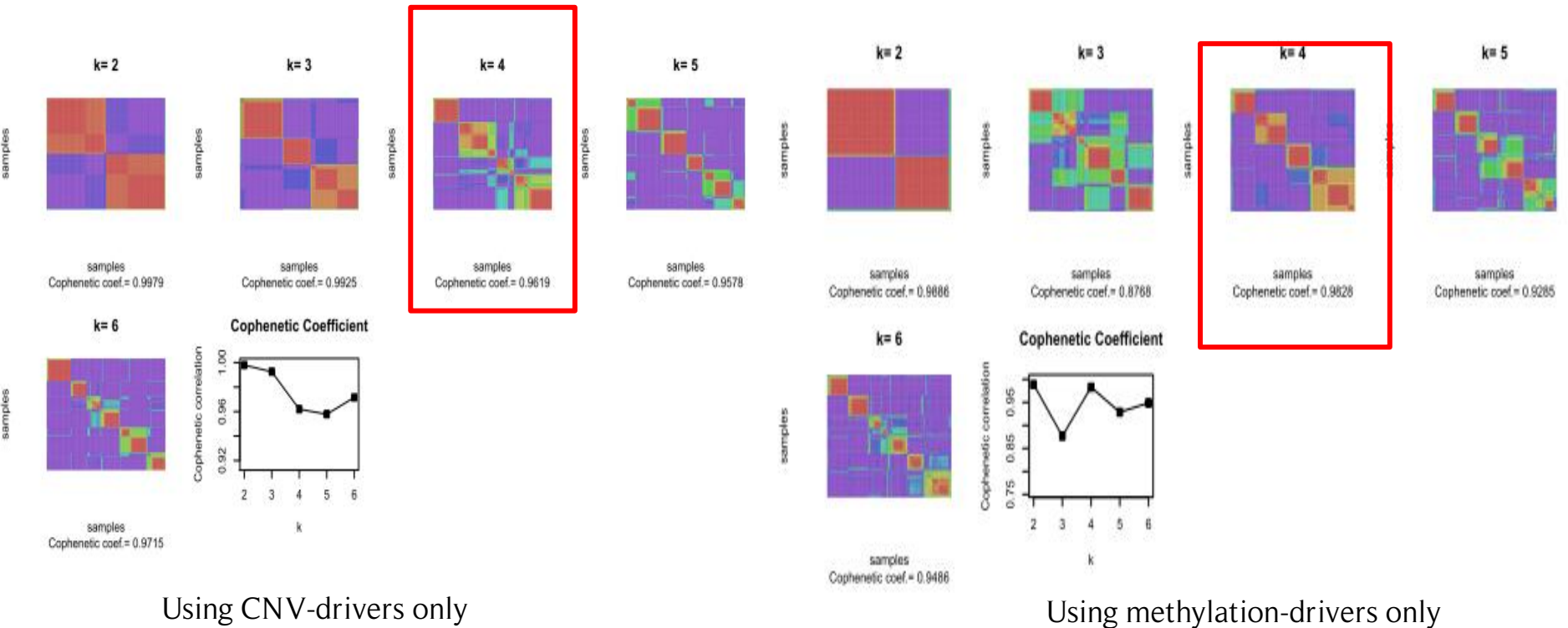


Random chosen 150 genes from 1500 gene sets

Driver gene set

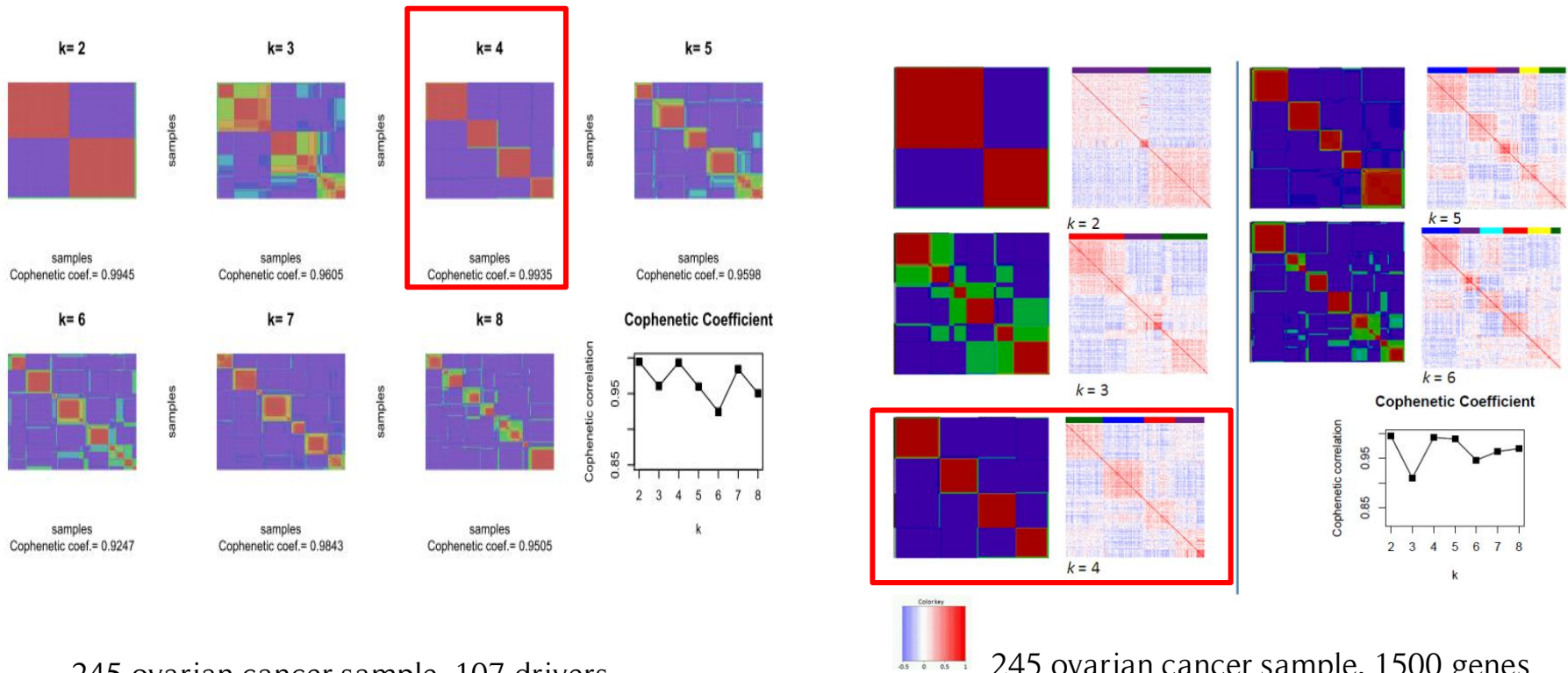
Comparing to random chosen gene set

Drivers' contribution to subtyping



Using one-source drivers only

Drivers' contribution to subtyping



245 ovarian cancer sample, 107 drivers

245 ovarian cancer sample, 1500 genes

Using independent samples to test

Results

- Data preparation
- Module network construction and drivers deriving
- Drivers' contribution to Expression subtype
- Drivers' function annotation and pathway analysis

Function annotation

Using IPA Ingenuity for gene function annotation and enrichment

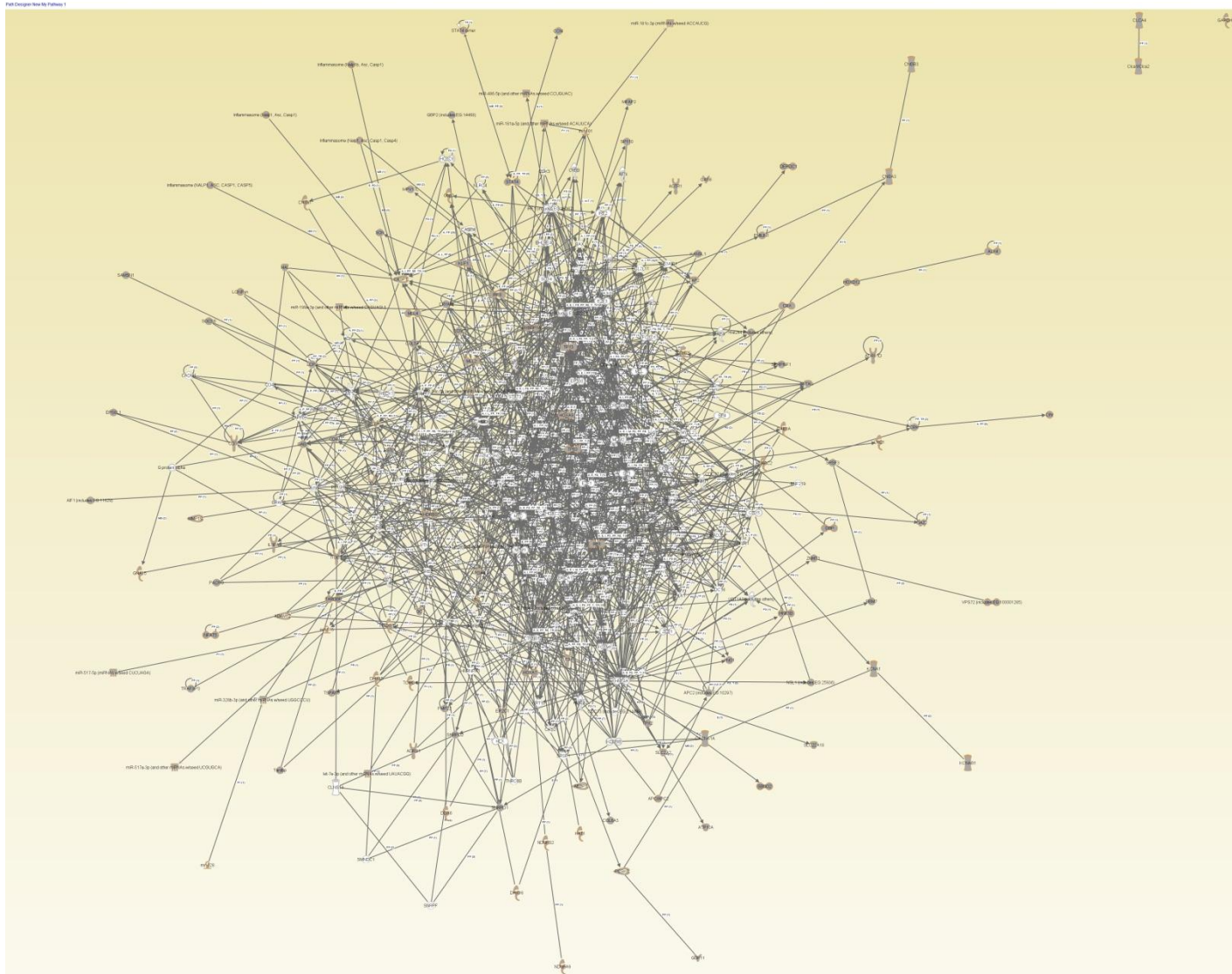
☐	Category	Functions Annotation	☐ p-Value	Molecules
☐	Cancer	Cancer	1.25E-10	AATK, ADAM12, ADAMTS13, ADRA1D, ADRB1, AEBP1, AGTR1, ALX4, ATP... all 75
☐	Cell Morphology	abnormal morphology of cells	5.08E-10	ADAMTS13, AEBP1, CACNA1A, CCND1, CD3E, CD4, CNGA3, CNTN1, CPLX3 all 30
☐	Cellular Development	proliferation of keratinocyte cancer	2.34E-07	EMILIN1, MMP19, RB1, SERPINF1, TP63, UCN all 6
☐	Cellular Growth and Proliferation	proliferation of keratinocyte cancer	2.34E-07	EMILIN1, MMP19, RB1, SERPINF1, TP63, UCN all 6
☐	Cancer	polycystic kidney disease	2.63E-07	AGTR1, Ccnb1/Gm5593, CCNC, CCND1, FSTL1, HNF1B, miR-181a-5p (and ... all 9
☐	Developmental Disorder	polycystic kidney disease	2.63E-07	AGTR1, Ccnb1/Gm5593, CCNC, CCND1, FSTL1, HNF1B, miR-181a-5p (and ... all 9
☐	Hereditary Disorder	polycystic kidney disease	2.63E-07	AGTR1, Ccnb1/Gm5593, CCNC, CCND1, FSTL1, HNF1B, miR-181a-5p (and ... all 9
☐	Renal and Urological Disease	polycystic kidney disease	2.63E-07	AGTR1, Ccnb1/Gm5593, CCNC, CCND1, FSTL1, HNF1B, miR-181a-5p (and ... all 9
☐	Cell Death and Survival	apoptosis	5.02E-07	AATK, ADAM12, ADRA1D, ADRB1, AGTR1, AIF1 (includes EG:11629), ALX4, . all 57
☐	Organismal Development	morphology of body region	6.43E-07	AEBP1, ALX4, CCND1, CD47, CNGA3, CNGB3, CPLX3, CRX, FSTL1, GDF11, .. all 19
☐	Cancer	head and neck tumor	7.57E-07	ADAM12, AEBP1, AGTR1, BIRC5, CACNA1A, CCND1, COL6A3, HOXA9, LCK all 21
☐	Cancer	liver cancer	7.90E-07	ADAM12, BIRC5, CCNB1, CCND1, F3, GC, let-7, let-7a-5p (and other miR... all 19
☐	Gastrointestinal Disease	liver cancer	7.90E-07	ADAM12, BIRC5, CCNB1, CCND1, F3, GC, let-7, let-7a-5p (and other miR... all 19
☐	Hepatic System Disease	liver cancer	7.90E-07	ADAM12, BIRC5, CCNB1, CCND1, F3, GC, let-7, let-7a-5p (and other miR... all 19
☐	Hepatocellular Carcinoma	liver cancer	7.90E-07	ADAM12, BIRC5, CCNB1, CCND1, F3, GC, let-7, let-7a-5p (and other miR... all 19

Function annotation enrichment of all drivers

Pathway analysis

- Generate the driver-related pathway
- Sub-networks of the overall pathway
- A high-confidence driver set based on pathway analysis

Pathway analysis--global pathway



Driver-related global pathway

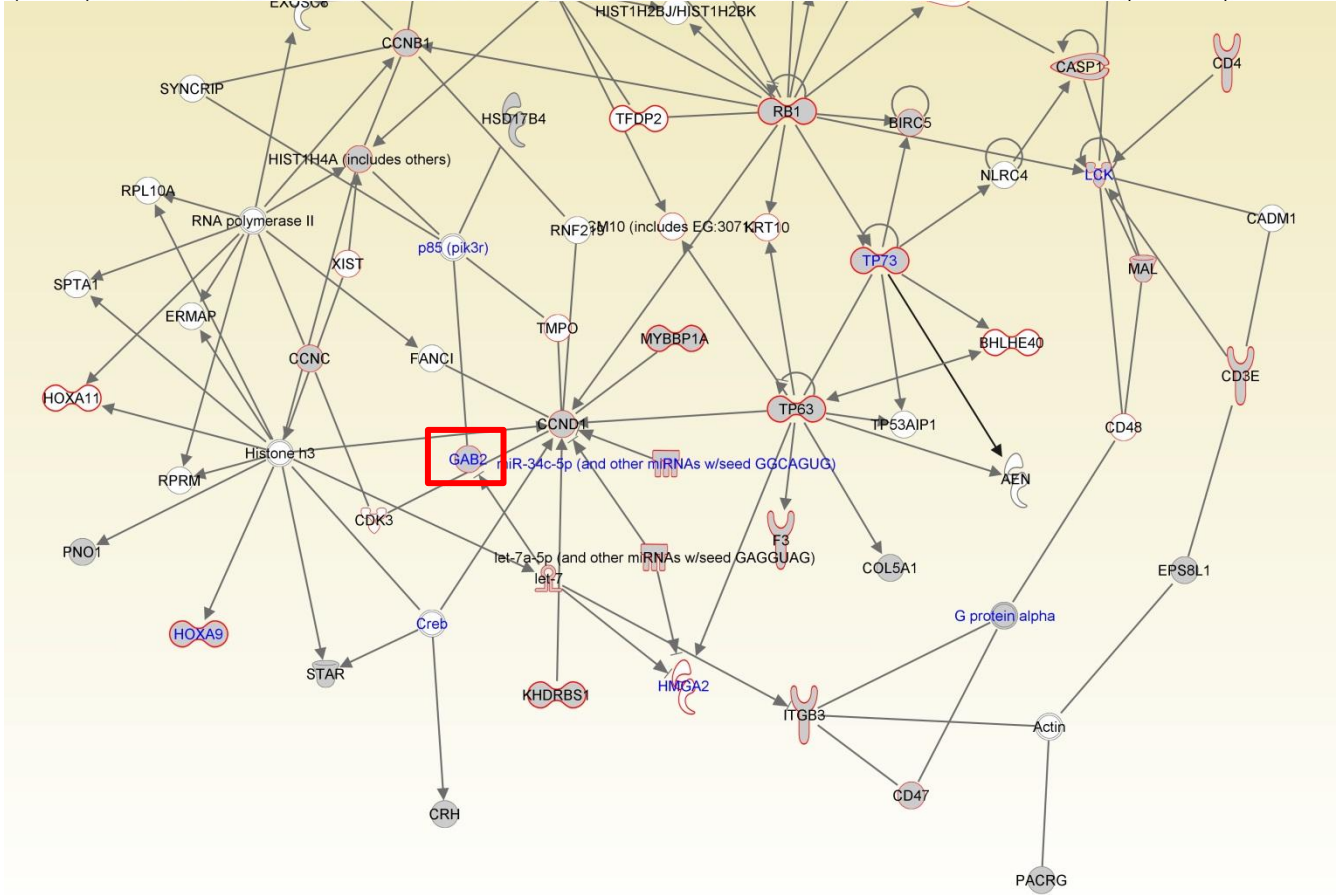
Pathway analysis--subnetworks

The global pathway can be divided into several subnetworks

	A	ID	Molecules in Network	Score	Focus Molec	Top Functions
☐	1		Actin, AEN, ALG5, ANAPC2 , BHLHE40, BIRC5 , CADM1, CASP1 , CASP4, CCNB1 , CCNC , CCND1 , CD4 , CD47 , CD48, CD3E , CDK3, COL5A1 , Creb, CRH , CYP27B1, E2f, EPS8L1 , ERMAP, EXOSC8, F3 , FANCI, G protein alpha, GAB2 , HIST1H2AB/HIST1H2AE, HIST1H2BJ/HIST1H2BK, HIST1H4A (includes others) , Histone h3, HMGA2, HOXA9 , HOXA11, HSD17B4 , ITGB3 , KHDRBS1 , KRT10, LCK , LCP2 , let-7 , let-7a-5p (and other miRNAs w/seed GAGGUAG) , MAL , MCM10 (includes EG:307126), MFAP1, miR-34c-5p (and other miRNAs w/seed GGCAGUG) , MYBBP1A , NLRC4, p85 (pik3r), PACRG , PGRMC1, PNO1 , RAB1A , RB1 , RNA polymerase II, RNF219, RPL10A, RPRM, SPTA1, STAR , SYNCRIP, TFDP2, TMPO, TP63 , TP73 , TP53AIP1, TPX2 , XIST	48	34	Cancer, Cell Cycle, Hematological Disease
☐	2		ACSL1, ADAM15, ADRB1 , AJAP1, AKAP5, AKAP13, CALD1, CD44, CDH1, CEBPA, CRYAB , CTNNB1, DDX6 , DDX21, DHX15 , DLG1, DLK1, EFNA1, EFTUD2, EIF2C1 , EIF2C2, ESR1, EXTL2, EZH2, GAS6 , GC , GEMIN4, GPR64, GPR124 , GRB2, HOXC6, ICOSLG, KANSL1 , KCNA1 , KCNK10 , KLF1 , MBD2, mir-26, MLL2 , MLL4 , MLL, MTMR6 , NET1 (includes EG:10276), NFAT5 , PCSK6 , PDX1 (includes EG:18609), PGR, PRPF8, PTPRF, RPLP2, S100P, SAA1, SART1, SERPINA5, SERPINF1 , SLC2A2 , SMNDC1, SNRNP200, SNRPD1 , SNRPD2 , SNRPF, SRF, STRN4, TFF2, TRAF3IP3 , TRD@, TSPAN8 , UGCG, WNT11, ZMAT5	29	24	Cellular Movement, RNA Post-Transcriptional Modification, Renal and Urological System Development and Function
☐	3		AATK , ADAM12 , ADRA1B, ADRA1D , AKR1C4, APC/APC2 , APOA4 , BHLHE40, CBX4, CKAP2, DAPK3, DBP , DMAP1, EBAG9, FKBP4, GDF11 , GLIPR1, GNB2L1, GRIN1, HDAC3, HK2, HNF1B , HNF4A, HOXA7 , IGFBP3, IL15 (includes EG:16168), IL5RA , IRF8 , JAK2, LEP, MAOA, mir-27, mir-132, mir-181 , mir-515 , miR-181a-5p (and other miRNAs w/seed ACAUUCA) , miR-486-5p (and other miRNAs w/seed CCUGUAC) , MMP19 , MST1 (includes EG:15235), NAMPT, NDC80, NEK2, NLRC4, NQO1, NR3C1, NSL1 (includes EG:25936) , NSMAF, PDE4B, PDE6B, PDK1, PLK2, PPP1R13L, PPP4C, PTN, SLC19A1, SLC2A12, SON , SP1, SRSF3 , STAT3, STAT4 , TAP2 , TARDBP , TNFSF9, TP53 (includes EG:22059), TP53I3, TYK2, UGT1A9 (includes others), VPS72 (includes EG:100001285) , ZBTB5	27	23	Cell Death and Survival, Cell Cycle, Neurological Disease
☐	4		DEPDC1 , FOXO1	2	1	Cancer, Cardiovascular System Development and Function, Cell Morphology
☐	5		ICK , SOX11	2	1	Digestive System Development and Function, Cell Death

Pathway analysis--subnetwork1

	Molecules in Network	Score	Focus Molec	Top Functions
1	Actin, AEN, ALG5, ANAPC2 , BHLHE40, BIRC5 , CADM1, CASP1 , CASP4, CCNB1 , CCNC , CCND1 , CD4 , CD47 , CD48, CD3E , CDK3, COL5A1 , Creb, CRH , CYP27B1, E2f, EPS8L1 , ERMAR, EXOSC8, F3 , FANCI, G protein alpha, GAB2 , HIST1H2AB/HIST1H2AE, HIST1H2B/J/HIST1H2BK, HIST1H4A (includes others) , Histone h3, HMGA2, HOXA9 , HOXA11, HSD17B4 , ITGB3 , KHDRBS1 , KRT10, LCK , LCP2 , let-7 , let-7a-5p (and other miRNAs w/seed GAGGUAG) , MAL , MCM10 (includes EG:307126), MFAP1, miR-34c-5p (and other miRNAs w/seed GGCAGUG) , MYBBP1A , NLRC4, p85 (pik3r), PACRG , PGRMC1, PNO1 , RAB1A , RB1 , RNA polymerase II, RNF219, RPL10A, RPRM, SPTA1, STAR , SYNCRIP , TFDP2, TMPO, TP63 , TP73 , TP53AIP1, TPX2 , XIST	48	34	Cancer, Cell Cycle, Hematological Disease



Top-related function : **Cancer**
8.95E-7

- Driver gene
- Ovarian cancer related
- Top function related

Pathway analysis--Gab2 case

Gab2: Overexpression of Gab2 promotes the migration and invasion in ovarian cancer cells

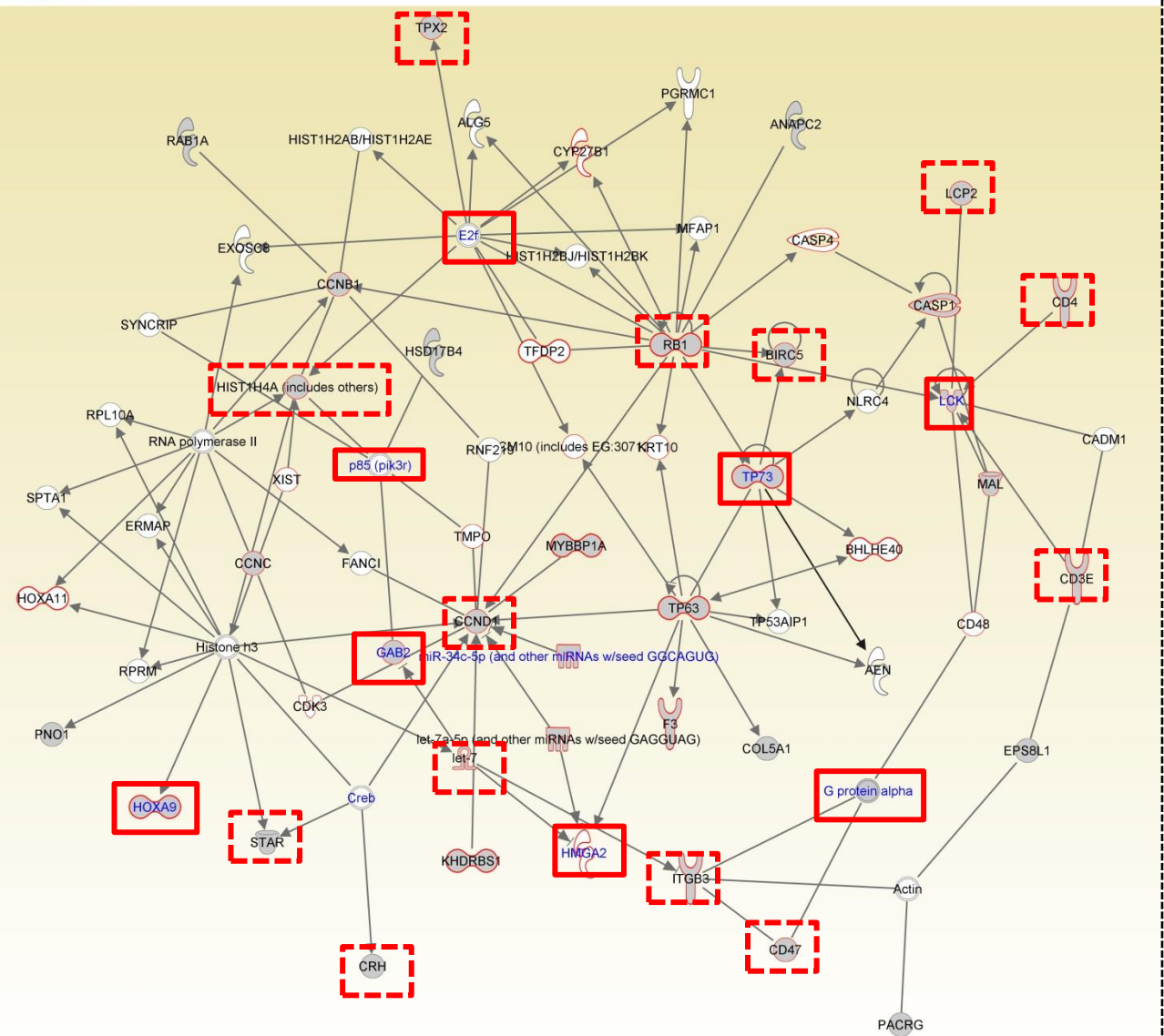
Wang Y et al Oncogene ,2012

In our analysis, Gab2 is classified as cnv-amplified driver amplified in 36% sample

Pathway analysis-high-confidence driver set



Path Designer Network 1



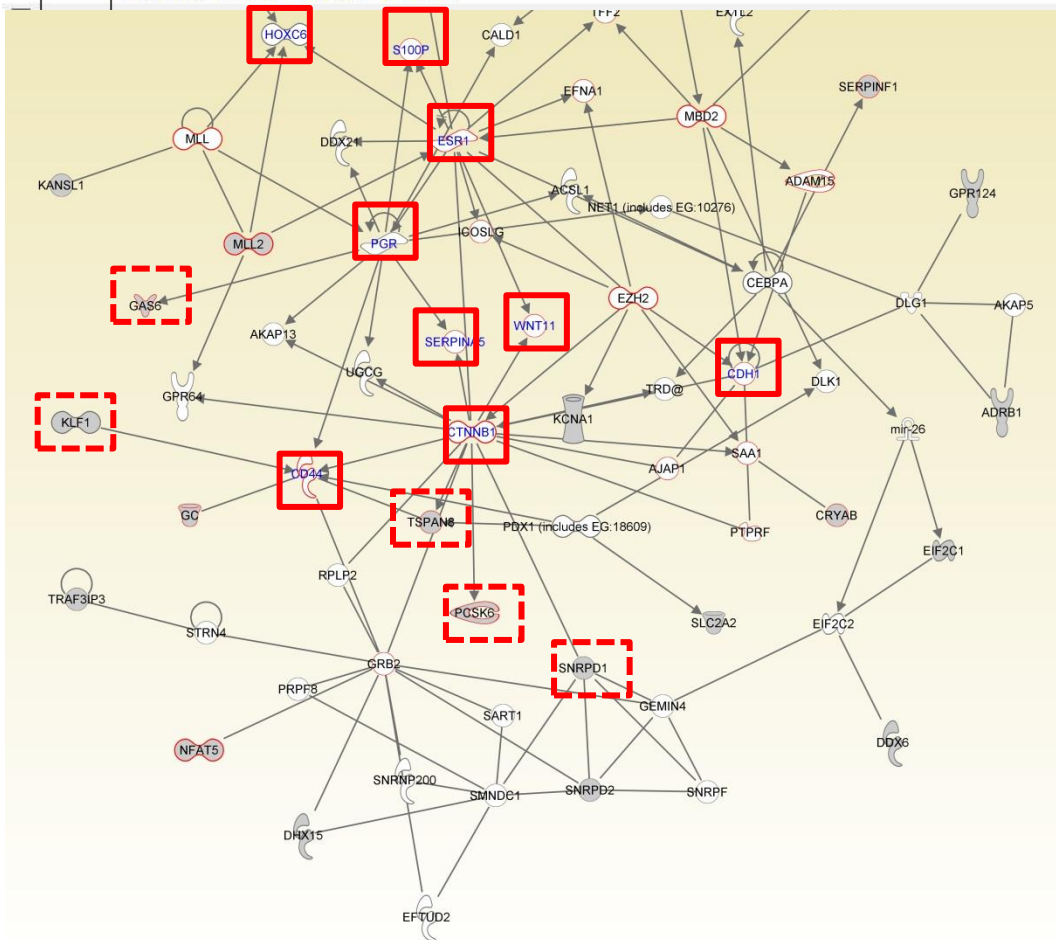
-  Known cancer related genes
-  New found driver genes which may be related to cancer

Pathway analysis- high-confidence driver set 1

Gene	Variation type	Frequency	Function
BIRC5	amp	41.50%	inhibitor of apoptosis (IAP) gene family
CCND1	methy	96%	highly conserved cyclin family
CD4	methy	97.30%	membrane glycoprotein of T lymphocytes
CD47	methy	58.70%	membrane transport and signal transduction
CD3E	del	26.80%	coupling antigen recognition to signal transduction pathway
CRH	methy	55.80%	secreted by hypothalamus in response to stress
ITGB3	methy	62.40%	participate in cell adhesion and cell-surface mediated signalling
LCP2	del	31.30%	promoting T cell development and activation
let-7	del	72.40%	target oncogenes RAS, HMGA2
RB1	del	52.30%	negative regulator of the cell cycle
STAR	methy	94.30%	steroid hormone synthesis
TPX2	amp	56.20%	microtubule-associated protein, activates the cell-cycle kinase

Pathway analysis--subnetwork2

2	ACSL1, ADAM15, ADRB1, AJAP1, AKAP5, AKAP13, CALD1, CD44, CDH1, CEBPA, CRYAB, CTNNB1, DDX6, DDX21, DXH15, DLG1, DLK1, EFNA1, EFTUD2, EIF2C1, EIF2C2, ESR1, EXTL2, EZH2, GAS6, GC, GEMIN4, GPR64, GPR124, GRB2, HOXC6, ICOSLG, KANSL1, KCNA1, KCNK10, KLF1, MBD2, mir-26, MLL2, MLL4, MLL, MTMR6, NET1 (includes EG:10276), NFAT5, PCSK6, PDX1 (includes EG:18609), PGR, PRPF8, PTPRF, RPLP2, S100P, SAA1, SART1, SERPINA5, SERPINF1, SLC2A2, SMINDC1, SNRNP200, SNRPD1, SNRPD2, SNRPF, SRF, STRN4, TFF2, TRAF3IP3, TRD@, TSPAN8, UGCG, WNT11, ZMAT5	29	24	Cellular Movement, RNA Post-Transcriptional Modification, Renal and Urological System Development and Function
---	--	----	----	--



Top-related function : Cellular Movement (invasion,migration,etc) 1.65E-5



Driver gene



Ovarian cancer related



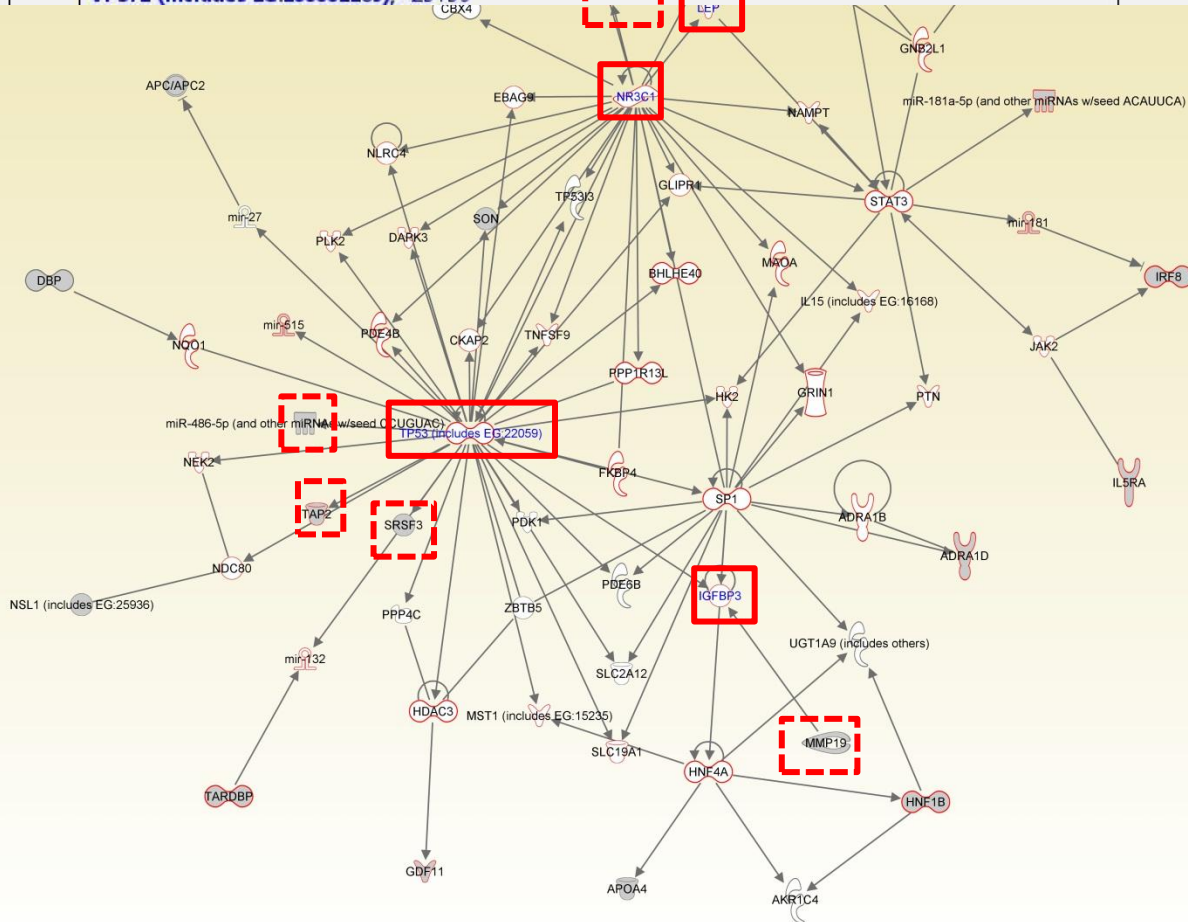
Top function related

Pathway analysis-high-confidence driver set 2

Gene	Variation type	Variation frequency	Function
GAS6	methy	77.50%	involved in the stimulation of cell proliferation
DBP	del	41.30%	involved in the regulation of some circadian rhythm genes
KLF1	methy	66.50%	induces high-level expression of adult beta-globin and other erythroid genes
MLL4	methy	99.60%	beta-globin locus transcription regulation
PCSK6	amp	19.00%	process latent precursor proteins into their biologically active products
SNPRD1	methy	86.50%	charged protein scaffold to promote SNRNP assembly
TSPAN8	amp	22.70%	mediate signal transduction events that play a role in the regulation of cell development, activation, growth and motility
ZMAT5	amp	16.20%	zinc finger, matrin-type 5

Pathway analysis--subnetwork3

3	AATK, ADAM12, ADRA1B, ADRA1D, AKR1C4, APC/APC2, APOA4, BHLHE40, CBX4, CKAP2, DAPK3, DBP, DMAP1, EBAG9, FKBP4, GDF11, GLIPR1, GNB2L1, GRIN1, HDAC3, HK2, HNF1B, HNF4A, HOXA7, IGFBP3, IL15 (includes EG:16168), IL5RA, IRF8, JAK2, LEP, MAOA, mir-27, mir-132, mir-181, mir-515, miR-181a-5p (and other miRNAs w/seed ACAUUCA), miR-486-5p (and other miRNAs w/seed CCUGUAC), MMP19, MST1 (includes EG:15235), NAMPT, NDC80, NEK2, NLRC4, NQO1, NR3C1, NSL1 (includes EG:25936), NSMAF, PDE4B, PDE6B, PDK1, PLK2, PPP1R13L, PPP4C, PTN, SLC19A1, SLC2A12, SON, SP1, SRSF3, STAT3, STAT4, TAP2, TARDBP, TNFSF9, TP53 (includes EG:22059), TP53I3, TYK2, UGT1A9 (includes others), VPS72 (includes EG:100001285), ZBTB5	27	23	Cell Death and Survival, Cell Cycle, Neurological Disease
---	--	----	----	---



Top-related function : **Cell Death and Survival 1.1E-9**

- Driver gene
- Ovarian cancer related
- Top function related

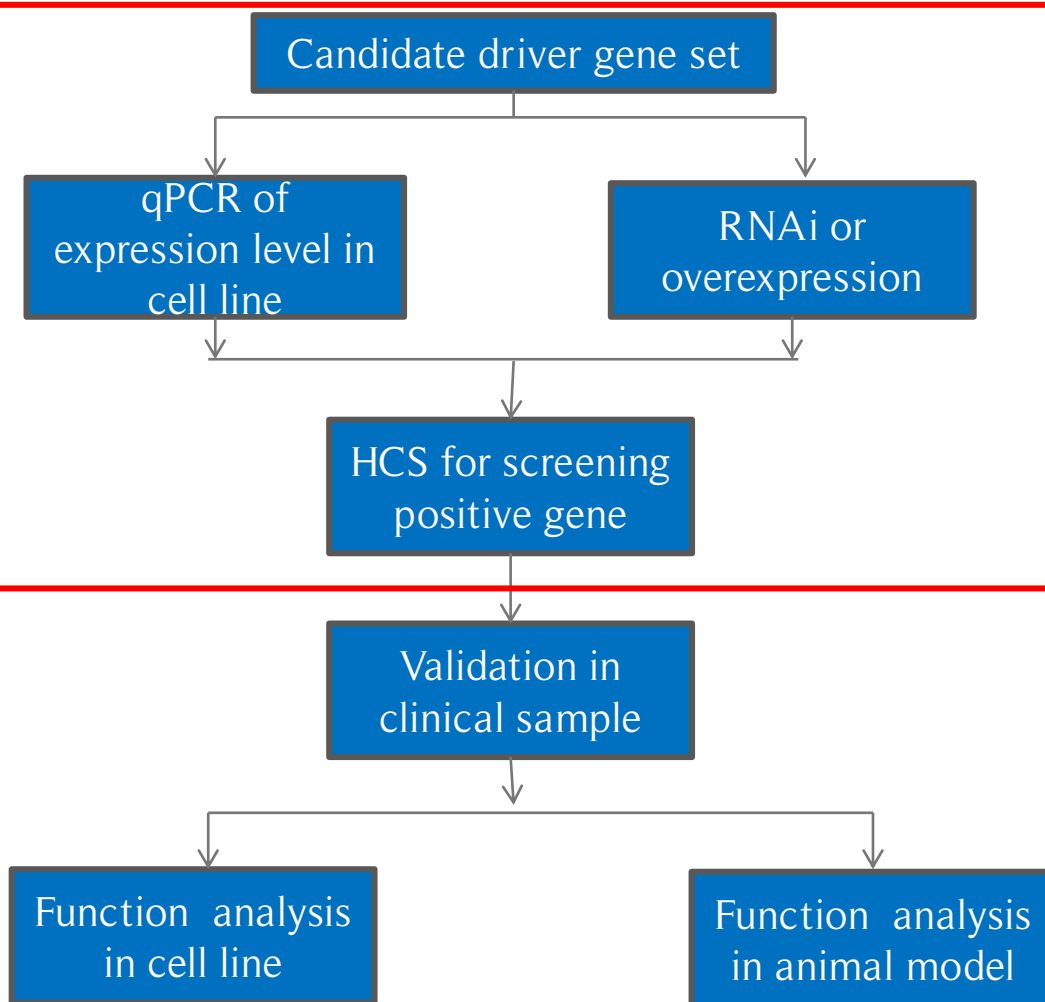
Pathway analysis-high-confidence driver set 3

Gene	Variation type	Variation frequency	Function
AATK	amp	23.30%	pre-requisite for the induction of growth arrest and/or apoptosis of myeloid precursor cells
mir-486-5p	amp	24.90%	induces replicative senescence of human adipose tissue-derived mesenchymal stem cells
MMP19	methy	88.30%	breakdown of extracellular matrix in normal physiological processes
SRSF3	amp	33.70%	mRNA splicing
TAP2	methy	72.00%	transport various molecules across extra- and intra-cellular membranes

According to SPD database AATK and MMP19 are secreted protein

- By Integrating both of the CNV-and Methylation-driven network, we identified 117 driver genes, which are highly related to functions cancer progresses and development
- The drivers we found are highly correlated with ovarian cancer and can be used to identify cancer subtype correctly
- Drivers' function annotation and regulatory role in pathway indicates that several novel genes may relate to ovarian cancer
- A 26 high-confidence driver set was derived depending on regulatory relationship with known ovarian cancer genes

Candidate gene function experiment



For now, the HSC(High Cotent Screening) step has been finished

8 genes with function relevant to cell growth and proliferation are selected

Positive gene list

Gene	Positive cell line	Variation type
SRSF3	HEY, SKOV3	amp
BIRC5	HEY	amp
GAS6	HEY	methy
CD4	HEY	methy
CD3E	HEY	del
PNO1	SKOV3	amp
CYC1	SKOV3	amp
TAP2	SKOV3	methy

Acknowledgement

Dr.Ding Guohui,Dr.Dong Xiao



Gene function and role in cancer

SRSF3:

Function: a member of the serine/arginine (SR)-rich family of pre-mRNA **splicing factors**

Potential role in cancer: Downregulation of splicing factor SRSF3 induces p53 β an alternatively spliced isoform of p53 that promotes **cellular senescence**

Tang, Y et.al 2012 *Oncogene*

Gene function and role in cancer

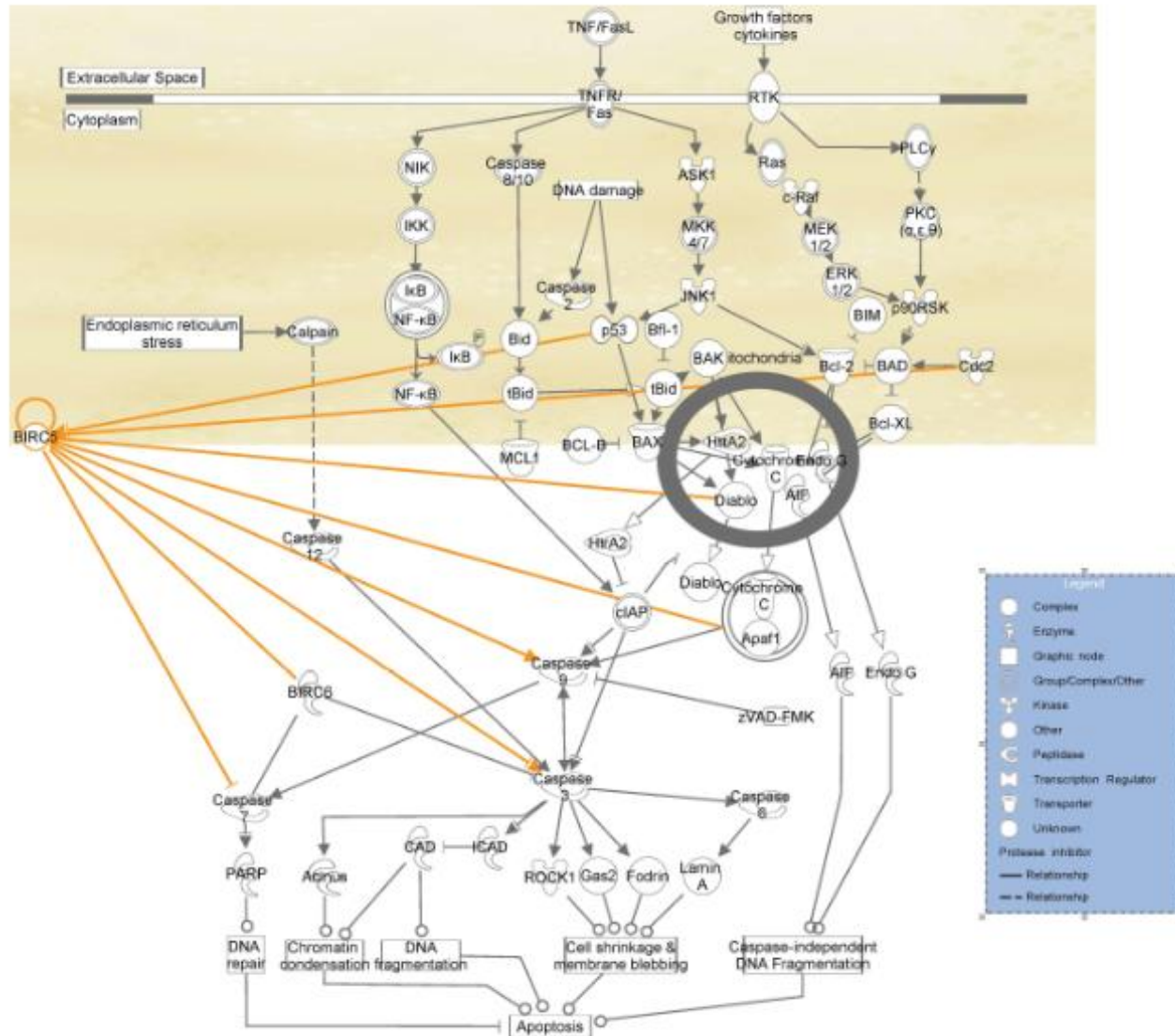
BIRC5:

Function: also called Survivin, a member of the inhibitor of apoptosis (IAP) family, leading to negative regulation of apoptosis or programmed cell death

Potential role in cancer: Overexpression in several cancer types and promote growth

N. A. Vayshlya et.al 2008 Molecular Biology
Boidot R et.al 2008 Genes Chromosomes Cancer

BIRC5 interaction with apoptosis pathway



Gene function and role in cancer

GAS6:

Function: a gamma-carboxyglutamic acid (Gla)-containing protein thought to be involved in the stimulation of cell proliferation

Potential Role in cancer:

Has been reported coexpress with Axl and overexpression correlate with poor prognosis in several cancer types.

Hutterer M et.al 2008 Clin Cancer Res
Sun W et.al 2004 Oncology

Gas6 expression correlating with favorable prognosis has been reported in human breast cancer and in RCC

Mc Cormack O et.al 2008 Br J Cancer
Gustafsson A et.al 2009 Clin Cancer Res

Gene function and role in cancer

CD4,CD3E:

Function:

CD4 encodes a membrane glycoprotein of T lymphocytes that interacts with major histocompatibility complex class II antigens

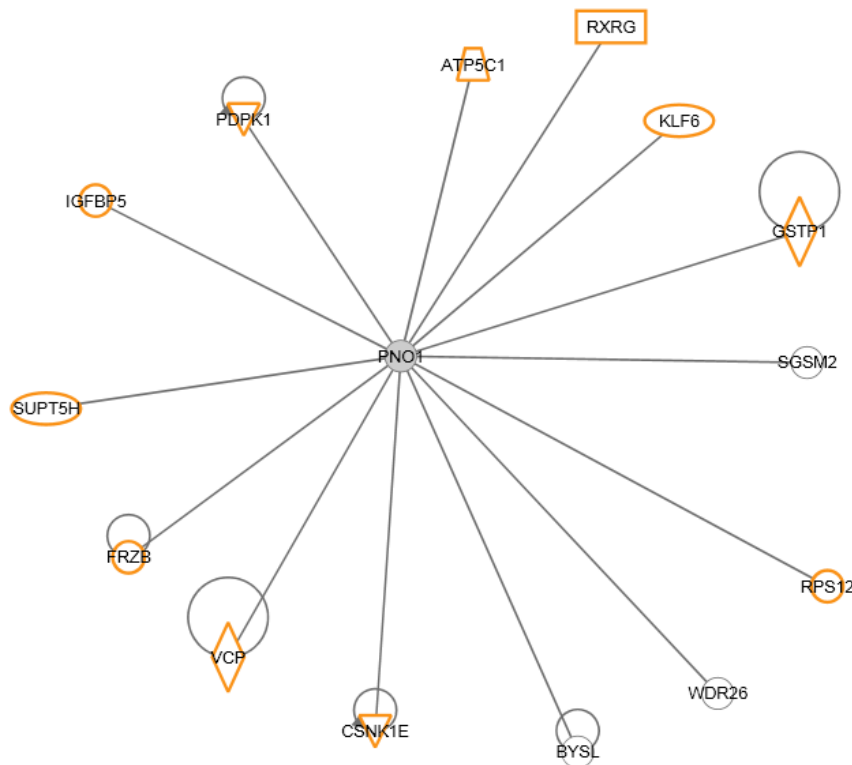
CD3E encoded the CD3-epsilon polypeptide forms the T-cell receptor-CD3 complex which plays an important role in coupling antigen recognition to several intracellular signal-transduction pathways

Potential role in cancer: Unknown

Gene function and role in cancer

PNO1

Function: partner of NOB1 homolog, function unknown



Have protein interaction with several cancer related gene

Gene function and role in cancer

CYC1

Function: cytochrome c-1, related with electron transport chain in mitochondrial inner membrane

Potential role in cancer: Unknown

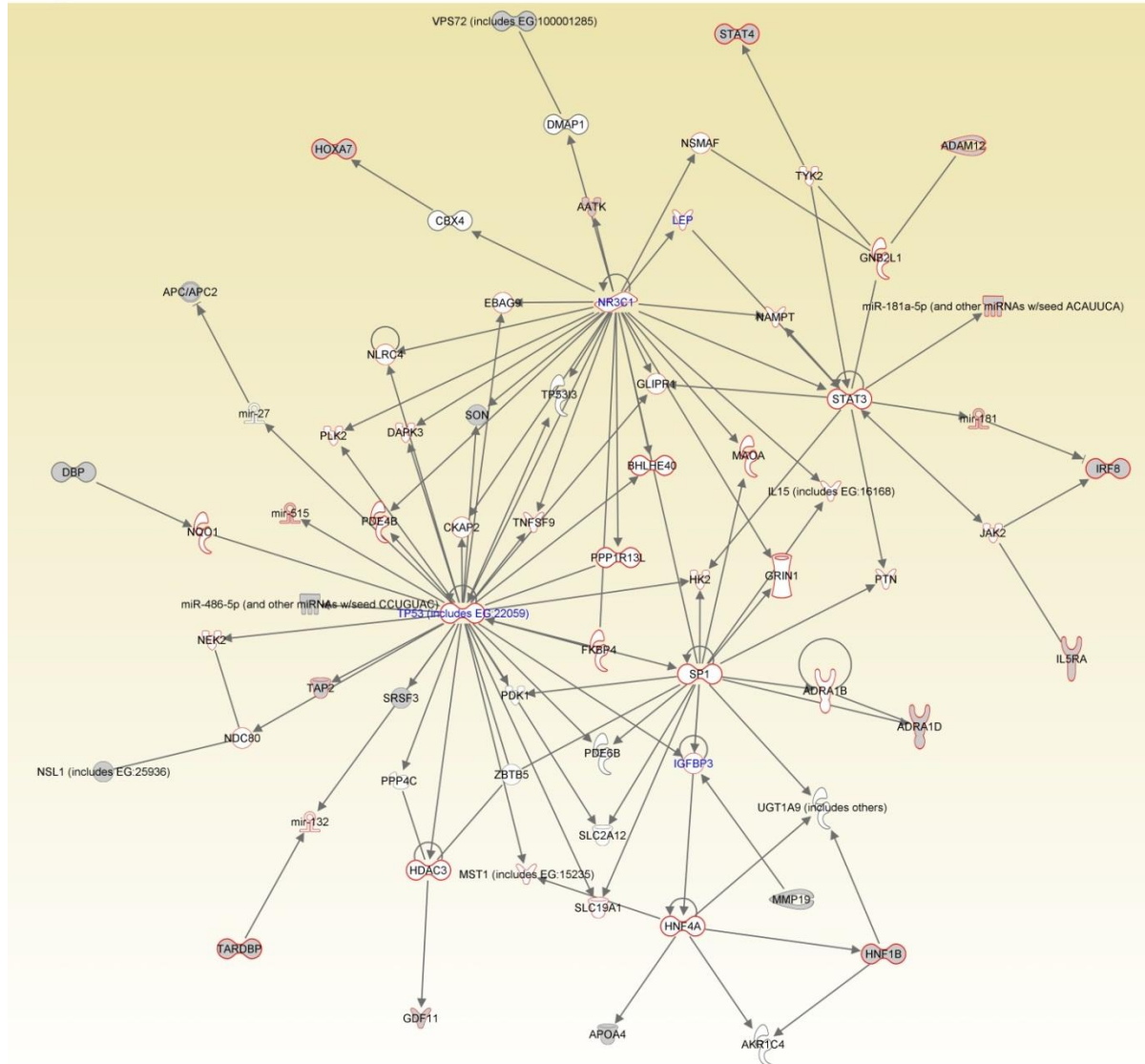
Gene function and role in cancer

TAP2

Function: Protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters.

Potential role in cancer: TAP2 allele is predictors of cervical cancer risk

Gostout BS et.al 2003 Gynecol Oncol



TAP2 has interaction with p53 in the subnetwork3