



G-DOC – Enabling Systems Medicine through Innovations in Informatics

PATIENT CARE
RESEARCH
EDUCATION
COMMUNITY

Subha Madhavan, Ph.D.
Director
Innovation Center for Biomedical Informatics
Georgetown University Medical Center

CBIIT Speaker Series

July 11, 2012



A Comprehensive Cancer Center Designated
by the National Cancer Institute

<http://lombardi.georgetown.edu>
Lombardi CancerLine: 202.444.4000



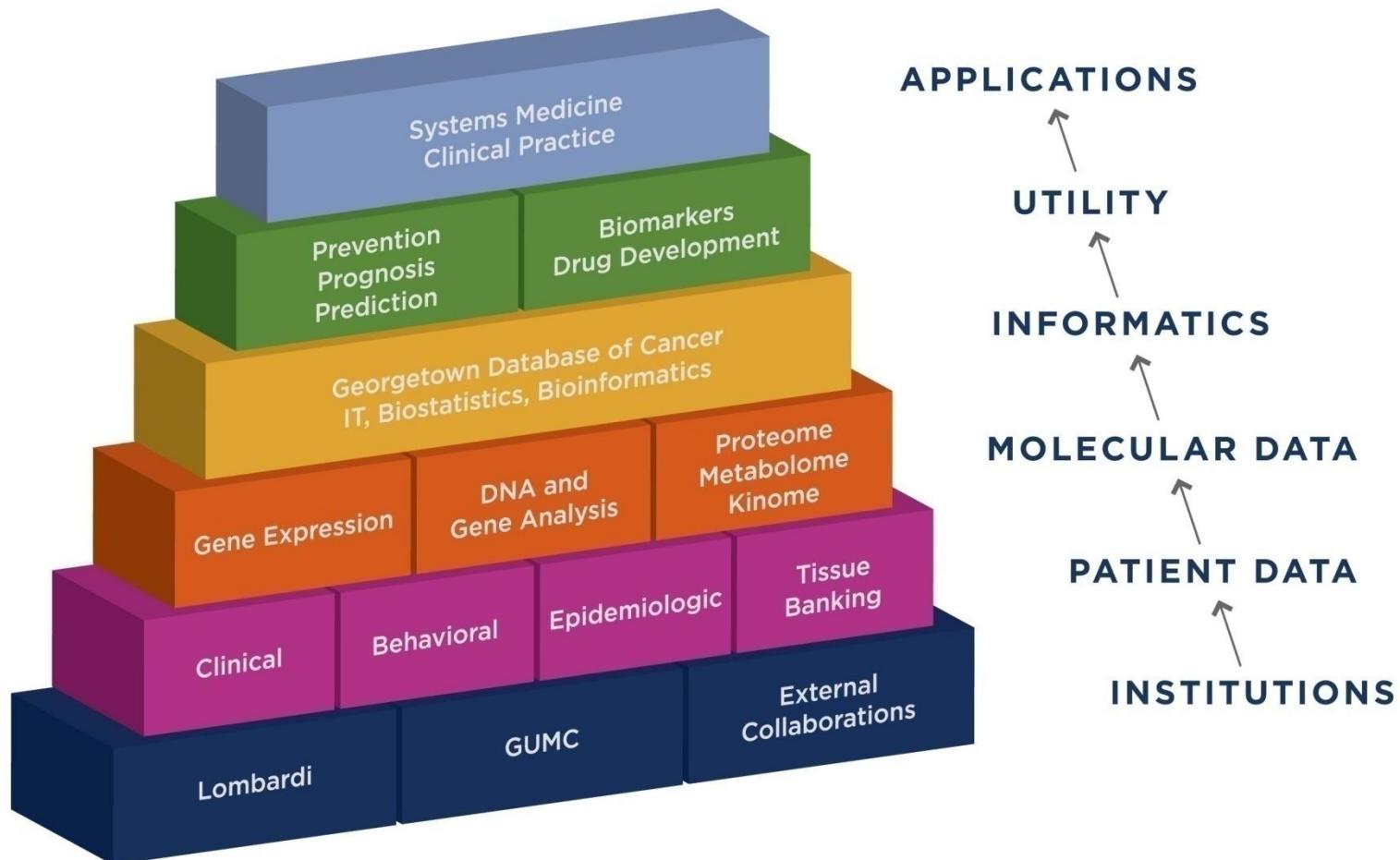
Systems Medicine Defined

- The new and emerging field of Systems Medicine, an **application of Systems Biology approaches to biomedical problems** in the clinical setting, leverages complex computational tools and high dimensional data to derive personalized assessments of disease risk.
- Systems Medicine offers the potential for more **effective individualized** diagnosis, prognosis, and treatment options.
- Achieving this goal requires the effective use of **petabytes of data**, which necessitates the development of new types of tools.

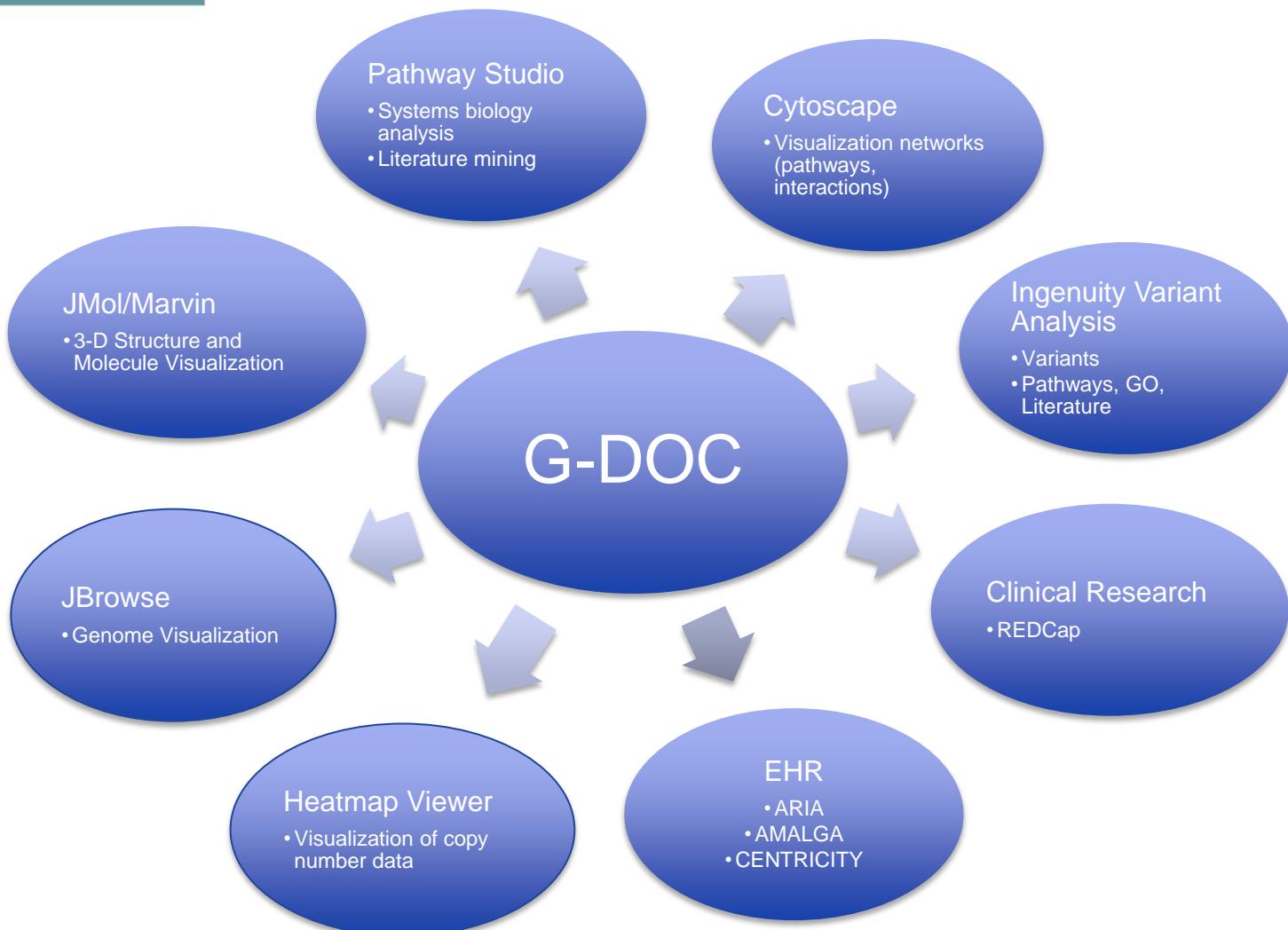
Driving Factors

- Information continuum (care -> research -> back to care): Connect the dots
- Incorporation of “omics-based evidence” in Clinical Research and in Care settings (EHRs, PHRs)
- Collect data once and use it multiple times – clinical care, secondary use for research
- Connect research platforms to accelerate scientific discovery and validation progress
- Efficiently utilize molecular and clinical information to ultimately transform patient care

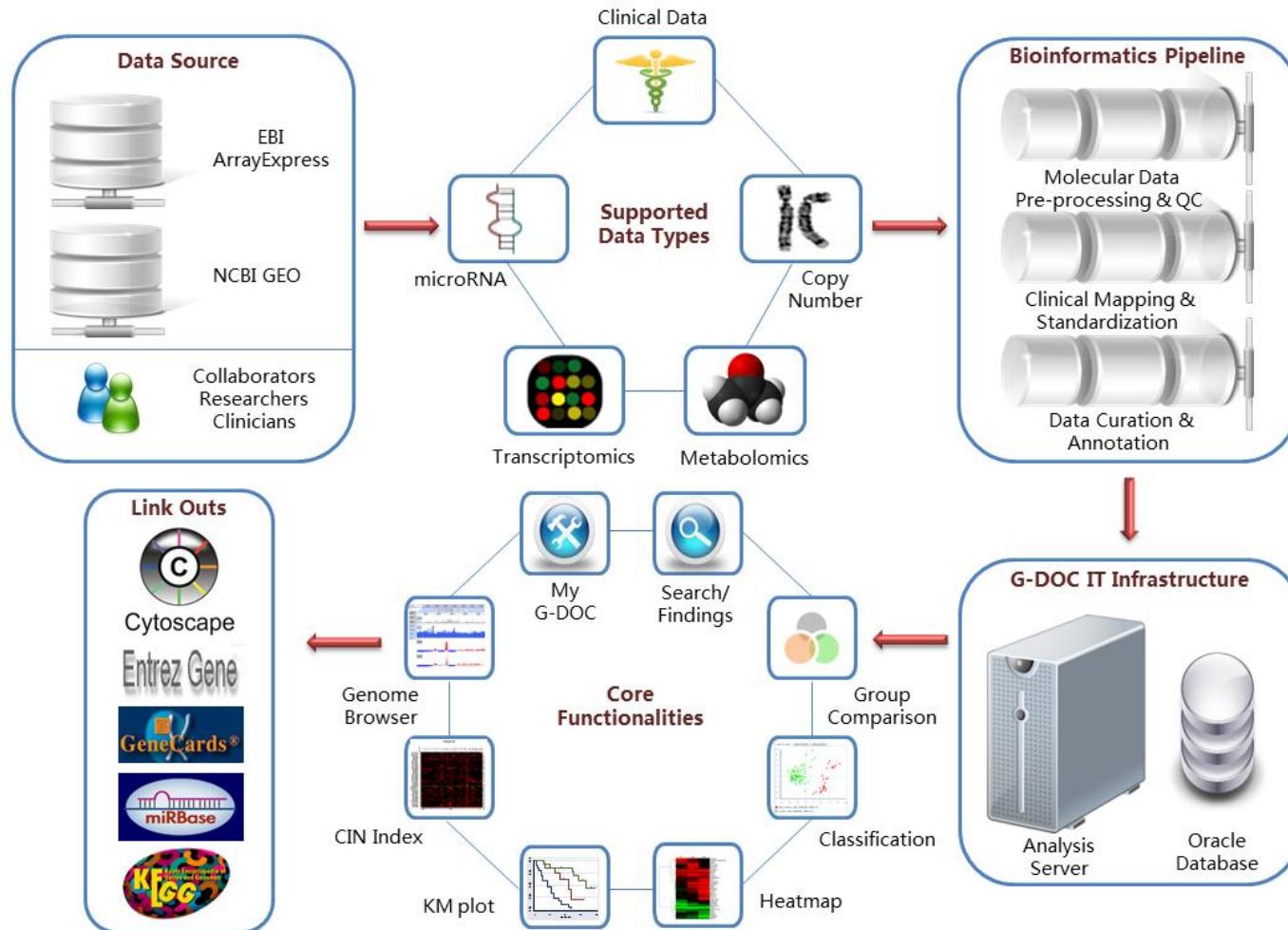
Vision For Georgetown Database of Cancer (G-DOC)



G-DOC Suite Of Tools



G-DOC Modular Architecture



Overview **Search** **Analyze** **My G-DOC**

WELCOME

The Georgetown Database of Cancer (G-DOC) is a cutting-edge data integration platform and knowledge discovery system for the oncology and translational research communities. G-DOC users can access public and proprietary clinical and -omics data aggregated from across the Medical Center, along with a comprehensive set of advanced analysis and visualization tools, to generate and test hypotheses across biomedical disciplines.

**Cancer/Study Overview**

Disease	Study Count	Patient Count	Biospecimen Count	Available Data Types
BREAST CANCER	19	3319	3690	   

FINDINGS

List of one potential new target for treating a subset of hepatocellular carcinomas based on Chiang, et al. 2008 - loaded on: [Wed Aug 17, 2011](#)

NEWS

List of genes that can be used to classify G2 breast tumors into G2a (low grade) and G2b (high grade) subsets, which are similar in survival

PUBLICATIONS

Cancer/Study Overview				
Disease	Study Count	Patient Count	Biospecimen Count	Available Data Types
BREAST CANCER	19	3319	3690	
COLON CANCER	9	662	1125	
LIVER CANCER	3	280	468	
PANCREATIC CANCER	1	52	51	
STOMACH CANCER	1	197	165	
TOTAL	33	4510	5499	

FINDINGS

NEWS	List of one potential new target for treating a subset of hepatocellular carcinomas based on Chiang, et al. 2008 - loaded on: Wed Aug 17, 2011
PUBLICATIONS	List of genes that can be used to classify C2 breast tumors into C2a (low grade) and C2b (high grade) subsets, which are similar in survival outcome to G1 and G3 tumors, respectively; validated in 2 separate cohorts based on Ivshina, et al. 2006 - loaded on: Wed Aug 17, 2011
	List of genes which are associated with predicting metastatic colon cancer patient respondents to Cetuximab (anti-EGFR) based on Khambata-Ford, et al. 2007 - loaded

[Home](#)[G-DOC ® Studies](#)[Search ▶](#)[Analyze ▶](#)[Help](#)[Notifications](#)[Saved Lists](#)[Saved Analyses](#)[Manage my groups / Request access](#)

ⓘ Welcome back, your last login was Tue Apr 3, 2012. You can check if you have been granted access to new [lists](#) or [analyses](#) since your last login

 search G-DOC®

(enter published findings ⓘ genes, proteins, cancer type, studies, investigators, authors ...)

Getting Started with G-DOC

Most users prefer to start using G-DOC to compare and analyze how groups of subjects within a cancer study differ, either by attributes or via various 'omics' characteristics. The typical process of searching for unique lists before analysis is done for you in the Quick Start feature below.



Quick Start

Let G-DOC organize subjects (patients, cell line, animal models) by cancer type, enabling you to stratify 2 groups by outcome or experimental design (e.g Relapse, Treated vs. Non-treated) and quickly take you to the next step of performing an analysis.



Tutorials

Watch step-by-step movies of workflows that are available within the G-DOC application. More instruction is also available in the [help section](#).

Features

Search

[Browse Genome](#)[Compounds/Drug Targets](#)[Findings](#)[Gene Expression](#)[Studies](#)

Analyze

[Chromosomal Instability Index](#)[Classification](#)[Group Comparison](#)[HeatMap Viewer](#)[KM Clinical Plots](#)[KM Gene Expression Plots](#)

My G-DOC

[Notifications](#)

[Home](#)[G-DOC ® Studies](#)[Search ▶](#)[Analyze ▶](#)[Help](#)[Notifications](#)[Saved Lists](#)[Saved Analyses](#)[Manage my groups
/ Request access](#)

Studies

My Studies						
Study Name	Id	Description	Principal Investigator(s)	Disease	Subject Matter	Point(s) of Contact
BRC_CLARKE_9999_01	222	Breast cancer cell line from Clarke Lab	Robert Clarke, PHD,DSC	BREAST CANCER	CELL LINE	Ayesha Shajahan
BRC_CLARKE_LIU_9999_01	149	Clarke-Liu Data Set	Minetta Liu, MD Robert Clarke, PHD,DSC	BREAST CANCER	PATIENT	Rebecca Riggins
BRC_DESMEDT_2007_01	144	Strong Time Dependence of the 76-Gene Prognostic Signature	Public Data Source,	BREAST CANCER	PATIENT	Public Data Source
BRC_DESMEDT_2009_01	145	OGI: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1-98 trial	Public Data Source,	BREAST CANCER	PATIENT	Public Data Source
BRC_FINAK_2008_01	167	Tumor-associated stroma derived from primary clinical breast cancer samples	Public Data Source,	BREAST CANCER	PATIENT	Public Data Source
BRC_FINETTI_2009_01	146	Molecular profiling of ERBB2-amplified breast cancers	Public Data Source,	BREAST CANCER	PATIENT	Public Data Source
BRC_LIN_2007_01	223	Timecourse of estradiol (10nM) exposure in MCF7 breast cancer cells	Public Data Source,	BREAST CANCER	CELL LINE	Public Data Source
BRC_LOI_2008_01	141	Molecular profiling with ER and tamoxifen status	Public Data Source,	BREAST CANCER	PATIENT	Public Data Source

[Home](#)[G-DOC ® Studies](#)[Search](#) ▶[Analyze](#) ▶[Help](#)[Notifications](#)[Saved Lists](#)[Saved Analyses](#)[Manage my groups
/ Request access](#)

BRC_CLARKE_LIU_9999_01 Details

Study Details

Study Name	BRC_CLARKE_LIU_9999_01 (Id:149)
Study Abstract	Toxicity from endocrine therapy is usually more tolerable than from cytotoxic chemotherapy, and the proportional benefits are higher for postmenopausal women with ER+ disease. ER and/or PgR status are useful predictors of responsiveness to endocrine agents but treatment failure is seen in about 50% of cases. Partly for this reason, sequential combination chemotherapy and endocrine therapy may be prescribed. The ability to predict endocrine responsiveness more accurately could decrease the need for chemotherapy in many patients. A predictor that could further direct specific endocrine therapies (e.g., antiestrogen vs. aromatase inhibitor), i.e., identify individual patients with either a very high or very low risk of responding to one or the other drug, also would have widespread use.
Principal Investigator(s)	Minetta Liu, MD Robert Clarke, PHD,DSC
Disease	BREAST CANCER
Point(s) of Contact	Rebecca Riggins

Data Type Details

MICROARRAY CLINIC		
Data-Type	Number of Elements	Search
Clinical Data	11 Clinical Elements	Search

[Home](#)[G-DOC ® Studies](#)[Search ▶](#)[Analyze ▶](#)[Help](#)[Notifications](#)[Saved Lists](#)[Saved Analyses](#)[Manage my groups
/ Request access](#)

Input Search | Molecule 'Sketch' Search

Enter name for a gene, protein, molecule:

< molecular weight <

[reset](#) | [search](#)

Page: 1 [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next >](#)

NAME: No name available for the molecule at this time.

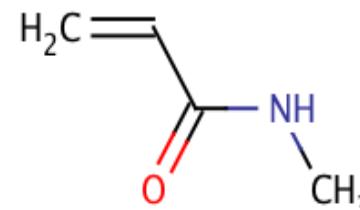
This compound is accessible to PUBLIC



Targets: [EGFR](#)
(EGFR)

Property	Value
Formula	
Molecular Weight	85.1045
Refractivity	
Solubility	
pH	
EC50 [nM]	
IC50 [nM]	

Property	Value
Donor Atoms	1
AcceptorAtoms	1
Clog P	-0.583
Rotatable Bonds	0
ED50 [nM]	
Other Assay	
Chiral	



NAME: No name available for the molecule at this time.

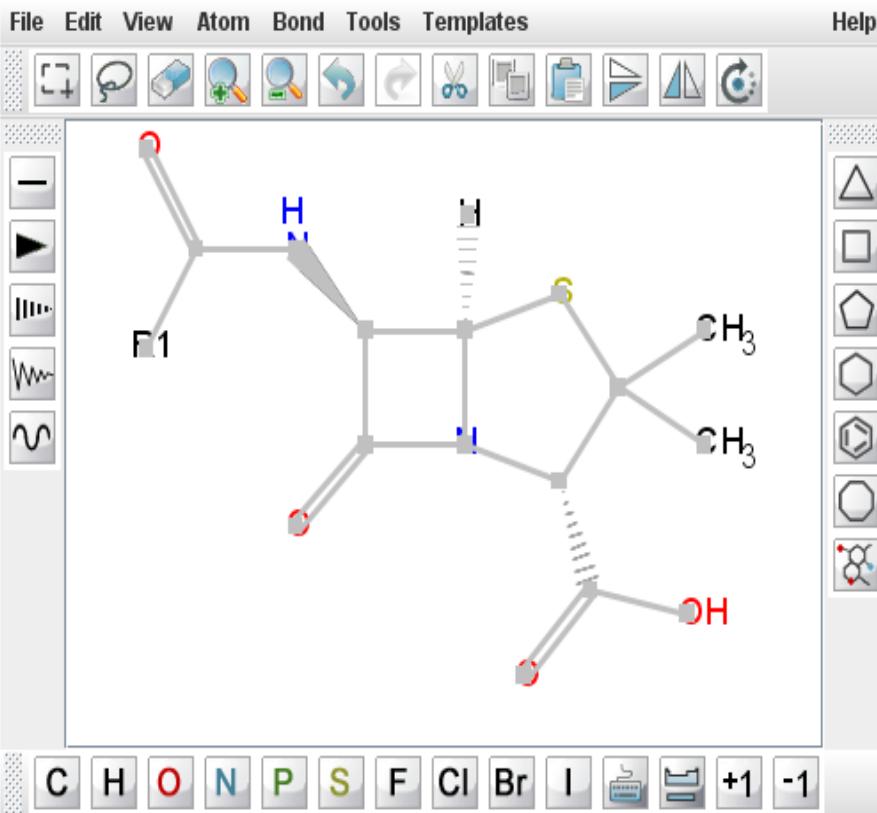
This compound is accessible to PUBLIC



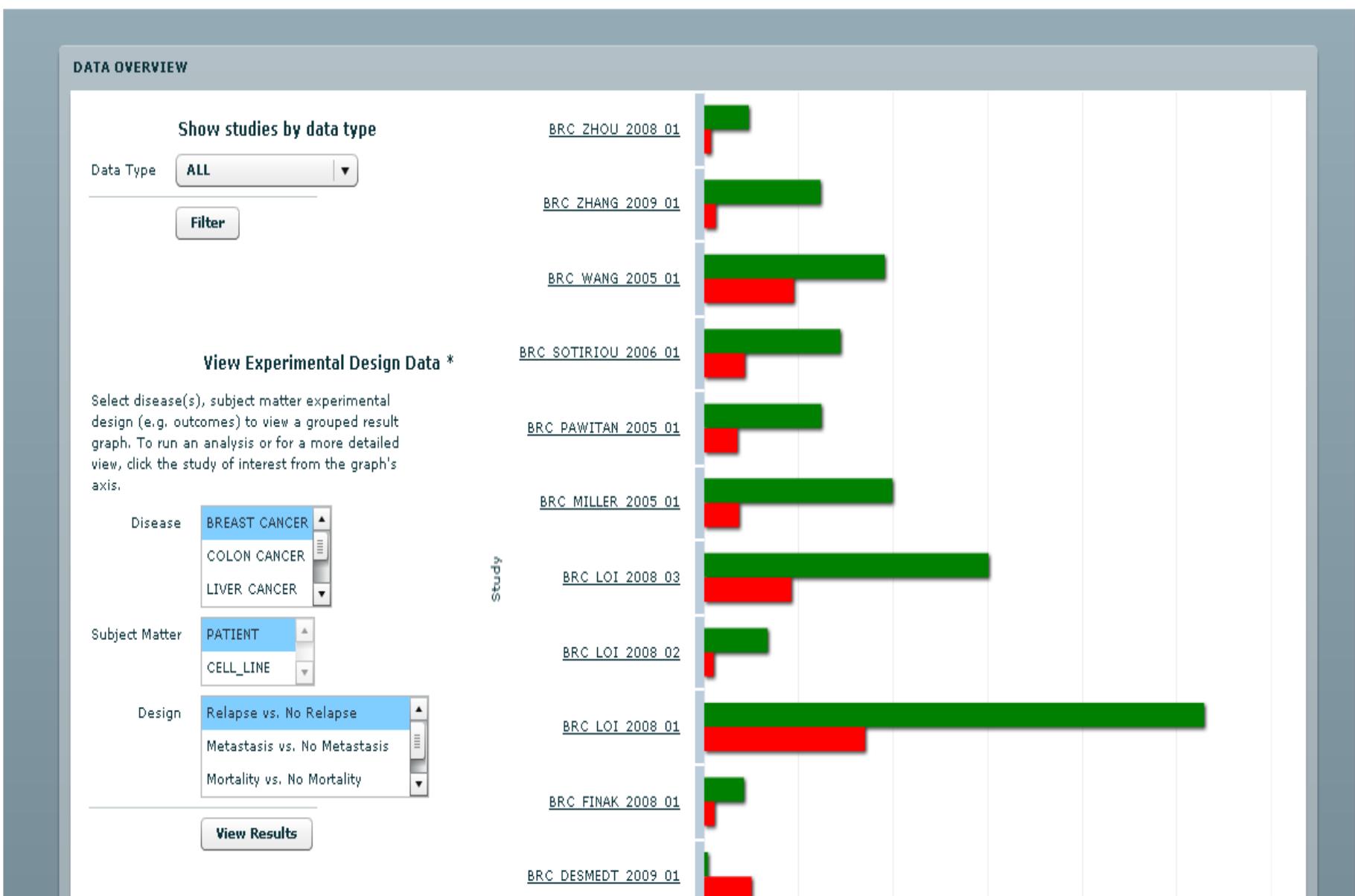
Targets: [EGFR](#)
(EGFR)

[Home](#)[G-DOC® Studies](#)[Search ▶](#)[Analyze ▶](#)[Help](#)[Notifications](#)[Saved Lists](#)[Saved Analyses](#)[Manage my groups / Request access](#)

Molecule 'Sketch' Search | [Input Search](#)



G-DOC ® Quick-Start



G-DOC ® Quick-Start

DATA OVERVIEW

Show studies by data type

Data Type: ALL

Filter

View Experimental Design Data *

Select disease(s), subject matter experimental design (e.g. outcomes) to view a grouped result graph. To run an analysis or for a more detailed view, click the study of interest from the graph's axis.

Disease: BREAST CANCER

Subject Matter: PATIENT

Design: Relapse vs. No Relapse

View Results

BRC_ZHOU_2008_01

BRC_ZHANG_2009_01

BRC_WANG_2005_01

BRC_SOTIRIOU_2006_01 options:

Analyze these 2 groups:

- [Chromosomal Instability Index](#)
- [Classification](#)
- [Group Comparison](#)
- [HeatMap Viewer](#)
- [KM Clinical Plots](#)
- [KM Gene Expression Plots](#)

View report:

- [-No Relapse](#)
- [-Relapse](#)

[Take me to report page for BRC_SOTIRIOU_2006_01](#)

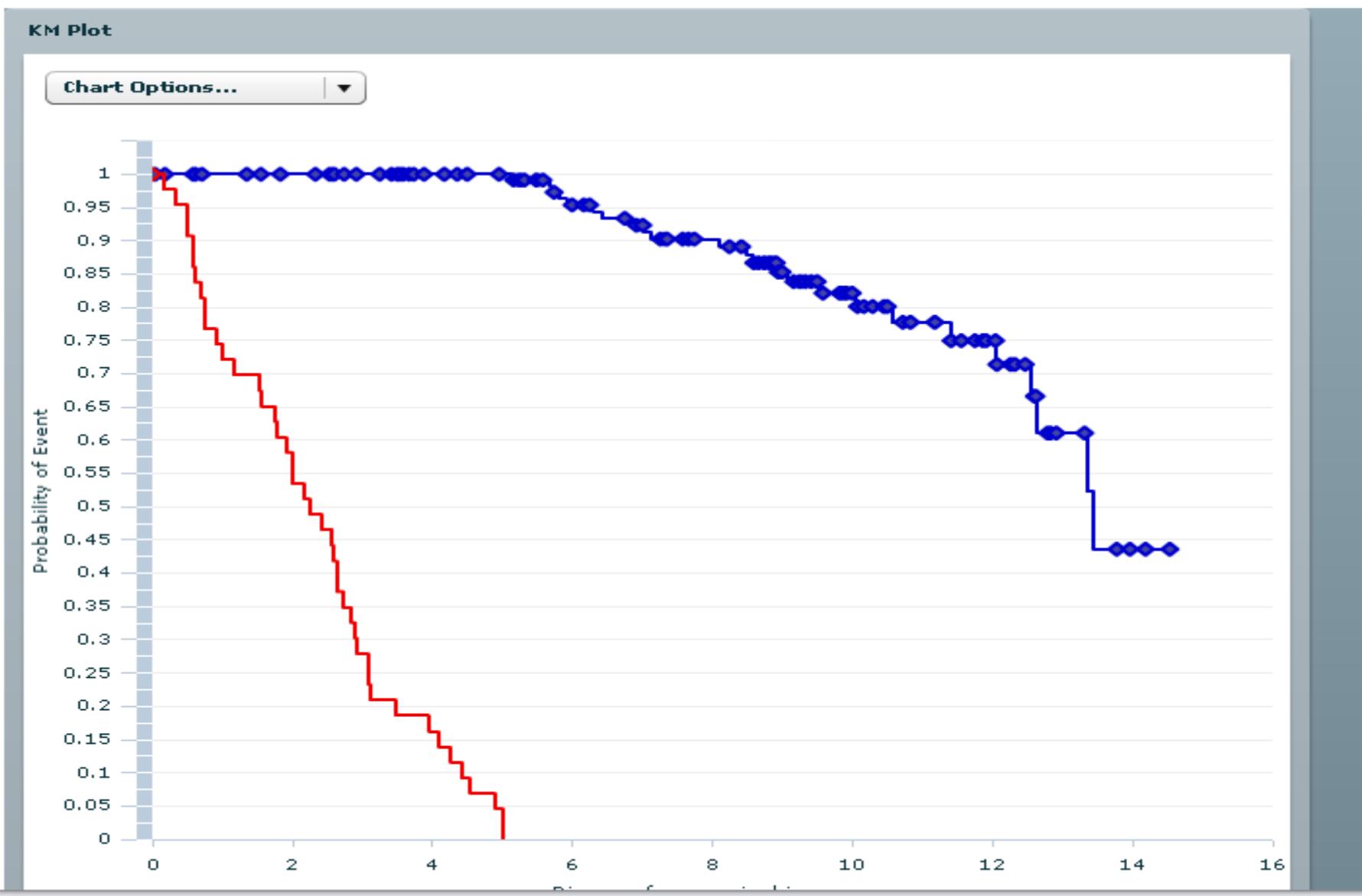
BRC_SOTIRIOU_2006_01

BRC_FINAK_2008_01

BRC_DESMEDT_2009_01

KM Plot Results

Current Study: BRC_SOTIRIOU_2006_01



Gene Expression KM Plot

Reporter: 202704_at, Fold Change: 1

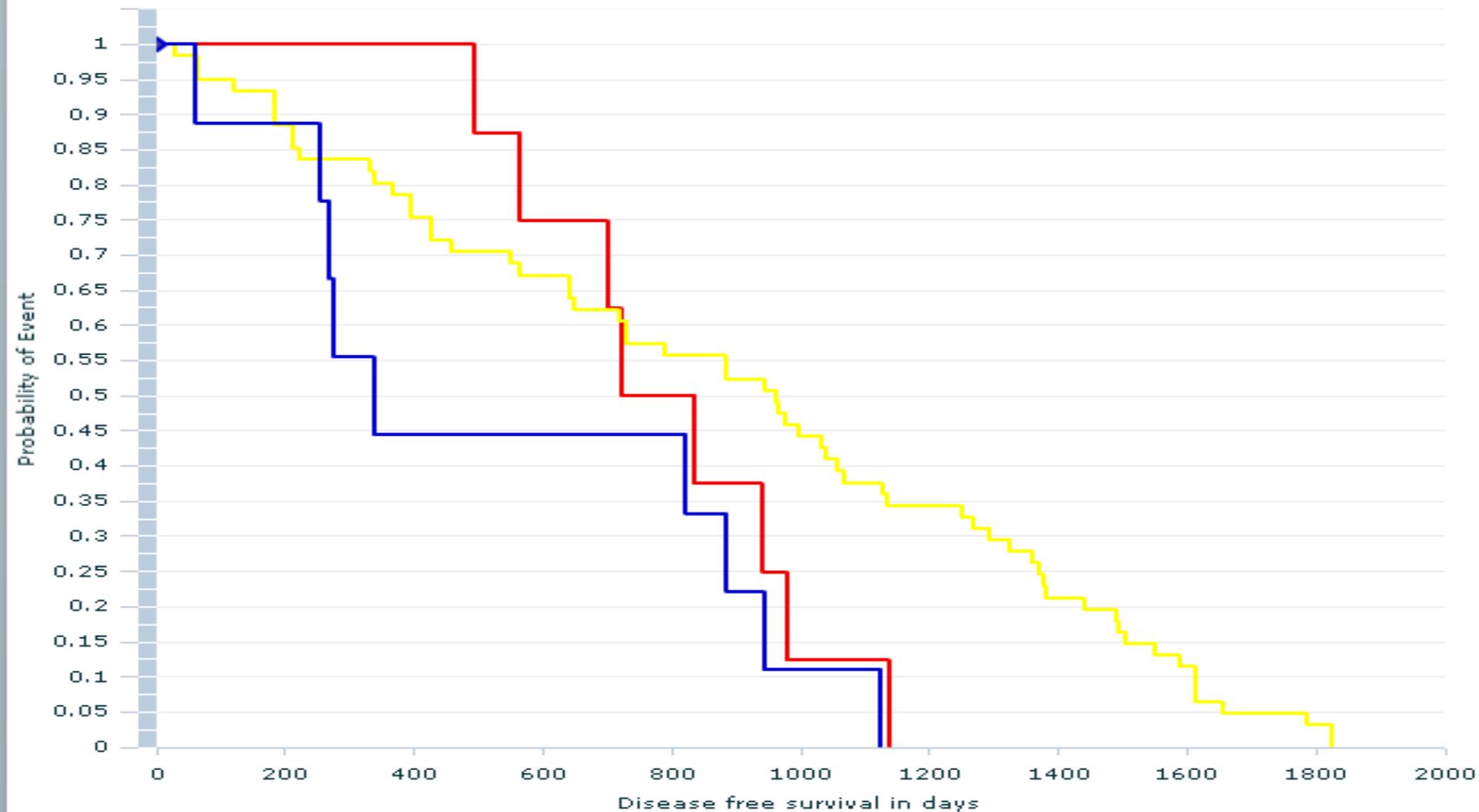
Fold Change

2

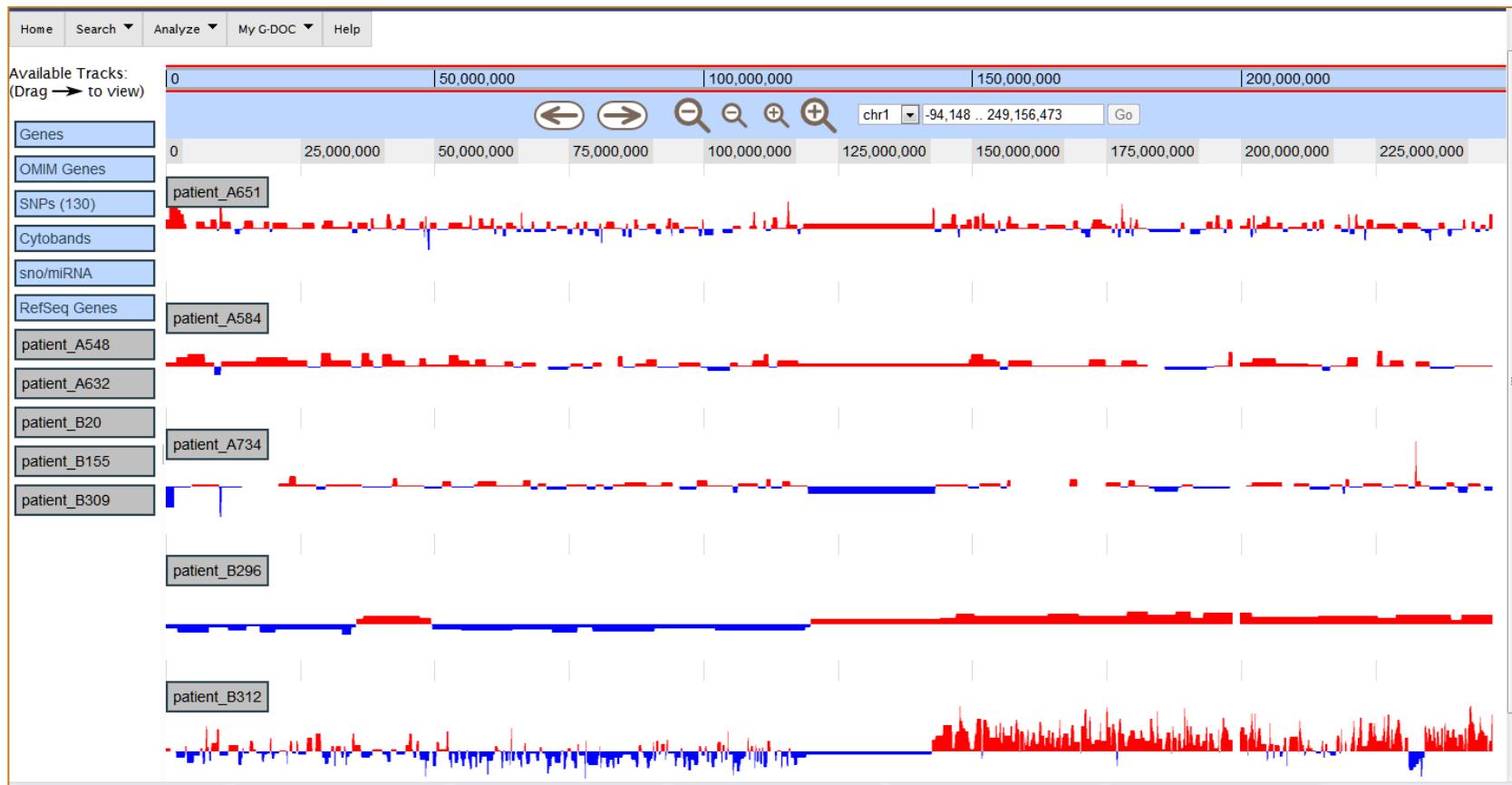
Reporters

202704_at | ▾

[Redraw plot](#)



DNA Copy Number Segments, Chr1



Correlate Abnormality/Event With Clinical Parameters

Lombardi Comprehensive Cancer Center
at Georgetown University

Logged in as: acs224
Logout

search gdoc

AGE_CASE_START: 77
CHEMOTHERAPY: Yes
CHEMO_REGIMENT: Folic acid, 5-Fluorouracil, Irinotecan, Xelodan
GENDER: FEMALE
PRIMARY_DISEASE: malignant neoplasm of sigmoid colon
PTNM_M: M0 - No distant metastasis
PTNM_N: N0 - No regional lymph node involvement
PTNM_N_POSITIVE: 0
PTNM_N_TOTAL: 30
PTNM_T: T3 - Size and/or extent of the primary tumor
RELAPSE: YES
SURGERY_TO_RELAPSE/FU: 1.3
TUMOR_DIGNITY: MALIGNANT
TUMOR_STAGE: Stage II
VITAL_STATUS: ALIVE

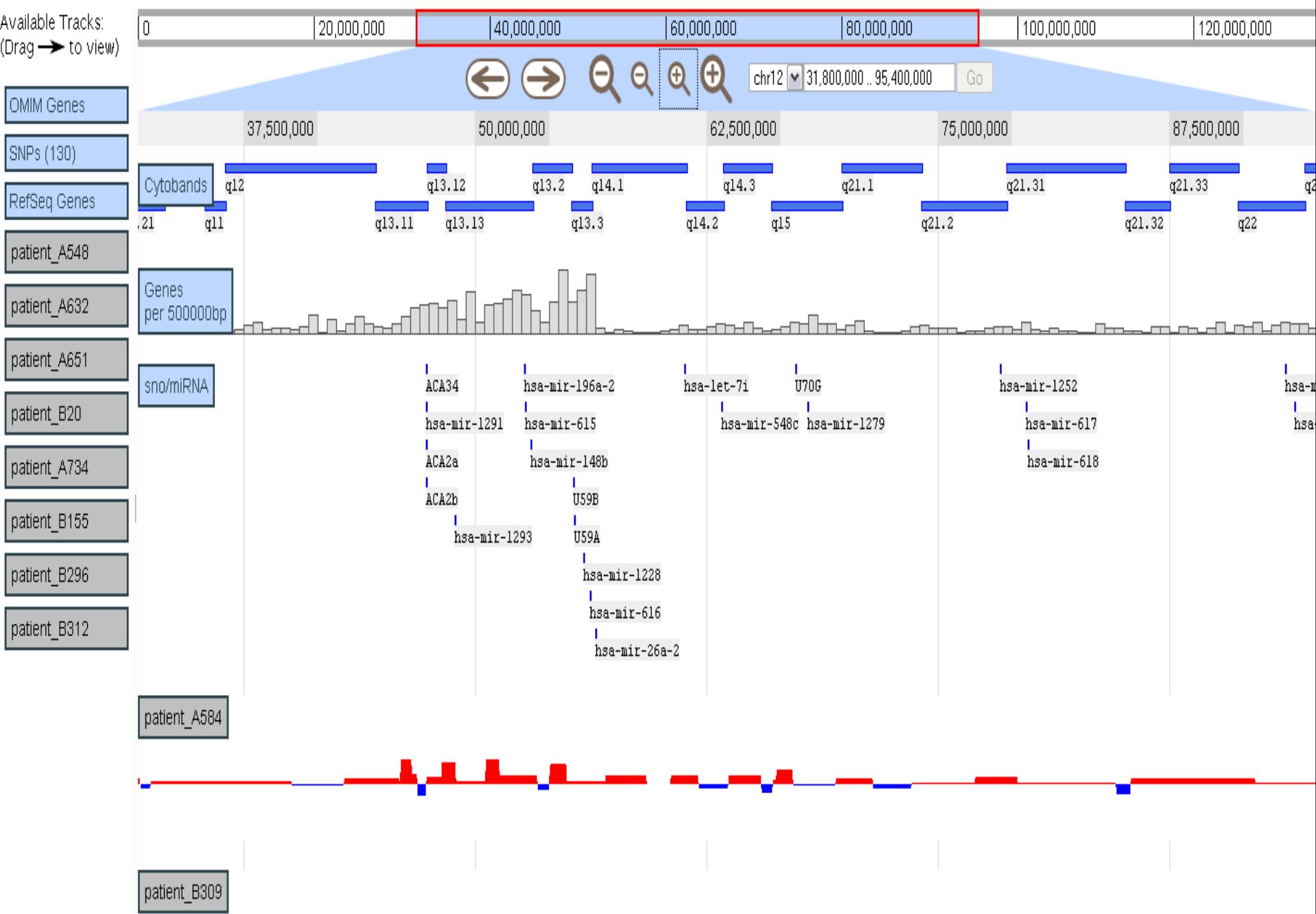
Available Tracks:
(Drag → to view) 0

OMIM Genes
RefSeq Genes
sno/miRNA
SNPs (130)
Patient 113769
Patient 113767
Patient 113773
Patient 113771
Patient 113775
Patient 113777
Patient 113765
Patient 113779

Cytobands
Genes

80,000,000 | 100,000,000 | 120,000,000
25,183,604 .. 25,381,604 Go
25,300,000 | 25,350,000

The screenshot shows the gdoc software interface. On the left, a sidebar lists various genomic tracks: OMIM Genes, RefSeq Genes, sno/miRNA, SNPs (130), Patient 113769, Patient 113767, Patient 113773, Patient 113771, Patient 113775, Patient 113777, Patient 113765, and Patient 113779. The 'Genes' track is currently selected. A large central panel displays clinical parameters for a specific case: AGE_CASE_START (77), CHEMOTHERAPY (Yes), CHEMO_REGIMENT (Folic acid, 5-Fluorouracil, Irinotecan, Xelodan), GENDER (FEMALE), PRIMARY_DISEASE (malignant neoplasm of sigmoid colon), PTNM_M (M0 - No distant metastasis), PTNM_N (N0 - No regional lymph node involvement), PTNM_N_POSITIVE (0), PTNM_N_TOTAL (30), PTNM_T (T3 - Size and/or extent of the primary tumor), RELAPSE (YES), SURGERY_TO_RELAPSE/FU (1.3), TUMOR_DIGNITY (MALIGNANT), TUMOR_STAGE (Stage II), and VITAL_STATUS (ALIVE). To the right of the clinical parameters is a genomic track viewer showing chromosomes 12 and 13 with specific genomic coordinates highlighted in blue (25,183,604 .. 25,381,604) and a zoomed-in view of chromosome 12 with coordinates 25,300,000 and 25,350,000. The top right corner shows the user is logged in as 'acs224'.



40 CRC Patients, Stage 2, >10 Years Follow-up (Samples provided by INDIVUMED Inc., Germany)

- 20 Relapse_Free Patients
- Tissue DNA: Tumor – 20; Normal – 20
- Tissue RNA: Tumor – 20; Normal – 20
- Biofluids microRNA: Serum – 20
- Biofluids Metabolites: Serum – 20; Urine – 20
- 20 Relapsed Patients
- Tissue DNA: Tumor – 20; Normal – 20
- Tissue RNA: Tumor – 20; Normal – 20
- Biofluids microRNA: Serum – 20
- Biofluids Metabolites: Serum – 20; Urine – 20
- Clinical Attributes: >100

Bottom Line:

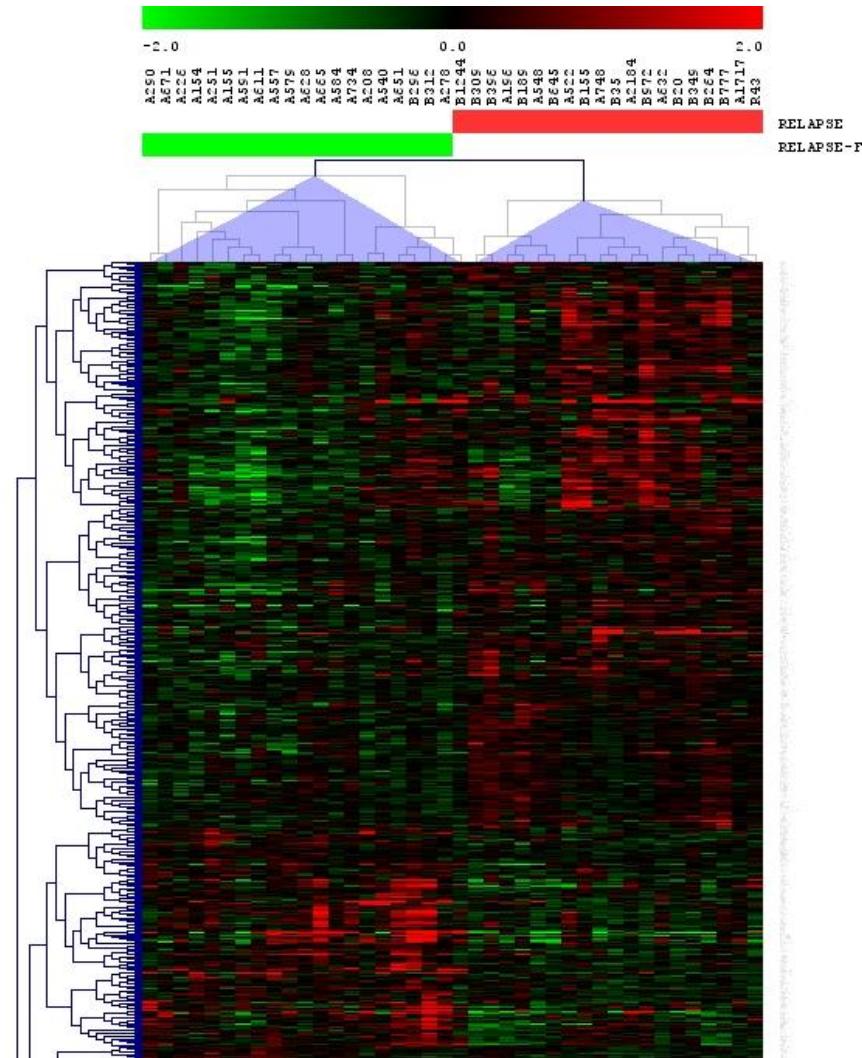
40 CRC patients: 20 with relapse vs. 20 relapse-free

What are the molecular correlates of Relapse?

Sample Type	Differentially Expressed
Tumor DNA CNV	37 cytobands
Tumor RNA Genes	720 reporters
Tumor RNA microRNA	34 microRNAs
Serum microRNA	8 microRNAs
Serum Metabolites	77 peaks
Urine Metabolites	47 peaks
DNA Exome-seq (EdgeBio)	Analysis In Progress

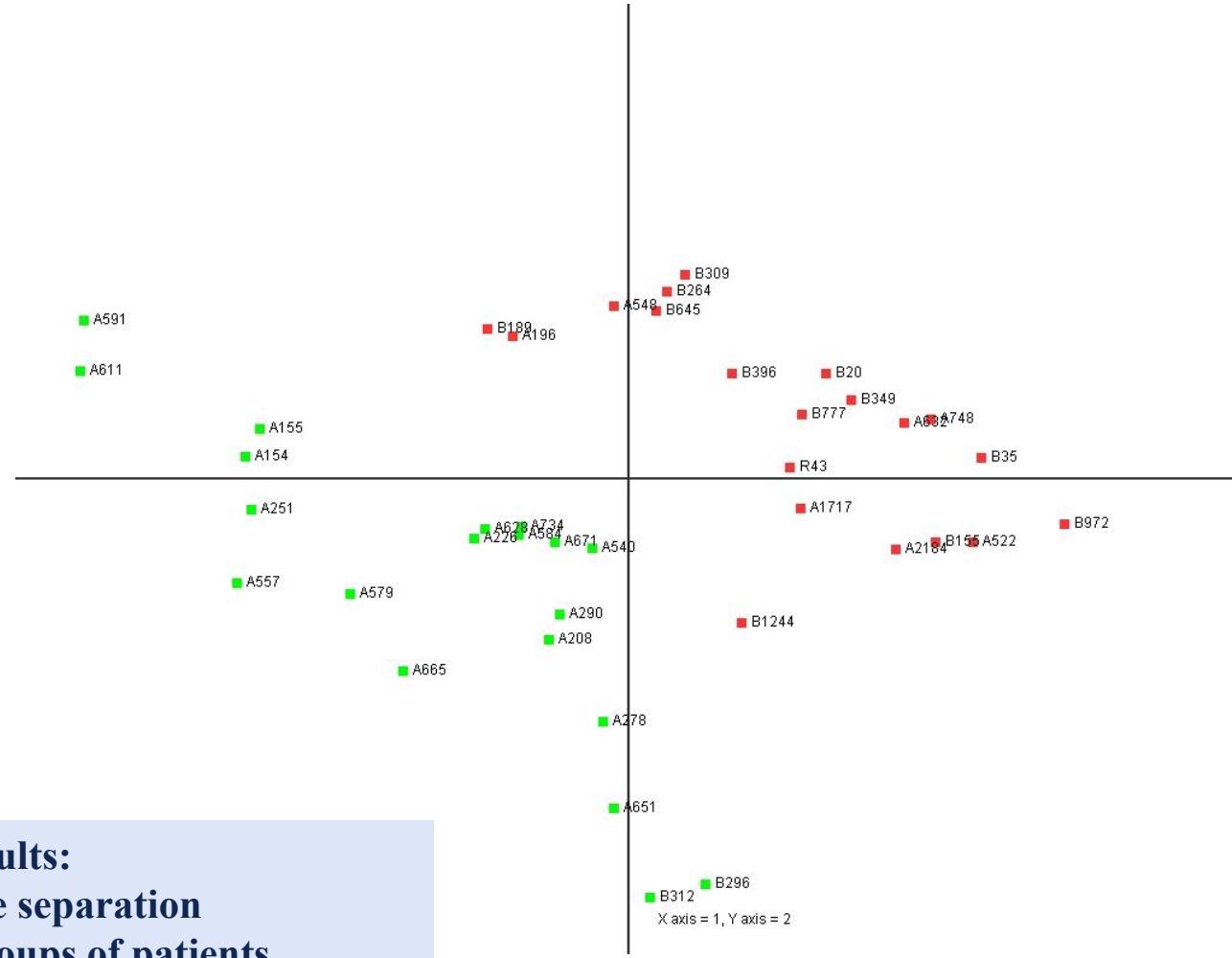
Gene Expression In Tumor Samples: Relapse vs. Relapse free

T-test p<0.05, 720 reporters



Genes In Tumor Samples: Relapse vs. Relapse free

PCA based on T-test p<0.05



PCA Results:
Complete separation
of two groups of patients
with one sample on a borderline

Enrichment Analysis: Diff. Expressed Genes Bio-Functions Most Affected In Relapse Group

Top Bio Functions

Diseases and Disorders

Name	p-value	# Molecules
Inflammatory Response	1.41E-10 - 1.04E-02	86
Infectious Disease	7.57E-10 - 1.04E-02	83
Gastrointestinal Disease	1.04E-09 - 1.03E-02	207
Genetic Disorder	2.54E-07 - 6.91E-03	288
Inflammatory Disease	1.38E-06 - 9.68E-03	155

Molecular and Cellular Functions

Name	p-value	# Molecules
Cell-To-Cell Signaling and Interaction	1.41E-10 - 1.04E-02	91
Antigen Presentation	3.04E-08 - 7.90E-03	45
Cellular Growth and Proliferation	1.01E-06 - 1.04E-02	122
Cellular Movement	1.67E-06 - 1.05E-02	86
Cellular Development	2.18E-06 - 1.04E-02	115

Physiological System Development and Function

Name	p-value	# Molecules
Hematological System Development and Function	1.41E-10 - 1.04E-02	98
Immune Cell Trafficking	1.41E-10 - 1.04E-02	67
Tissue Development	4.58E-08 - 1.04E-02	41
Humoral Immune Response	1.01E-06 - 1.03E-02	29
Tissue Morphology	1.20E-06 - 1.03E-02	68

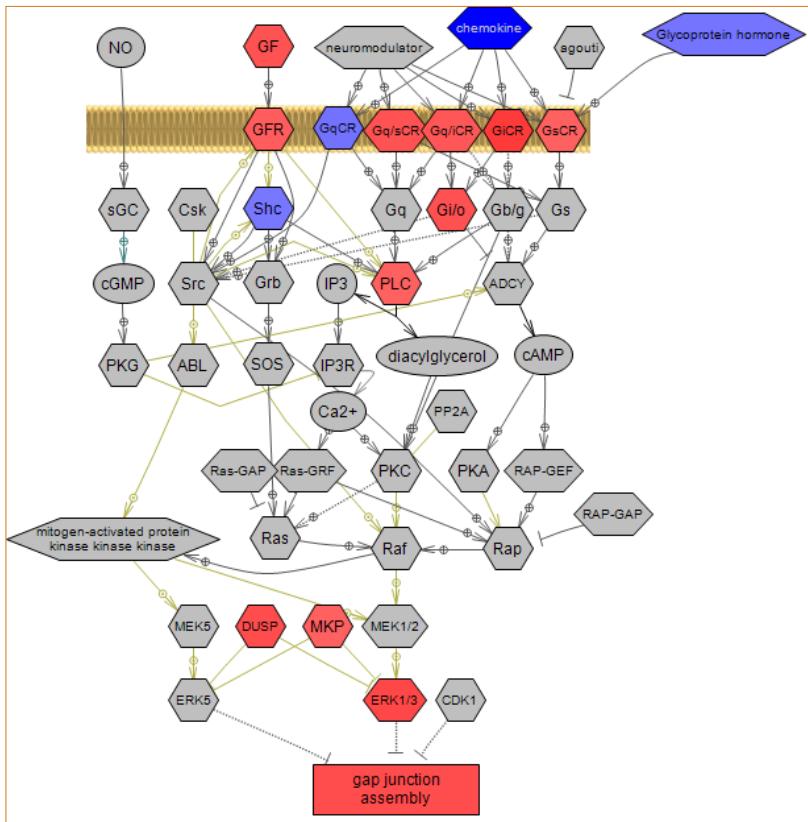
Detailed Pathway/Sub-network Analysis in Pathway Studio

Top 20 Pathways				
Name	p-value	Total Entities	Overlap	
Gap Junction Regulation	0.0065742	51	32	
EGFR/ERBB -> STAT signaling	0.0234547	20	3	
IL10R -> STAT signaling	0.024255	8	2	
IGF1R -> STAT signaling	0.0305485	9	2	
CSF3R -> STAT signaling	0.0305485	9	2	
IL7R -> STAT signaling	0.0305485	9	2	
EGFR -> ZNF259 signaling	0.0374082	10	2	
Translation Control	0.0414135	86	41	
VEGFR -> STAT signaling	0.0447929	11	2	
Atlas of Signaling	0.0457982	381	193	
EGFR -> CTNND signaling	0.0526631	12	2	
EGFR/ERBB2 -> CTNNB signaling	0.0609814	13	2	
PTPRC -> STAT6 signaling	0.0913066	3	1	
Adipocytokine Signaling	0.107811	52	31	
FcIgER -> NFATC1 signaling	0.108109	13	2	
Apoptosis Regulation	0.116453	69	25	
Purine metabolism	0.128625	155	7	
CCR2/5 -> STAT signaling	0.12898	20	2	

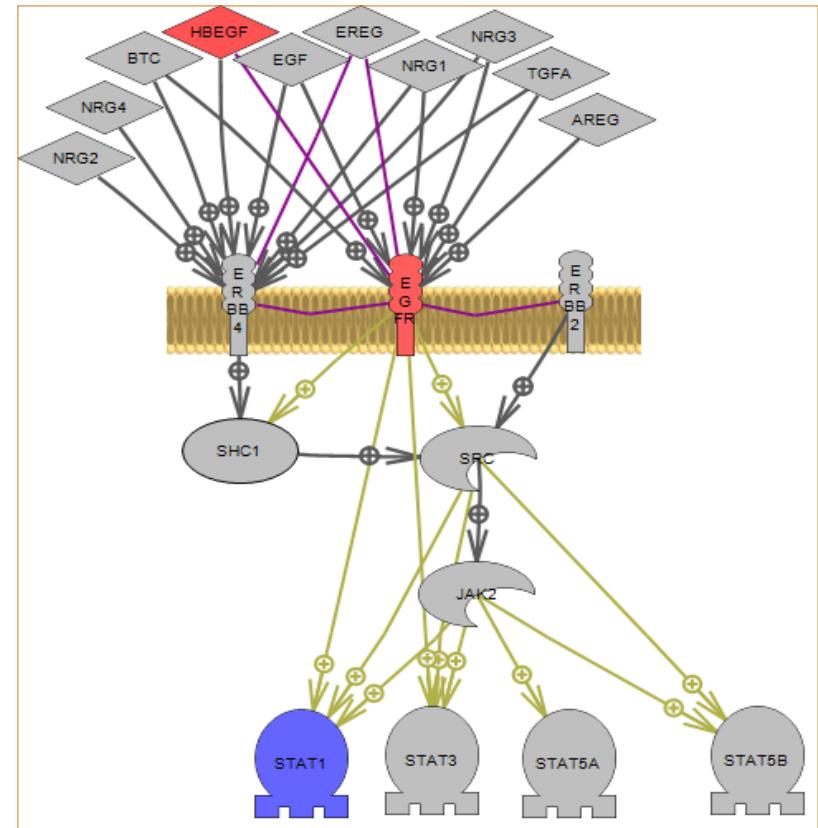
Differentially Expressed Genes : Enrichment Analysis

Pathway Studio: Top Signaling Pathways

Gap Junction Regulation Pathway



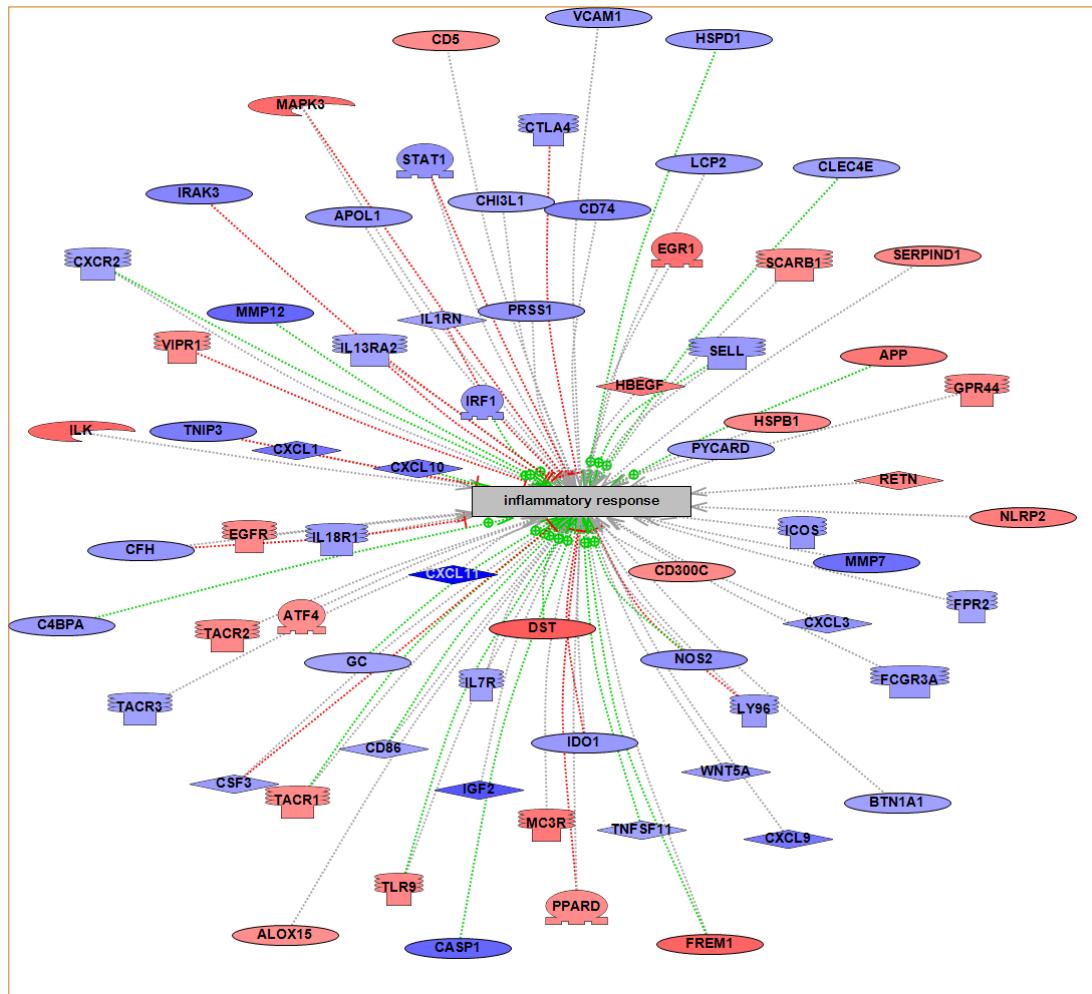
EGFR/ERBB -> STAT signaling



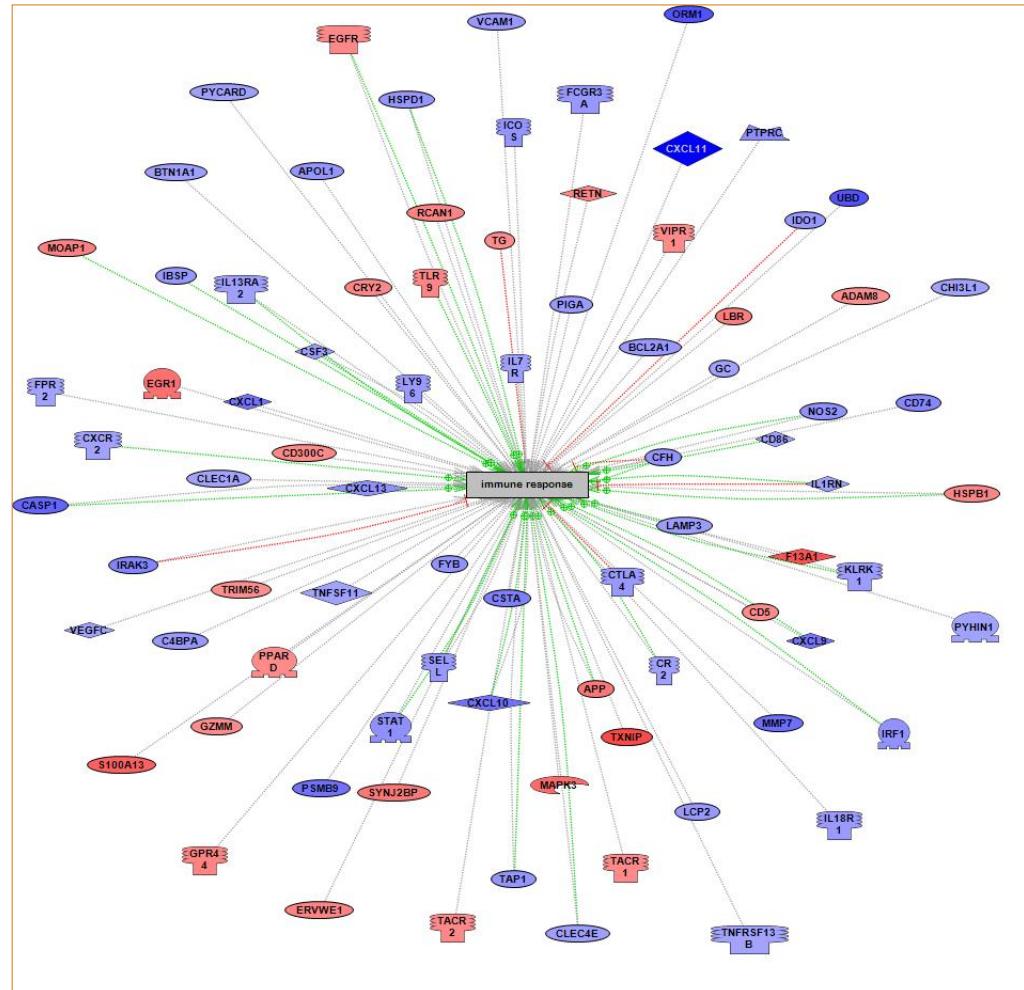
Sub-Network Enrichment: Cell Processes

Gene Set Seed	Overlapping Entities	p-value	Total # of Neighbors	Overlap
inflammatory response	EGFR,VCAM1,TNFSF11,IGF2,MAPK3,TACR3,CD5,NOS2,CSF3,HBEGF,HSPB1,CASP1,APP,RETN,ILK,HSPD1,EGR1,TACR1,CXCL10,CXCL11,CXCR2,SELL,PPARD,WNT5A,IRF1,ALOX15,CXCL1,PYCARD,SCARB1,IL1RN,IL7R,MMP7,TLR9,STAT1,LY96,TACR2,PRSS1,FPR2,ATF4,MMP12,FCGR3A,CFH,ICOS,CTLA4,IDO1,VIPR1,CD86,CXCL9,C4BPA,MC3R,GPR44,IL13RA2,CHI3L1,CXCL3,LCP2,SERPIN1,CD74,NLRP2,APOL1,GC,IRAK3,CLEC4E,BTN1A1,CD300C,TNIP3,DST,IL18R1,FREM1	8.13E-12	1237	68
immune response	TRIM56,EGFR,VCAM1,TNFSF11,MAPK3,TG,CD5,NOS2,CSF3,HSPB1,CASP1,APP,RETN,HSPD1,EGR1,TACR1,CXCL10,PTPRC,CXCL11,CXCR2,SELL,PPARD,IRF1,CXCL1,PYCARD,IL1RN,IL7R,MMP7,TLR9,STAT1,LY96,TXNIP,TACR2,BCL2A1,FP2,VEGFC,FCGR3A,CFH,ICOS,KLRK1,CTLA4,IDO1,ORM1,VIPR1,CD86,CXCL9,C4BPA,RCAN1,GPR44,TNFRSF13B,CXCL13,CRY2,F13A1,IL13RA2,CHI3L1,IBSP,PSMB9,LCP2,PIGA,ADAM8,GZMM,CR2,TAP1,ERVWE1,CD74,LBR,LAMP3,CSTA,UBD,APOL1,PYHIN1,GC,IRAK3,CLEC4E,FYB,BTN1A1,CD300C,S100A13,CLEC1A,MOAP1,IL18R1,SYNJ2BP	2.31E-09	1855	82
apoptosis	TAS2R10,EGFR,VCAM1,TNFSF11,IGF2,MAPK3,TG,CD5,NOS2,CSF3,AQP3,HBEGF,HSPB1,CASP1,ATP2A1,APP,RETN,ILK,LPXN,FOLH1,E2F1,HSPD1,EGR1,TACR1,GJA5,FBXO32,DUSP6,CXCL10,PTPRC,CXCL11,CXCR2,SELL,PPARD,WNT5A,IRF1,ALOX15,CXCL1,PYCARD,SCARB1,DNM1,JRS2,IL1RN,IL7R,MMP7,FOXA2,TLR9,STAT1,TSC2,TXNIP,BCL2A1,ACVR2B,GNAI1,TP53INP1,FPR2,CPE,FAAH,GPX3,ATF4,VEGFC,MMP12,FCGR3A,CFH,ICOS,BTG2,KLRK1,CTLA4,MAOA,IDO1,ORM1,VIPR1,KCNK5,CD86,CXCL9,C4BPA,RCAN1,LMNA,OGDH,CA3,THR8,GP44,SLURP1,SPRY2,GSTT1,TNFRSF13B,CXCL13,BCL2L14,SCIN,NUPR1,MKL1,PCSK9,SHBG,CHI3L1,CXCL3,IBSP,PSMB9,PIGA,RHOH,GZMM,CR2,EPHX1,EPMA2,GMAP4,TAP1,S,MAP6,CSRNP1,NOV,MSRA,ENTPD5,GR1A2,RASSF3,RIN2,CD74,BCL2L10,LBR,TSC22D1,FSCN1,TES,PARVA,TACC1,NEDD9,PRPF31,GPR87,CSTA,MNDA,PLK4,GBP1,SFRP5,SMG1,GALR2,AKR7A2,UBD,PIK3IP1,PACS2,PARP15,S100A6,TIMM8A,FLIP1L,CD3D,APOL1,AIM2,PINK1,MZF1,DEDD2,FAIM2,SMPD3,HAGH,PAFAH2,PPM1A,SALL1,MOAP1,GRIN3A,PTPN7,OIP5,MYCT1,ACER2,DMRT2,FREM1,SERPINB3,SERP2,CTRL,KCTD11,MUC17,DNASE1L1,TMEM109,CHAC1	6.71E-08	5105	165
pregnancy	EGFR,VCAM1,TNFSF11,IGF2,MAPK3,TG,CD5,NOS2,CSF3,AQP3,HBEGF,HSPB1,CASP1,RETN,HSPD1,EGR1,TACR1,GJA5,CXCL10,PTPRC,PPARD,IRF1,SCARB1,IL1RN,TLR9,STAT1,TSC2,FAAH,ATF4,VEGFC,FCGR3A,BTG2,CTLA4,IL11RA,MAOA,IDO1,ORM1,THOP1,SHBG,PSMB9,EPHX1,SERPIN1,ERVWE1,HSD3B1,PECR,GALR2,S100A6,GC,DST,ST3GAL6,KLRC3,STOX1	1.12E-07	1052	52
T-cell response	VCAM1,TNFSF11,TG,CD5,NOS2,CSF3,CASP1,APP,FOLH1,E2F1,HSPD1,EGR1,CXCL10,PTPRC,SELL,PYCARD,IL7R,TLR9,STAT1,TXNIP,VEGFC,FCGR3A,ICOS,KLRK1,CTLA4,IDO1,CD86,CXCL9,RCAN1,TNFRSF13B,TAP1,ENTPD5,CD74,BTN1A1,CLEC1A,MAGEC2	3.30E-07	650	37
T-cell function	TNFSF11,MAPK3,CD5,NOS2,CSF3,E2F1,HSPD1,EGR1,CXCL10,PTPRC,SELL,PYCARD,IL1RN,IL7R,TLR9,STAT1,AHNAK,ICOS,KLRK1,CTLA4,IDO1,VIPR1,KCNK5,CD86,TNFRSF13B,LCP2,CR2,NEDD9,CD3D	8.81E-07	460	29
antigen processing and presentation	TNFSF11,TG,CD5,NOS2,HSPD1,CXCL10,PTPRC,IRF1,DNM1,IL7R,STAT1,FCGR3A,ICOS,CTLA4,IDO1,CD86,THOP1,TNFRSF13B,PSMB9,CR2,TAP1,CD74,FSCN1,UBD,HLA-DMA,HLA-DMB	1.07E-06	388	26
calcium mobilization	EGFR,IGF2,TACR3,CD5,HBEGF,APP,TACR1,CXCL10,PTPRC,CXCL11,CXCR2,SELL,WNT5A,CXCL1,SCARB1,DNM1,TACR2,PRSS1,AHNAK,GNAI1,FPR2,FCGR3A,ICOS,KLRK1,CTLA4,ORM1,CXCL9,GPR44,SPRY2,TNFRSF13B,CXCL13,SCIN,CXCL3,LCP2,CR2,MCHR1,GALR2,RGS7,CD300E	1.36E-06	747	39
leukocyte migration	EGFR,VCAM1,MAPK3,CD5,NOS2,CSF3,HBEGF,HSPB1,APP,ILK,EGR1,CXCL10,CXCL11,CXCR2,SELL,PPARD,CXCL1,STAT1,FPR2,MMP12,ICOS,CTLA4,CD86,CXCL9,CXCL13,CXCL3,RHOH,VNN2	2.94E-06	462	28

Proteins Regulating Cell Processes of Inflammatory Response



Proteins Regulating Cell Processes of Immune Response



Gene Expression Findings:

- Strong Expression Pattern of Inflammatory Response:
 - In tumors as well as in normal samples

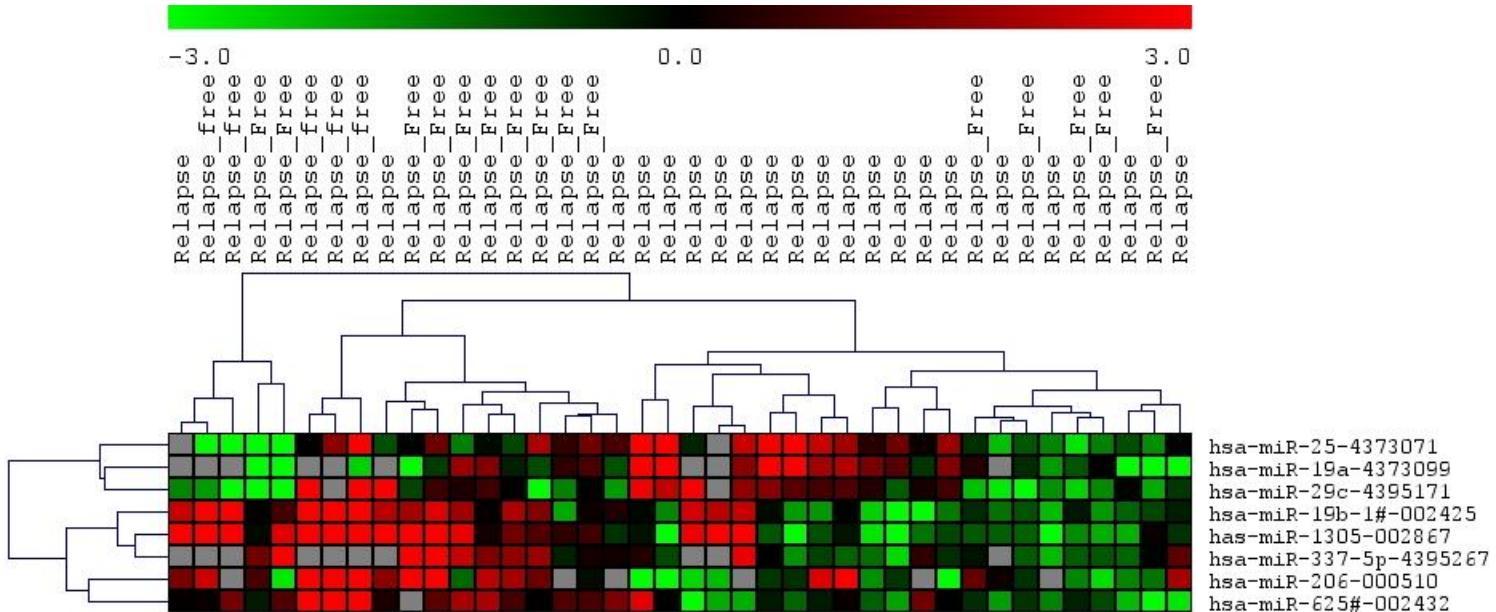
Possible source: Infiltrating white blood cells

Reference:

Schetter et al. Association of inflammation-related and microRNA gene expression with cancer specific mortality of colon adenocarcinoma.
Clin.Cancer.Res. , 2009, 15(18): 5878–5887.

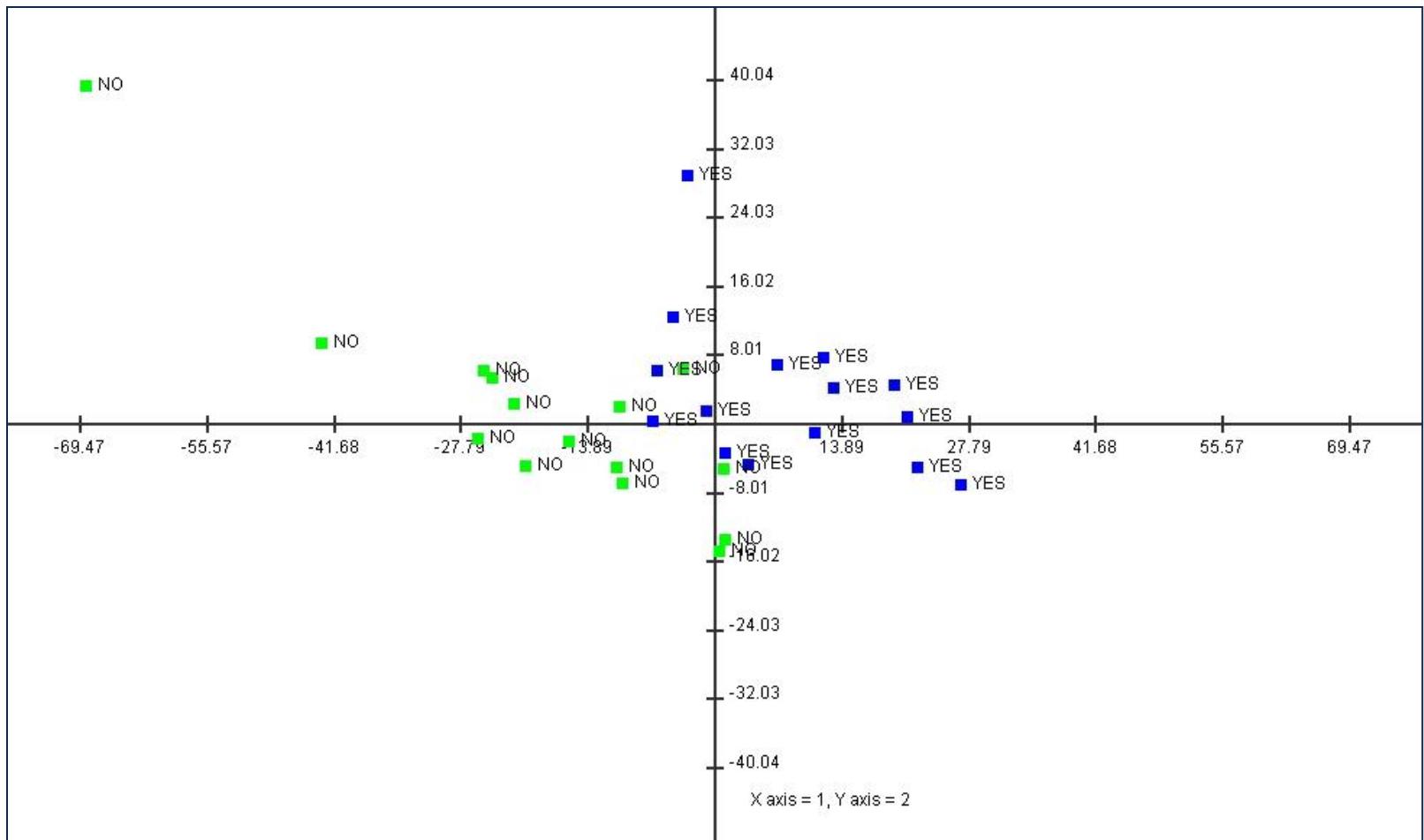
microRNA In Serum

40 samples, t-test p<0.05 8 miRNAs



microRNA In Serum

PCA 8 microRNAs



DNA Copy Number Analysis

Affy SNP 6.0 arrays

- Raw Data → Probe Level Copy Number: **1.6 million probes**
- Probe Level → Segment Level Copy Number: **100K segments**
- Segment LeveL → CIN Index: Whole Chromosome: **22 values**
Individual Cytobands: **~800 values**

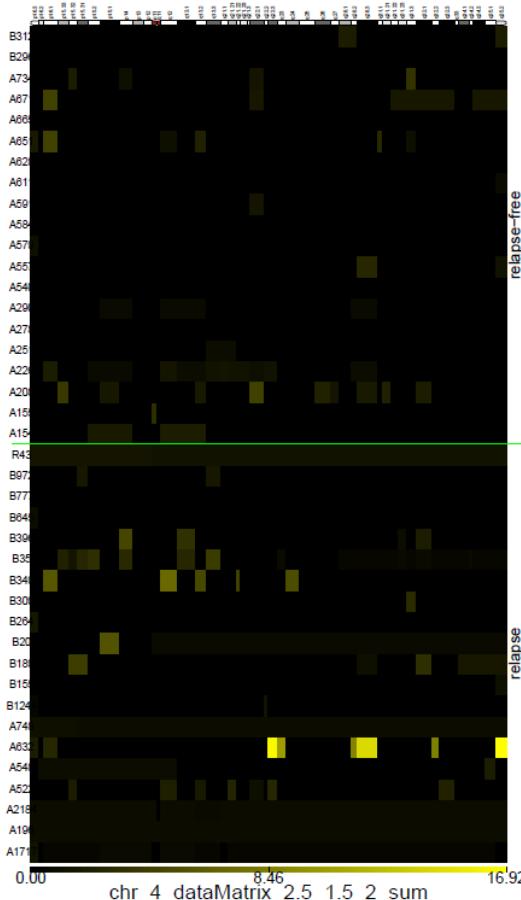
20 Relapse_Free Patients
Tissue DNA: Tumor – 20; Normal – 20

20 Relapse Patients
Tissue DNA: Tumor – 20; Normal – 20

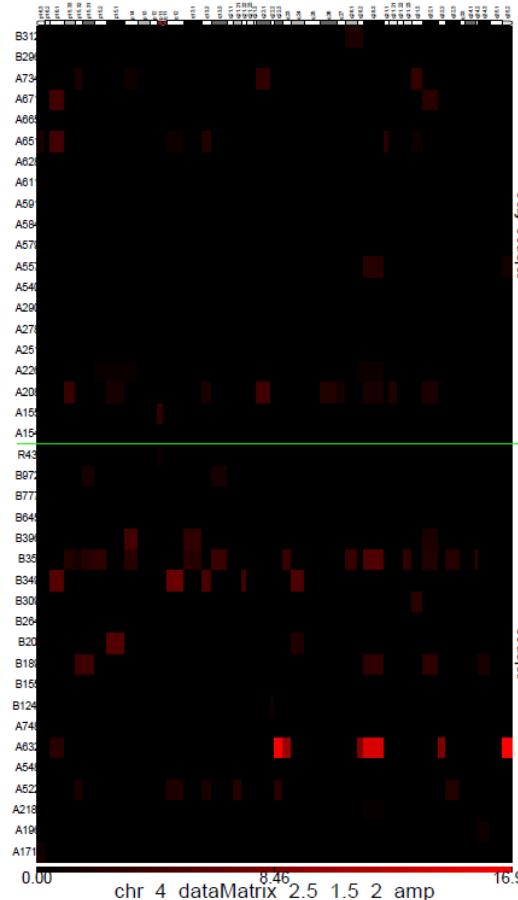
CIN Index, Cytoband level:

Chromosome # 4

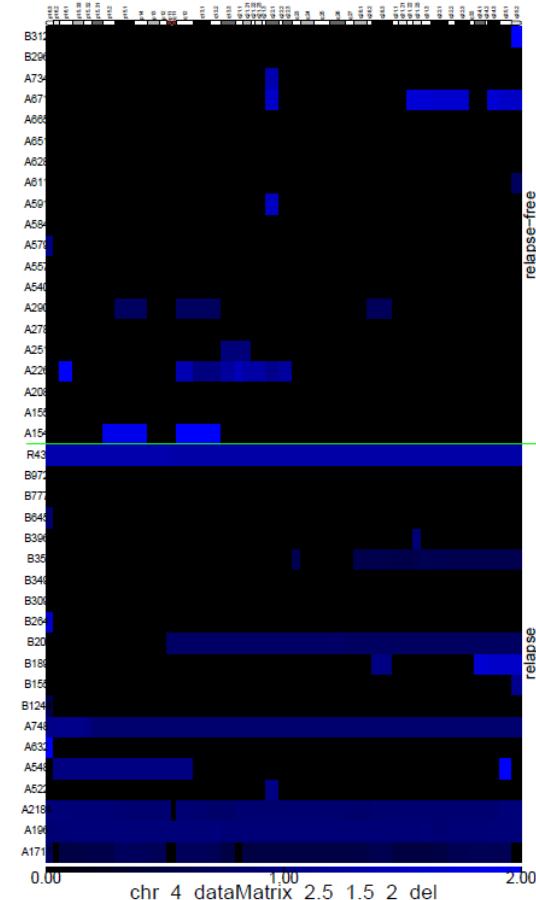
Overall



Gains



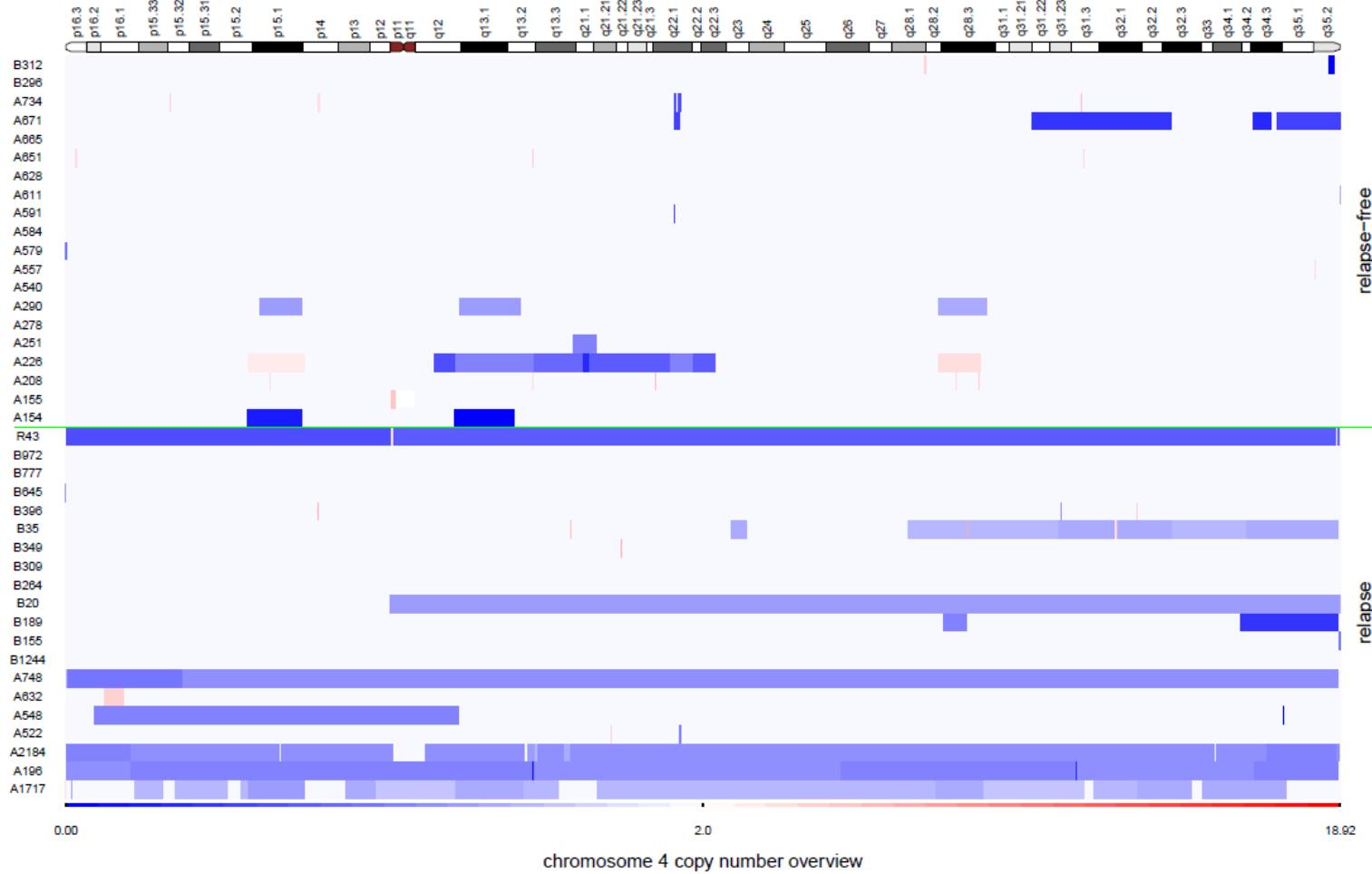
Losses



Copy Number – Segment level

Gains and Losses are shown

Chromosome 4



CIN Index - Cytoband level: Relapse vs Relapse_Free, t-test, P<0.05

Brosens et al. Cell Oncol (Dordr). 2011 Jun;34(3):215-23..
Deletion of chromosome 4q predicts outcome in stage II
colon cancer patients.

RESULTS: Stage II colon cancers of patients who had relapse of disease showed significantly more losses on chromosomes 4, 5, 15q, 17q and 18q.

In the microsatellite stable (MSS) subgroup ($n = 28$), only loss of chromosome 4q22.1-4q35.2 was significantly associated with disease relapse

Metabolites in Biofluids: Relapse vs Relapse_Free

Serum:

Serum Pos – 10 and 30 samples;

Serum Neg – 10 and 30 samples;

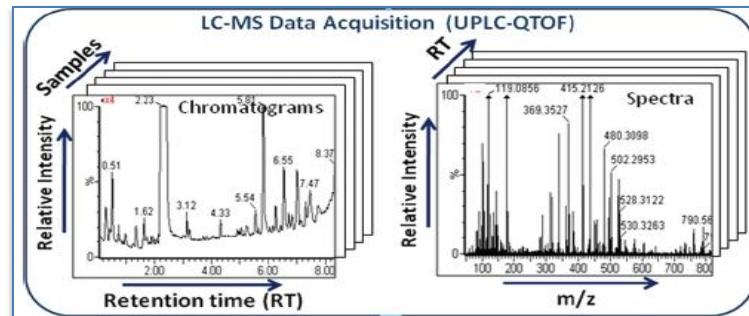
Urine:

Urine Pos - 40 samples

Urine Neg – 40 samples

Metabolomics Methods

Sample Preparation



LC-MS Data Preprocessing (MassLynx)
(Filtering, feature extraction, feature matching,
retention time correction & handling missing peaks)

	Sample 1	Sample N
RT ₁ -m/z ₁				
Ion Abundance				
...				
RT _n -m/z _n				

Linear Modeling
“moderated t-statistics”

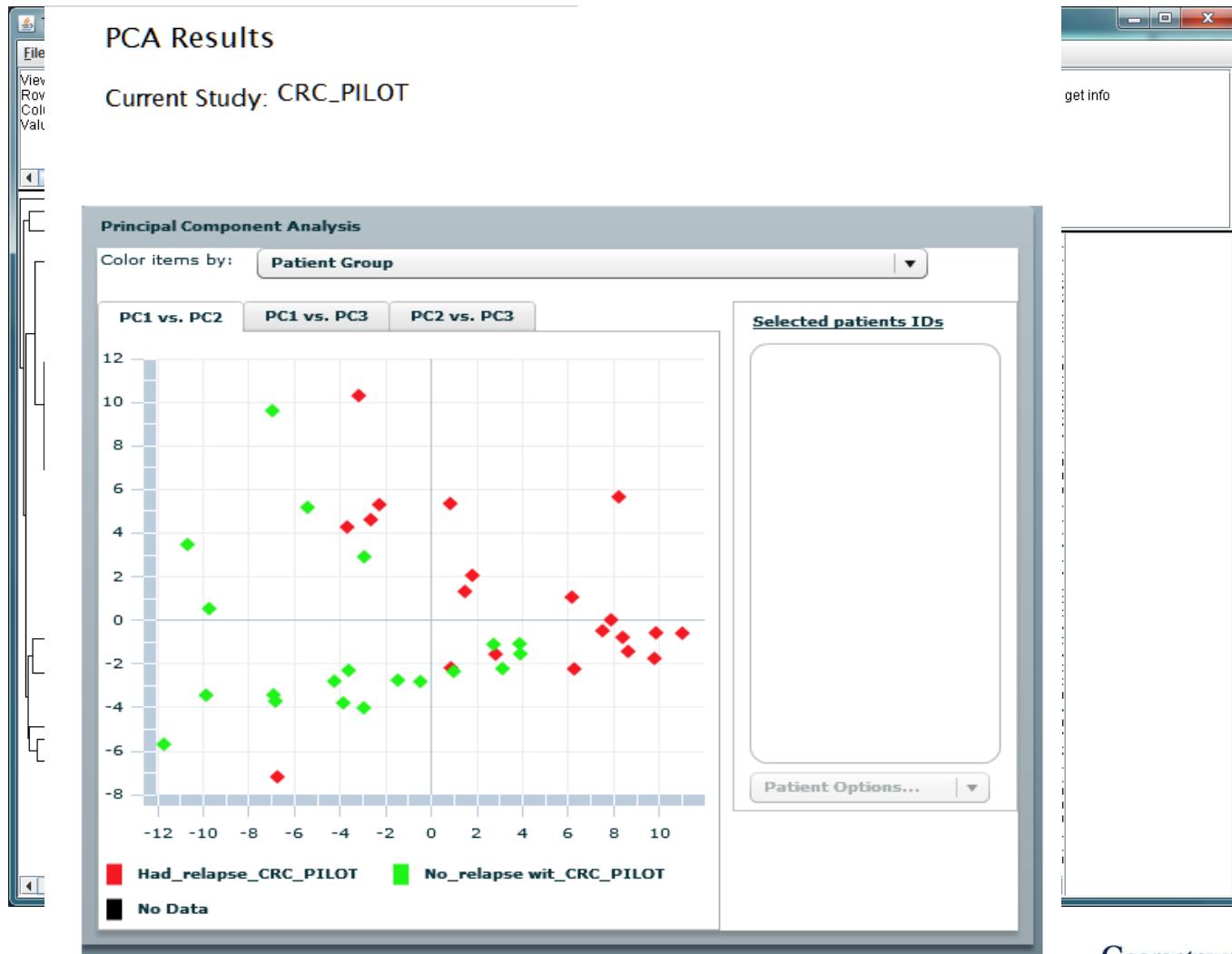
Feature Identification
(HMDB, KEGG, METLIN, METACYC, LMDB)
& Validation (MS/MS)*

Pathway Analysis
IPA , SMPDB

Network Analysis
(Multi-omics)

Metabolomics – Urine Positive

40 samples, 47 peaks p<0.01

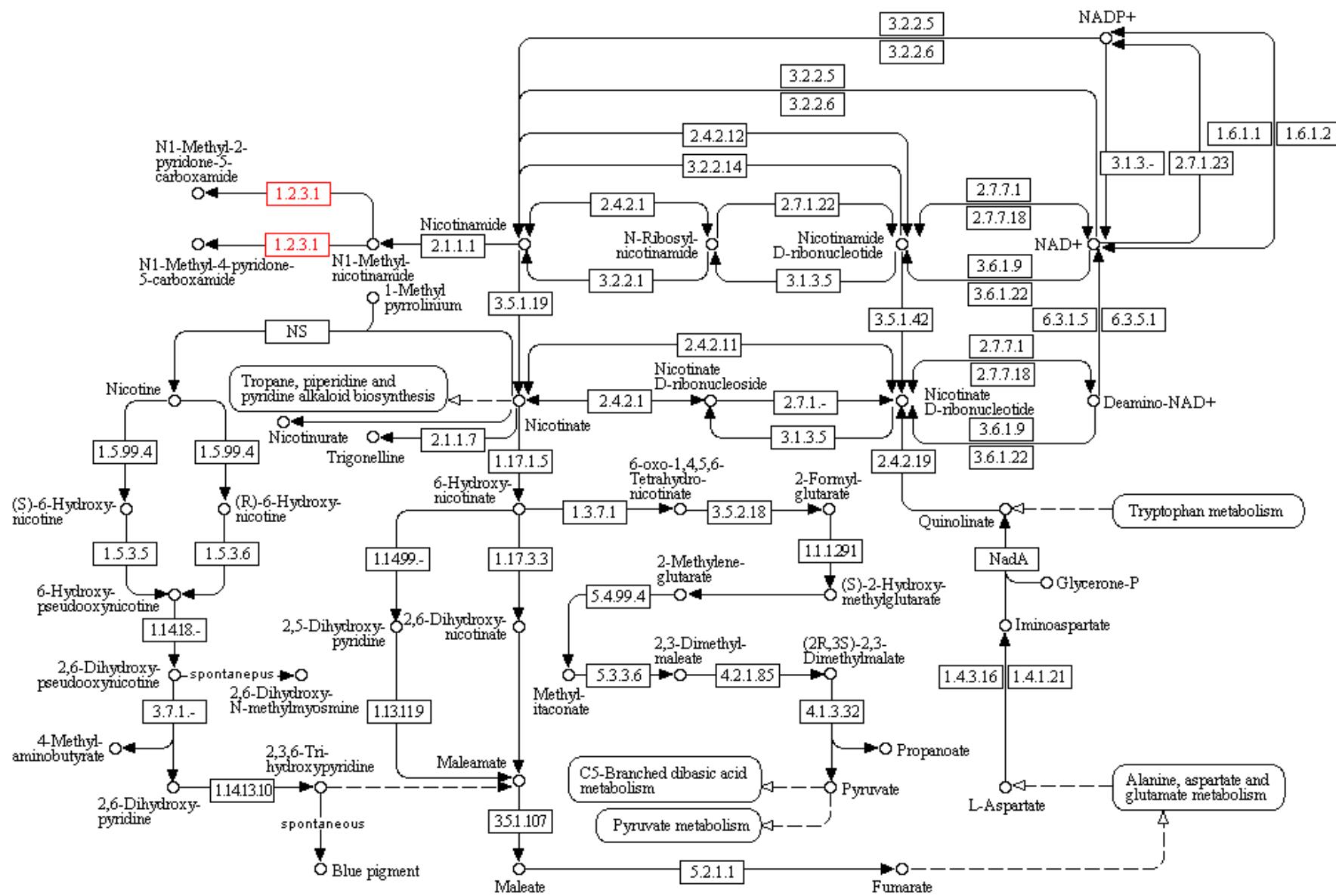


m/z	Putative Metabolites	KEGG	HMDB	METLIN	LMDB	METACYC
119.0815	L-2,4-diaminobutyric acid	o	o			
121.0318	3-Methylthiopropionic acid	o	o			
130.0495	1-Pyrroline-4-hydroxy-2-carboxylate	o	o	o		o
130.0495	5-Oxo-D-proline	o	o	o		o
130.0495	5-oxoproline	o	o			o
130.0495	L-1-Pyrroline-3-hydroxy-5-carboxylate	o	o	o		
130.0495	Pyroglutamic acid	o	o	o		
130.0495	pyrrolidone-carboxylate	o	o			o
130.0495	pyrroline-hydroxy-carboxylate	o	o			o
130.0497	1-Pyrroline-4-hydroxy-2-carboxylate	o	o	o		o
130.0497	5-Oxo-D-proline	o	o	o		o
130.0497	5-oxoproline	o	o			o
130.0497	L-1-Pyrroline-3-hydroxy-5-carboxylate	o	o	o		
130.0497	Pyroglutamic acid	o	o			
130.0497	pyrrolidone-carboxylate	o	o			o
130.0497	pyrroline-hydroxy-carboxylate	o	o			o
130.0499	Pyroglutamic acid	o	o			
130.0499	Pyrrolidonecarboxylic acid	o	o			
130.0499	Pyrroline hydroxycarboxylic acid	o	o			
135.0764	L-Canaline	o	o			
135.0803	cinnamyl alcohol	o	o			o
135.0803	phenylacetone	o	o			o
149.0267	2-Oxo-4-methylthiobutanoic acid	o	o			
153.0655	N1-Methyl-2-pyridone-5-carboxamide	o	o			
153.0655	N1-Methyl-4-pyridone-5-carboxamide	o	o			
153.0655	N-Methyl-2-pyridone-5-carboxamide	o	o	o		
153.0655	N-Methyl-4-pyridone-5-carboxamide	o	o	o		
153.0659	N1-Methyl-2-pyridone-5-carboxamide	o	o			
153.0659	N1-Methyl-4-pyridone-5-carboxamide	o	o			
165.0536	2-coumarate	o	o			o
165.0536	2-Hydroxycinnamate	o	o			
165.0536	2-Hydroxycinnamic acid	o	o			
165.0536	4-coumarate	o	o			o
165.0536	4-Hydroxycinnamic acid	o	o	o		
165.0536	cis-p-coumarate	o	o			o
165.0536	enol-phenylpyruvate	o	o			o
165.0536	m-Coumaric acid	o	o	o		

Enrichment Analysis for Differentially Identified Putative Metabolites

Pathway Name	SMPDB_SourceID	Metabolite hit	Description	P Value (Based on Hypergeometric test)
Phenylketonuria/ Phenylalanine Metabolism	SMP00206	Phenylpyruvic acid	Phenylpyruvic acid is a keto-acid that is an intermediate or catabolic byproduct of phenylalanine metabolism. Phenylalanine accumulation disrupts brain development, leading to mental retardation.	0.00224
		Phenylacetic acid	Phenyl acetate (or phenylacetate) is a carboxylic acid ester that has been found in the biofluids of patients with nephritis and/or hepatitis as well as patients with phenylketonuria (PKU).	
Nicotinate and Nicotinamide Metabolism	SMP00048	N1-Methyl-2-pyridone-5-carboxamide	N-methyl-2-pyridone-5-carboxamide (2PY) is one of the end products of nicotinamide-adenine dinucleotide (NAD) degradation. Increased serum 2PY concentrations are observed in chronic renal failure (CRF) patients, which along with the deterioration of kidney function and its toxic properties (significant inhibition of PARP-1), suggests that 2PY is an uremic toxin. (PMID 12694300)	0.01563
		N1-Methyl-4-pyridone-5-carboxamide	N1-Methyl-4-pyridone-5-carboxamide (4PY) is a normal human metabolite (one of the end products of nicotinamide-adenine dinucleotide (NAD) degradation). 4PY concentration in serum is elevated in non-dialyzed chronic renal failure (CRF) patients when compared with controls. (PMID 12694300)	
5-Oxoprolinuria / Glutathione Synthetase Deficiency	SMP00143 / SMP00337	Pyroglutamic acid	A cyclized derivative of L-glutamic acid,Elevated blood levels may be associated with problems of glutamine or glutathione metabolism	0.01581 / 0.04603
Prolidase Deficiency(PD) / Arginine and Proline Metabolism	SMP00207 / SMP00020	1-Pyrroline-4-hydroxy-2-carboxylate	pyrrole-2-carboxylate (PCA) in human urine may be formed in urine from a labile precursor, presumably delta(1)-pyrroline-4-hydroxy-2-carboxylate (PMID: 4430715)	0.05076 / 0.05422
		Pyrroline hydroxycarboxylic acid	Pyrroline hydroxycarboxylic acid is a metabolite identified in the urine of patients with type II hyperprolinemia. (OMIM 239510). The oxidation of pyrroline-carboxylate generates glutamate and pyrroline-hydroxycarboxylate, a reaction catalyzed by hydroxyproline oxidase (PMID: 500817)	

NICOTINATE AND NICOTINAMIDE METABOLISM



Insights into Colon Cancer Etiology via a Regularized Approach to Gene Set Analysis of GWAS Data.

Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, Hsu L.

Biostatistics and Biomathematics Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, USA.

Abstract

Genome-wide association studies (GWAS) have successfully identified susceptibility loci from marginal association analysis of SNPs. Valuable insight into genetic variation underlying complex diseases will likely be gained by considering functionally related sets of genes simultaneously. One approach is to further develop gene set enrichment analysis methods, which are initiated in gene expression studies, to account for the distinctive features of GWAS data. These features include the large number of SNPs per gene, the modest and sparse SNP associations, and the additional information provided by linkage disequilibrium (LD) patterns within genes. We propose a "gene set ridge regression in association studies (GRASS)" algorithm. GRASS summarizes the genetic structure for each gene as eigenSNPs and uses a novel form of regularized regression technique, termed group ridge regression, to select representative eigenSNPs for each gene and assess their joint association with disease risk. Compared with existing methods, the proposed algorithm greatly reduces the high dimensionality of GWAS data while still accounting for multiple hits and/or LD in the same gene. We show by simulation that this algorithm performs well in situations in which there are a large number of predictors compared to sample size. We applied the GRASS algorithm to a genome-wide association study of colon cancer and identified nicotinate and nicotinamide metabolism and transforming growth factor beta signaling as the top two significantly enriched pathways. Elucidating the role of variation in these pathways may enhance our understanding of colon cancer etiology.

Copyright © 2010 The American Society of Human Genetics. Published by Elsevier Inc. All rights reserved.

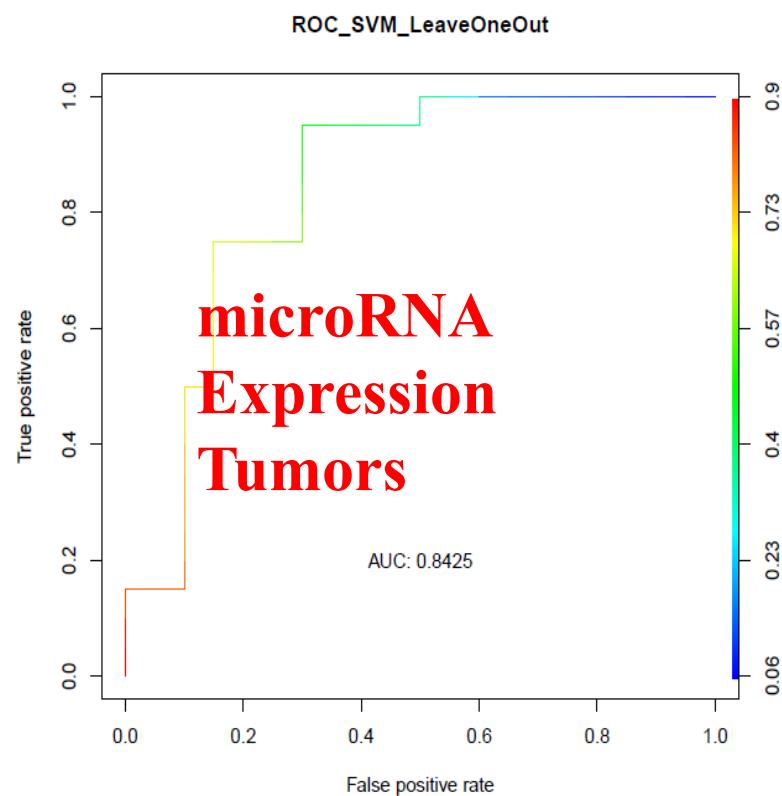
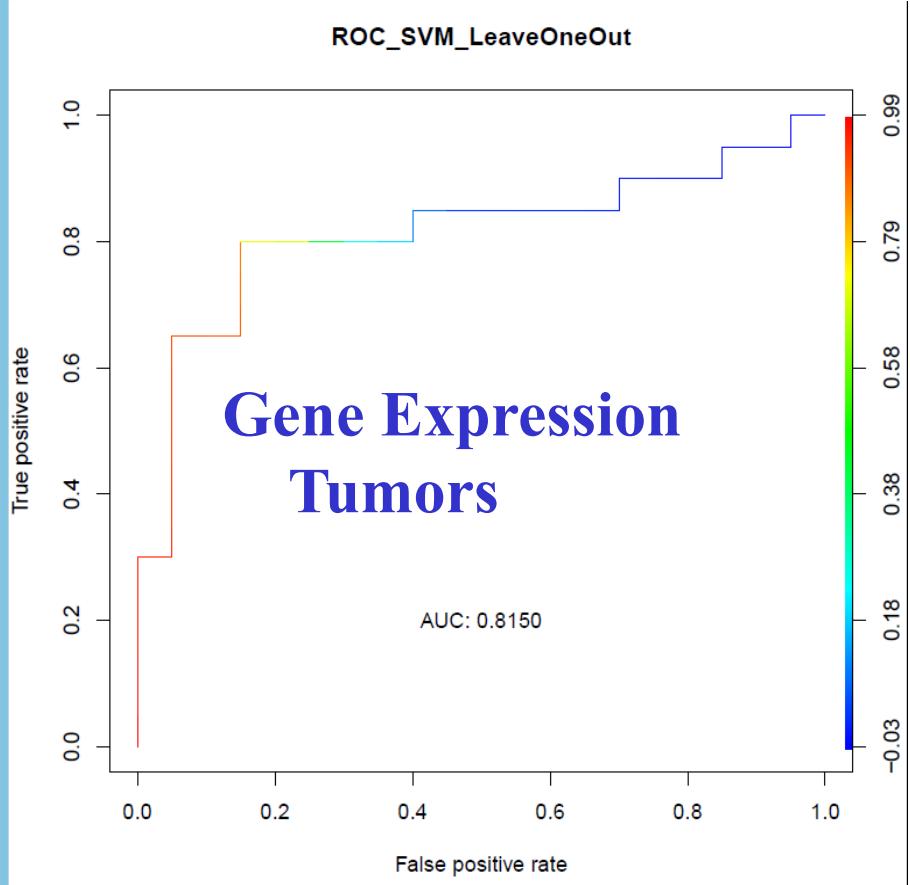
PMID: 20605128 [PubMed - as supplied by publisher]

Table 3. Top-Ranking KEGG Pathways Associated with Colon Cancer Risk in the Women's Health Initiative Sample

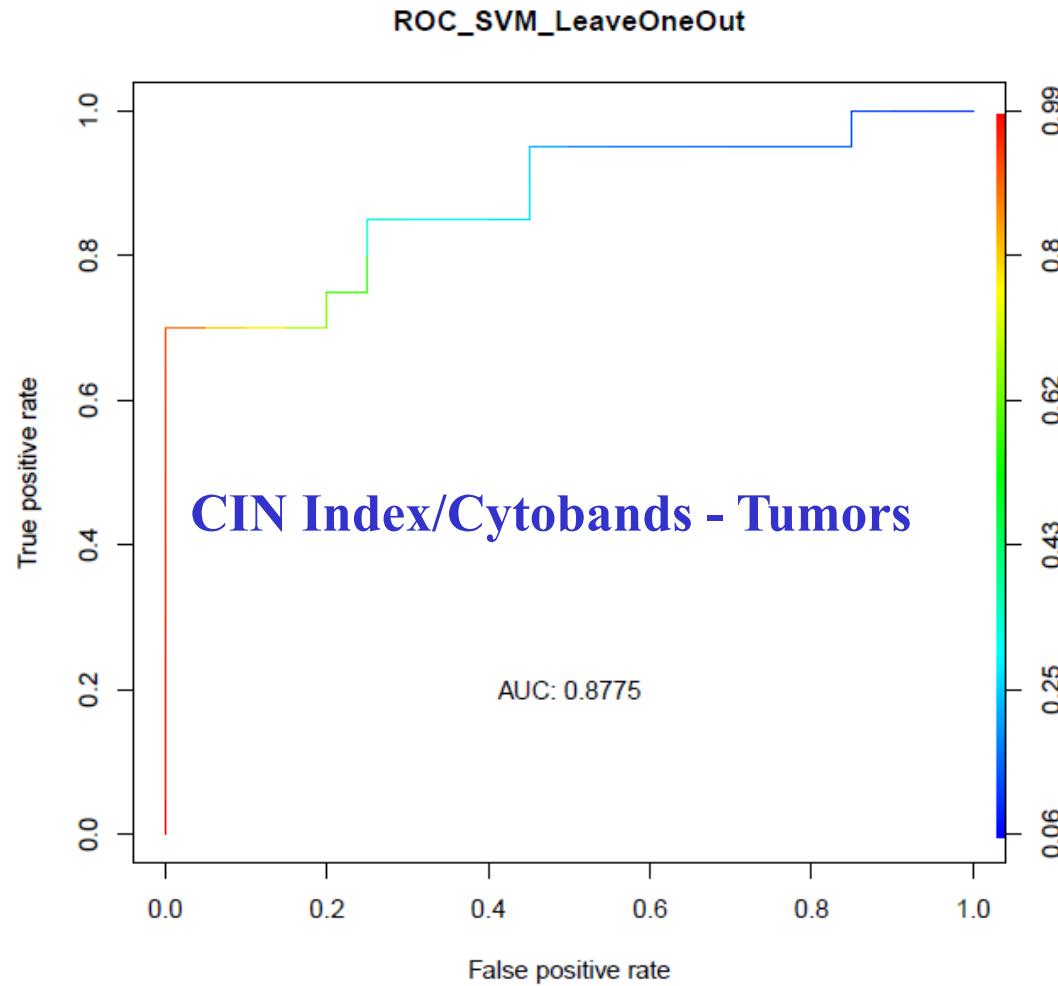
Rank	KEGG Number	Pathway Name	No. Genes/No. eigenSNPs	p Value
1	HSA00760	Nicotinate and nicotinamide metabolism	23/602	0.015
2	HSA04350	TGF-beta signaling	89/2912	0.035

Top-ranking KEGG pathways that are associated with colon cancer risk at significance level $\alpha = 0.05$. p values are calculated from Equation 6 based on 1000 permutations.

Classification Algorithm: Support Vector Machine (SVM)

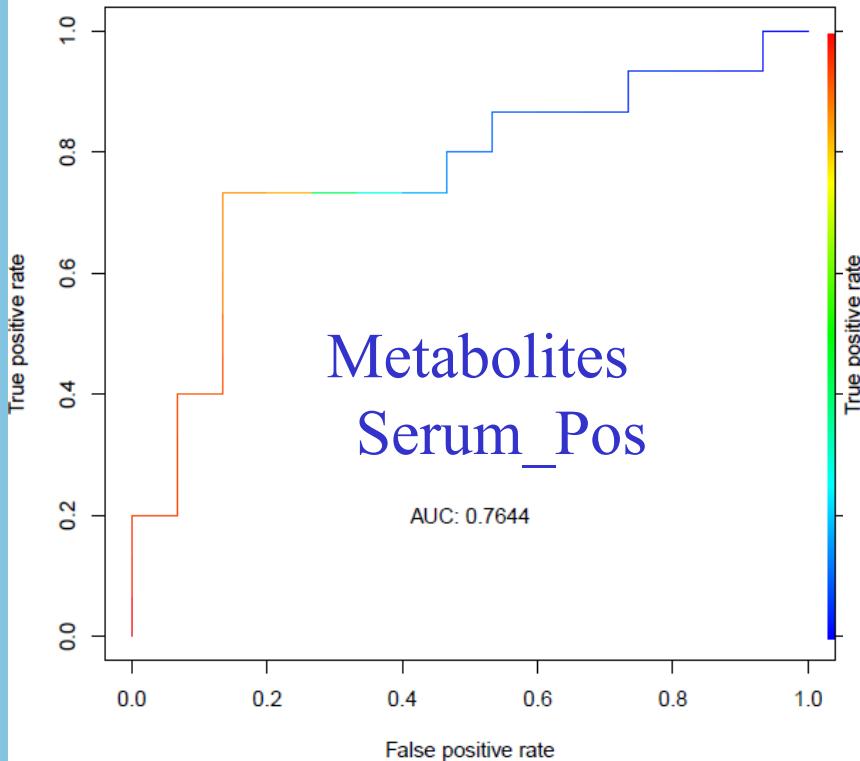


Classification Algorithm: Support Vector Machine (SVM)



Classification Algorithm: Support Vector Machine (SVM)

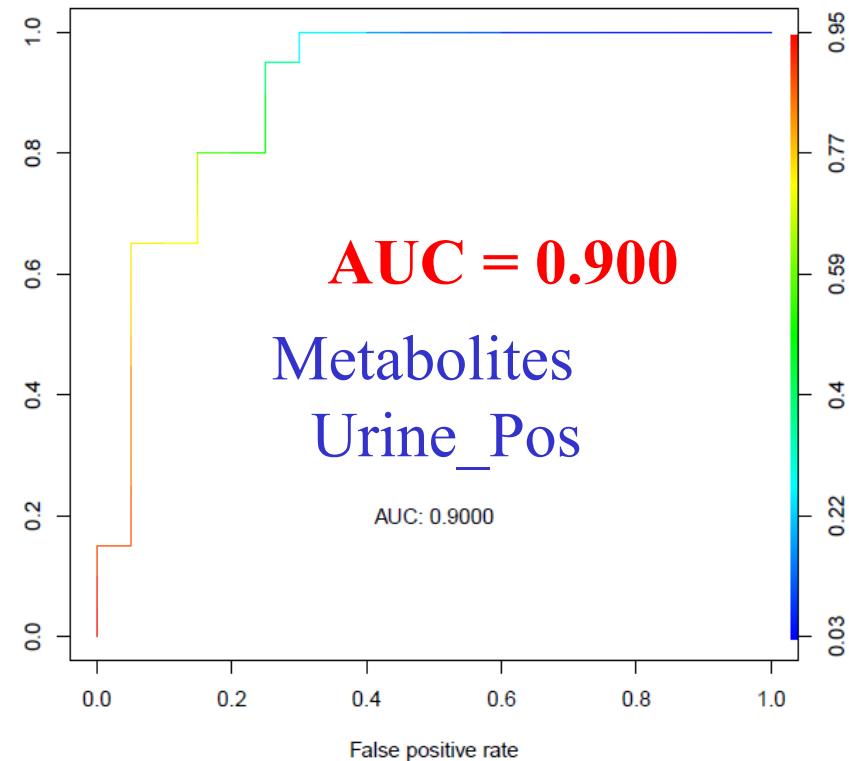
ROC_SVM_LeaveOneOut



Metabolites
Serum_Pos

AUC: 0.7644

ROC_SVM_LeaveOneOut



Metabolites
Urine_Pos

AUC: 0.9000

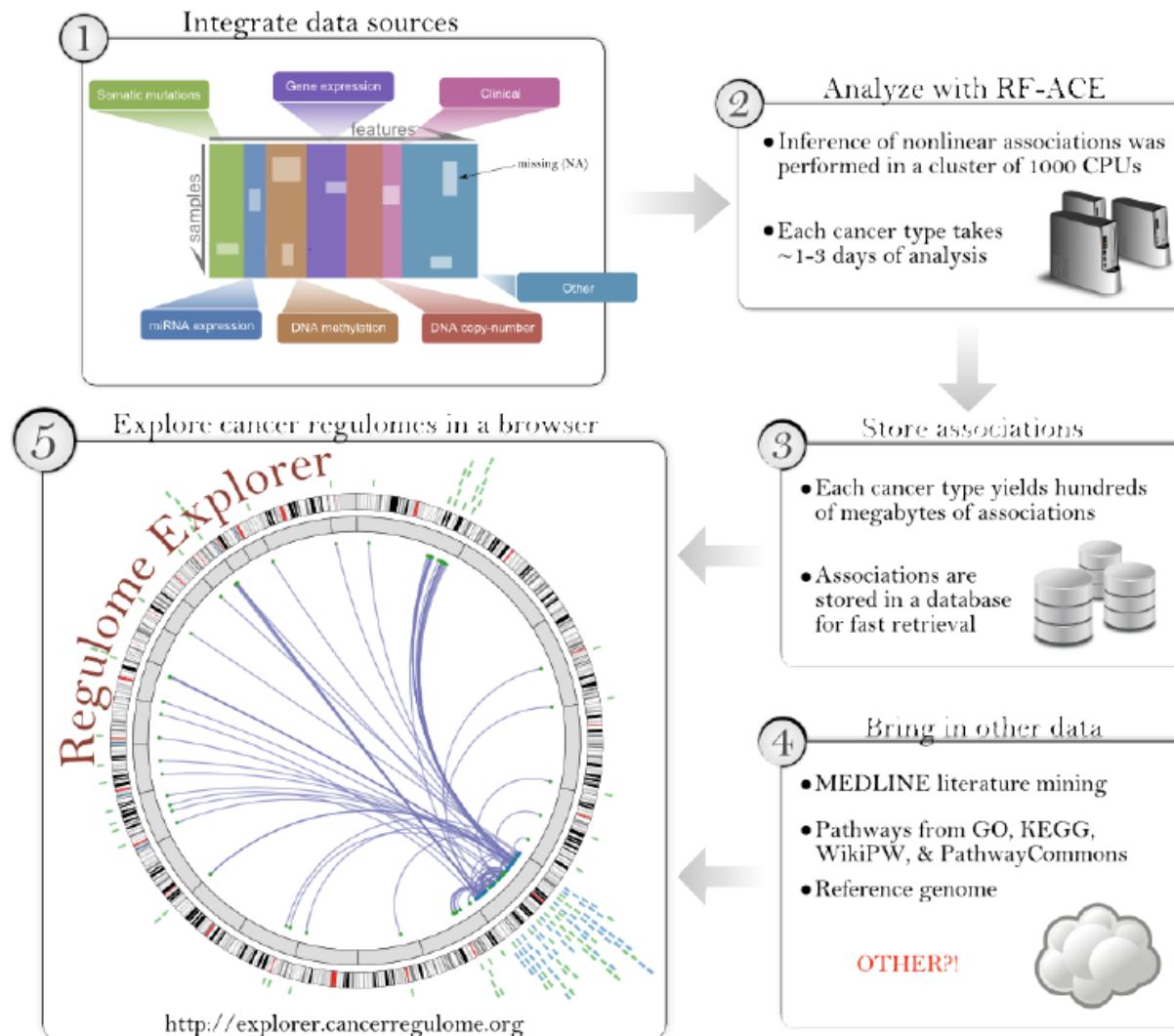
Results of ROC Analysis for SVM classification

Data Type/Tissue Type	Classifier	AUC
Metabolites/Serum_Pos	SVM	0.7
Gene Expression/Tumors	SVM	0.81
miRNA Expression/Tumors	SVM	0.84
CIN Index/Tumors	SVM	0.87
Metabolites/Urine_Pos	SVM	0.9

Critical associations involved in CRC Relapse

- Results with 5 molecular data sets plus clinical outcome:
Gene Expression, microRNA expression, CNV, Metabolites/Serum;
Metabolites/Urine
- Target: Relapse Status
- Workflow:
 - Combine 5 data matrixes (pre-filtered on significance of differences) –
 - Input to RF-ACE algorithm,
 - Analyze with RF-ACE with 2K permutations,
 - Results of RF-ACE - Upload to Regulome Explorer instance at AWS
 - Visualize results of analysis on Genome map and/or Network viewer
 - Identify genome locations with large number of highly correlated changes
 - i.e. “Hot Spots”
 - filter data based on importance of association with clinical outcome
 - Find interactions between different molecular features that are highly ranked on importance

Multi-Variate Analysis with RF-ACE and Mapping to Regulome Explorer



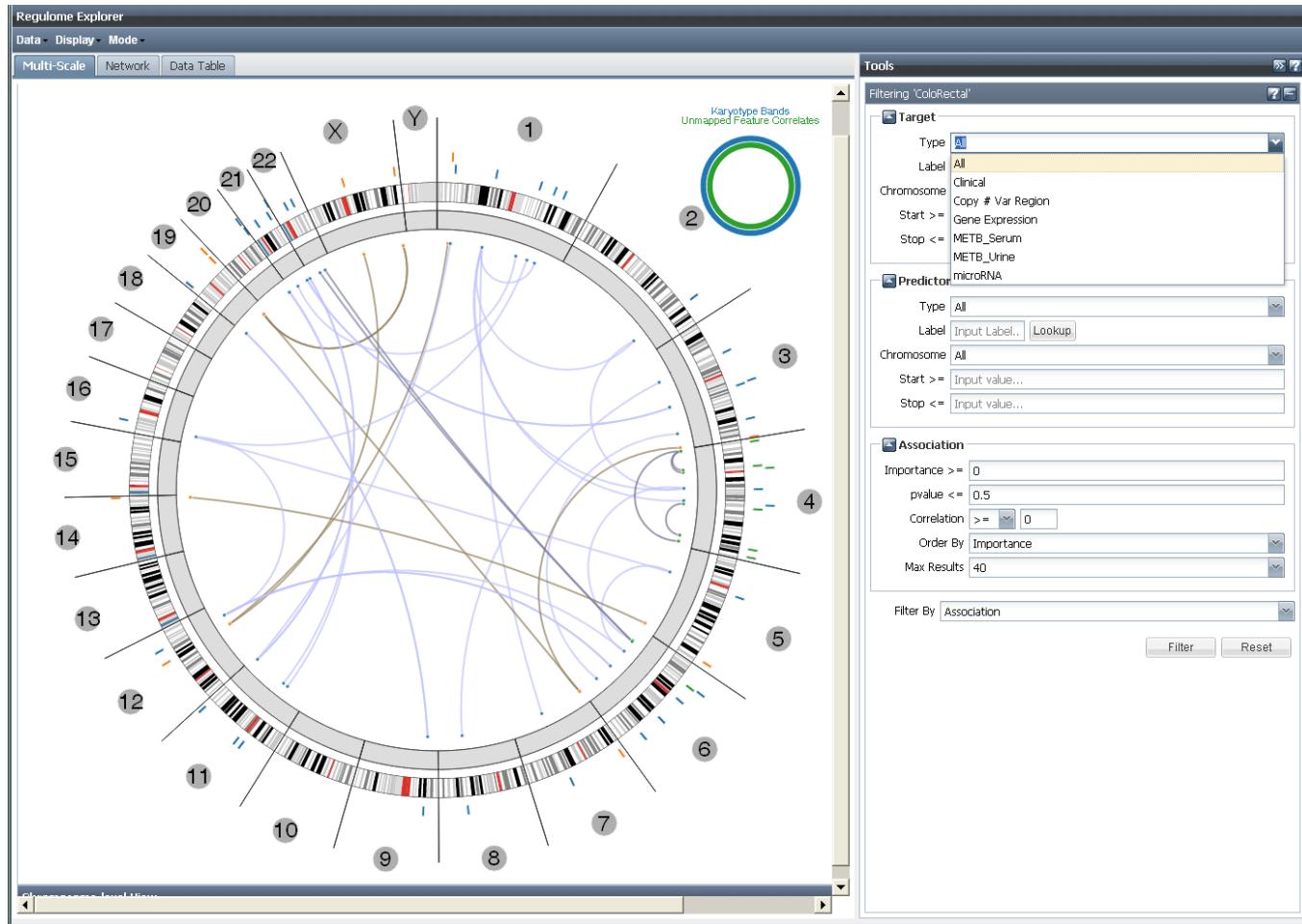
Multi-Variate Analysis with RF-ACE ; Visualization with Regulome Explorer

RF-ACE

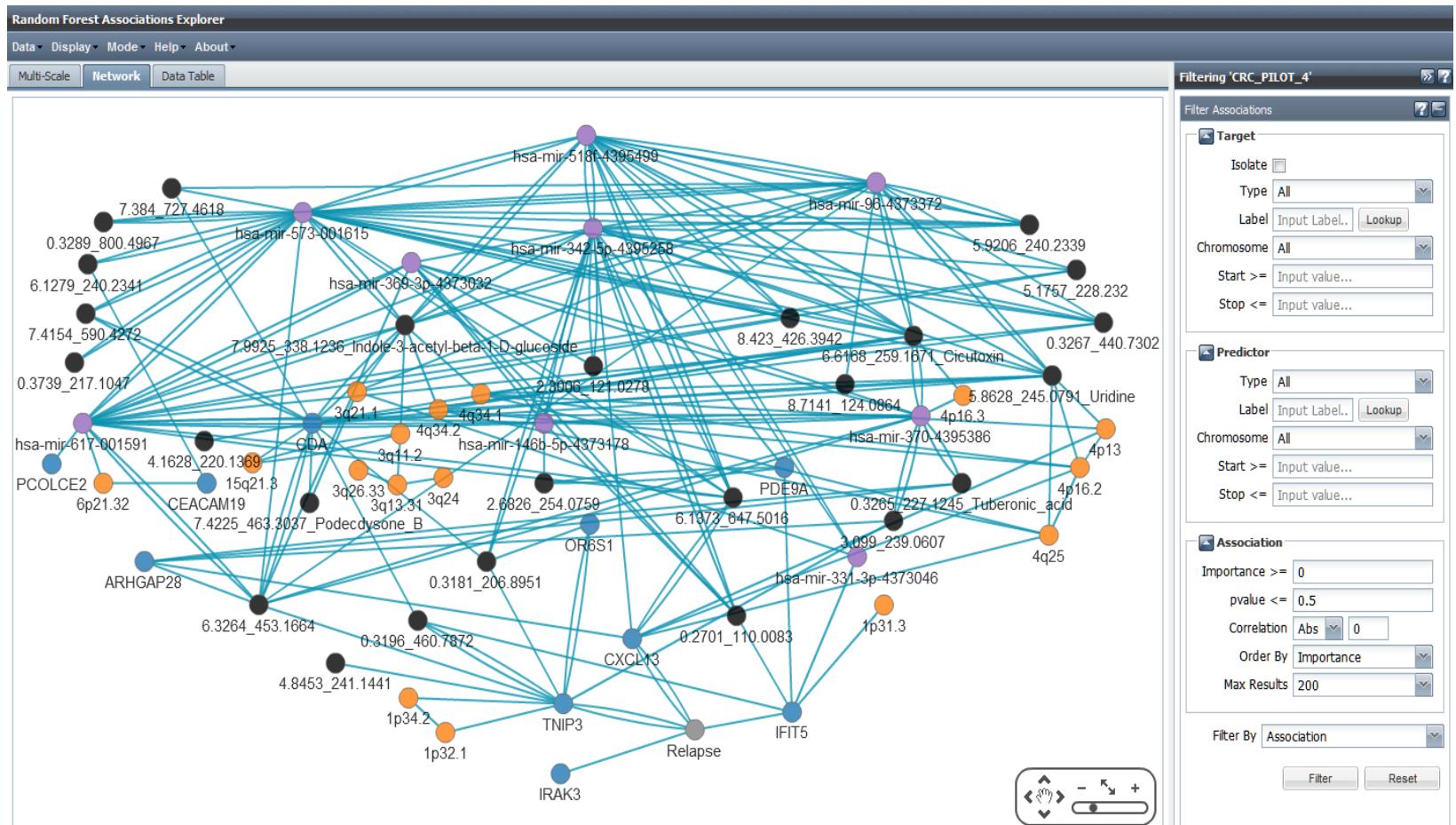
(Random Forests with Artificial Contrast Ensembles)

- RF implementation with added flexibility
 - support for string literals and various data formats
 - Easy interface with default parameter options
- Normalized importance score
- Inclusion of statistical testing framework
 - p-values for associations
- Better predictive power with Gradient Boosting Trees

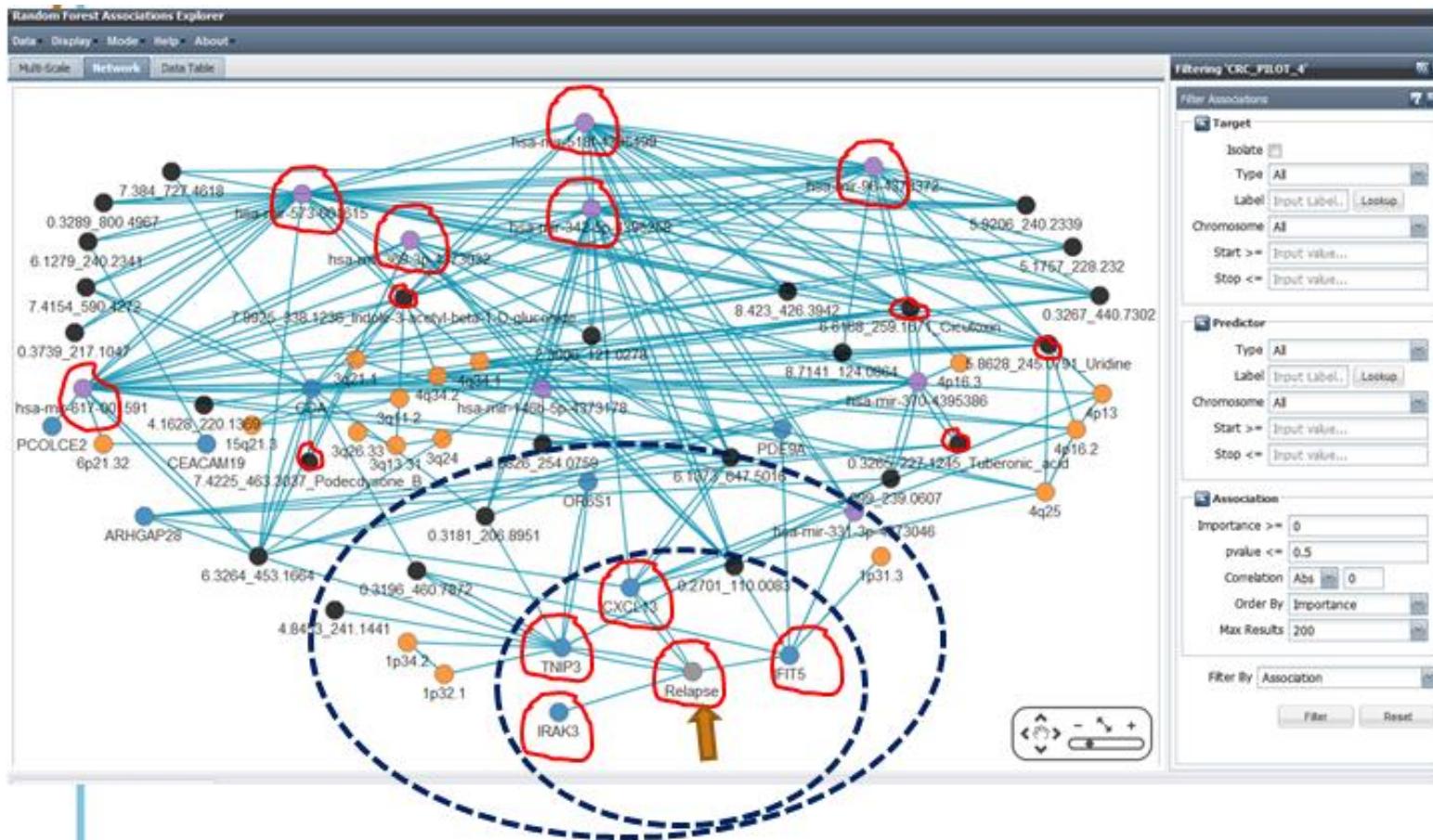
Genome View of Top 40 Molecular Features/Associations



Top 200 Features Associated With CRC Relapse



Top 200 Features Associated With CRC Relapse



Bayesian to RF-ACE comparison

Augmentd MArKov B (0.25)	RF-ACE
N:METB_Serum:0.3289_800.4967::::	N:MI RN:hsa-mir-342-5p-4395258:chr14:100575991:100576090:+:
N:METB_Urine:4.8453_241.1441::::	N:METB_Serum:6.1279_240.2341::::
N:GEXP:IFIT5:chr10:91174555:91178405:+:	N:METB_Urine:7.4154_590.4272::::
N:GEXP:TNIP3:chr4:122053785:122085280:-:	N:METB_Urine:5.9206_240.2339::::
N:METB_Serum:0.3267_440.7302::::	N:METB_Urine:7.9925_338.1236_Indole-3-acetyl-beta-1-D-glucoside::::
N:GEXP:PDE9A:chr21:44073924:44195403:+:	N:MI RN:hsa-mir-617-001591:chr12:81226311:81226408:-:
N:MI RN:hsa-mir-370-4395386:chr14:101377475:101377550:+:	N:METB_Urine:4.1628_220.1369_Darlingine::::
N:METB_Urine:6.6168_259.1671_Cicutoxin::::	N:METB_Urine:7.384_727.4618::::
N:GEXP:IRAK3:chr12:66583079:66641951:+:	N:METB_Urine:2.3006_121.0278::::
N:METB_Urine:7.4225_463.3037_Podecdysone_B::::	N:METB_Urine:6.3264_453.1664::::
N:METB_Serum:5.1757_228.232_Halaminol_A::::	N:METB_Urine:0.3265_227.1245_Tuberonic_acid::::
N:GEXP:OR651:chr14:21108854:21109850:-:	N:METB_Urine:4.8453_241.1441::::
N:GEXP:CXCL13:chr4:78527019:78532193:+:	N:METB_Urine:7.4225_463.3037_Podecdysone_B::::
N:METB_Serum:6.1373_647.5016_SM(d18_1/12_0)::::	N:GEXP:ARHGAP28:chr18:6837347:6912155:+:
N:METB_Serum:0.2701_110.0083::::	N:METB_Urine:5.8628_245.0791_Uridine::::
N:MI RN:hsa-mir-342-5p-4395258:chr14:100575991:100576090:+:	N:METB_Serum:0.3196_460.7872::::
N:METB_Urine:7.384_727.4618::::	N:METB_Serum:5.1757_228.232_Halaminol_A::::
N:METB_Urine:6.3264_453.1664::::	N:GEXP:PDE9A:chr21:44073924:44195403:+:
N:CNVR:4q34.1:chr4:172200000:176600000::	N:GEXP:CDA:chr1:20915622:20945061:-:
N:GEXP:CDA:chr1:20915622:20945061:+:	N:GEXP:PCOLCE2:chr3:142537176:142607738:-:
N:METB_Urine:5.9206_240.2339::::	N:MI RN:hsa-mir-96-4373372:chr7:129414531:129414609:-:
N:GEXP:ARHGAP28:chr18:6837347:6912155:+:	N:METB_Urine:2.6826_254.0759::::
N:METB_Urine:2.3006_121.0278::::	N:METB_Urine:6.6168_259.1671_Cicutoxin::::
N:METB_Urine:0.3265_227.1245_Tuberonic_acid::::	N:GEXP:IFIT5:chr10:91174555:91178405:+:
N:MI RN:hsa-mir-573-001615:chr4:24521814:24521913:-:	N:GEXP:TNIP3:chr4:122053785:122085280:-:
N:CNVR:1p31.3:chr1:60900000:68700000::	N:MI RN:hsa-mir-331-3p-4373046:chr12:95702195:95702289:+:
N:GEXP:PCOLCE2:chr3:142537176:142607738:-:	N:MI RN:hsa-mir-370-4395386:chr14:101377475:101377550:+:
N:CNVR:6p21.32:chr6:31900000:33600000::	N:GEXP:CXCL13:chr4:78527019:78532193:+:
N:METB_Urine:4.1628_220.1369_Darlingine::::	N:GEXP:OR651:chr14:21108854:21109850:-:
	N:METB_Serum:6.1373_647.5016_SM(d18_1/12_0)::::
	N:METB_Serum:0.2701_110.0083::::
	N:GEXP:CEACAM19:chr19:45175203:45186762:+:
	N:METB_Serum:0.3181_206.8951::::
	N:METB_Serum:0.3289_800.4967::::
	N:METB_Serum:0.3267_440.7302::::
	N:METB_Serum:0.3739_217.1047::::
	N:MI RN:hsa-mir-573-001615:chr4:24521814:24521913:-:

Last stop: Somatic Mutation Analysis

Variant Analysis™

Feedback

My Samples | My Analyses > CRC test4 two groups R vs RF [x]

Filter Cascade

Variants	Genes
71949	8777

Predicted Deleterious

23252	6918

Genetic Analysis

6676	4012

Biological Context

38	28

Add Filter

Legend [hide]

Function	Call Quality
-	<20
Identical to Reference Genome	20 -> 30+
Heterozygous Variant	30+ (black)
Heterozygous/Ambiguous	30+ (orange)
Homozygous Variant	30+ (dark blue)
Copy Number Gain/Heterozygous	30+ (light orange)
Copy Number Gain/Homozygous	30+ (light blue)
Hemizygous	30+ (dark orange)
Nullizygous	30+ (yellow)
Gene Fusion	30+ (red)
No genotype	30+ (grey)

Variants | Genes | Groups/Complexes | Pathways | Processes | Diseases | Overview

Add Filter Export Find processes Top 100 processes

Name	p-value	#Genes	#Variants	#Cases ▾	%Cases	#Controls	%Controls
migration of cells	1.367E-8	15	24	17	85	0	0
inflammatory response	4.587E-11	13	23	16	80	0	0
proliferation of blood cells	1.229E-10	13	22	16	80	0	0
migration of blood cells	6.755E-10	13	22	16	80	0	0
leukocyte migration	6.755E-10	13	22	16	80	0	0
cell movement of leukocytes	1.270E-10	13	22	16	80	0	0
cell movement of blood cells	6.947E-10	13	22	16	80	0	0
proliferation of mononuclear leukocytes	5.399E-10	12	21	16	80	0	0
proliferation of lymphocytes	4.060E-10	12	21	16	80	0	0
proliferation of immune cells	1.073E-9	12	21	16	80	0	0
cell death of blood cells	1.436E-9	12	21	16	80	0	0
proliferation of T lymphocytes	6.771E-10	11	20	16	80	0	0
cell death of immune cells	1.368E-8	11	20	16	80	0	0
activation of mononuclear leukocytes	4.200E-11	11	20	16	80	0	0
activation of leukocytes	2.552E-9	11	20	16	80	0	0
activation of blood cells	1.193E-8	11	20	16	80	0	0
growth of immune cells	3.564E-15	11	17	16	80	0	0
growth of blood cells	2.992E-14	11	17	16	80	0	0
cellular homeostasis	1.678E-8	14	24	15	75	0	0
hematopoiesis	2.823E-9	12	21	15	75	0	0
development of blood cells	3.105E-11	12	21	15	75	0	0
infiltration of leukocytes	1.284E-11	11	20	15	75	0	0

migration of cells genes

- ALB albumin
- ASTN1 astrotactin 1
- CCR2 chemokine (C-C motif) receptor 2
- CD14 CD14 molecule
- FN1 fibronectin 1
- GATA2 GATA binding protein 2
- IL10 interleukin 10
- IL17A interleukin 17A
- IL1B interleukin 1, beta
- IL1RN interleukin 1 receptor antagonist
- IL21 interleukin 21
- NFKB1 nuclear factor of kappa light polypeptide gene enhancer in B-cells 1

Filter

CRC Exome Data – Variant Analysis

Welcome Yuriy Gusev | Logout FAQ

Variant Analysis™

Feedback

My Samples | My Analyses rg_test1 [x]

Filter Cascade

Variants	Genes
71949	8777
↓	
Common Variants	
47882	8338
↓	
Predicted Deleterious	
17709	6389
↑↓	
Cancer Driver Variants	
3569	1326
↑↓	
Genetic Analysis	
1667	927
↑↓	
Biological Context	
1638	913
↑	
Add Filter	

Legend [hide]

Function Call Quality <20...>30+
 Loss normal Gain
 Identical to Reference Genome
 Heterozygous Variant
 Heterozygous/Ambiguous
 Homozygous Variant
 Copy Number Gain/Heterozygous
 Copy Number Gain/Homozygous
 Hemizygous
 Nullzygous
 Gene Fusion
 No genotype

Variant: chr1 - 43804340

View Path to Phenotype
 Variant Findings (2)

Gene Symbol: MPL
 myeloproliferative
 leukemia virus
 oncogene

dbSNP ID: 12731981

Clinical Assessment: Benign

Gene Region: Exonic

Protein Variant: V114M

Transcript Variant: 340G>A

Translation Impact: nonsynonymous

SIFT Function: Tolerated

Variation Type: SNP

Cytoband: 1p34.2

CG Public Genomes: 3.7%

Frequency: 1000 Genomes: 2.24%

More Details

Search for gene name/symbol 1638 variants

Chr...	Position	Gene Region	Gene Symbol	Protein Variant	Case Samples	Control Samples	Translation Impact	SIFT Function	Regulatory Site	Regulator	Variant Findings	dbSNP ID
1	43212750	Exonic, Intron/exon	LEPRE1	G750R	- - - - -	- - - - -	nonsynonymous					1165
1	43804340	Exonic	MPL	V114M	- - - - -	- - - - -	nonsynonymous	Tolerated			2	1273
1	43826794	Exonic	CDC20	V361I	- - - - -	- - - - -	nonsynonymous	Tolerated				1801
1	44071114	Exonic	PTPRF	R1121T, R1130C	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	45213367	Exonic	KIF2C	K104R	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	45267579	Exonic	PLK3	L210M	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	45798078	Exonic	MUTYH	V230G, V231G, I232V	- - - - -	- - - - -	nonsynonymous	Damaging				
1	46073362	Exonic, Intron/exon	NASP	Q196R, Q280R	- - - - -	- - - - -	nonsynonymous	Tolerated				7518
1	46080809	Exonic	NASP	N259K, N533K	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	46655680	Exonic	POMGNT1	E551K	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	46657846	Exonic	POMGNT1	R488Q	- - - - -	- - - - -	nonsynonymous	Damaging				
1	47042324	Exonic, ncRNA	MKNK1	I79M	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	47048945	Exonic, ncRNA	MKNK1	K31E	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	47737780	Exonic	STIL	M784R	- - - - -	- - - - -	nonsynonymous	Damaging				
1	47746485	Exonic	STIL	E549K	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	47746675	Exonic	STIL	L485F	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	52301812	Exonic	NRD1 (includes L171M, L235M, S39C)	- - - - -	- - - - -	- - - - -	nonsynonymous	Damaging				
1	52344172	Exonic, Promoter	NRD1 (includes S39C)	- - - - -	- - - - -	- - - - -	nonsynonymous	Damaging	Promoter Loss	Myf		
1	52840543	Exonic	ORC1 (includes F772S, F777S)	- - - - -	- - - - -	- - - - -	nonsynonymous	Damaging				
1	52851608	Exonic	ORC1 (includes T461M, T466M)	- - - - -	- - - - -	- - - - -	nonsynonymous	Damaging				3087
1	53535829	Exonic	PODN	L130P, L149P	- - - - -	- - - - -	nonsynonymous	Damaging				
1	53542912	Exonic, Intron/exon	PODN	N240S, N258S	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	53544513	Exonic	PODN	T350M, T473M	- - - - -	- - - - -	nonsynonymous	Tolerated				1256
1	53547748	Exonic	PODN	G492V, G615V	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	54518725	Exonic, Promoter	TCEANC2, TM	A46V	- - - - -	- - - - -	nonsynonymous	Tolerated	Promoter Loss	FOXC1		4129
1	55319801	Exonic	DHCR24	K376R	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	55337168	Exonic	DHCR24	P244L	- - - - -	- - - - -	nonsynonymous	Tolerated				
1	55344114	Exonic	DHCR24	P250L	- - - - -	- - - - -	Tolerated					

Gene Level Summaries – APC Gene (known)

Screenshot of the Ingenuity Variant Analysis (IVA) software interface showing gene level summaries for the APC gene.

Filter Cascade:

Variant Type	Count	Gene Count
Common Variants	47882	8338
Predicted Deleterious	17709	6389
Cancer Driver Variants	3569	1326
Genetic Analysis	1667	927
Biological Context	1638	913

Legend:

- Function: - Identical to Reference Genome, - Heterozygous Variant, - Heterozygous/Ambiguous, - Homozygous Variant, - Copy Number Gain/Heterozygous, - Copy Number Gain/Homozygous, - Nullizygous, - Gene Fusion, - No genotype
- Call Quality: <20, 20-30+, 30+

Table: Top 200 genes associated with APC

Name	#Variants	#Cases	%Cases	#Controls	%Controls
TG	13	14	70	2	10
APC	13	11	55	1	5
CFB	7	11	55	2	10
MUC4	20	10	50	1	5
PKHD1	10	10	50	5	25
BTNL2	9	10	50	2	10
HSPG2	12	9	45	4	20
THBS2	6	8	40	1	5
IKBKB	3	8	40	0	0
BIRC6	9	7	35	0	0
ABCB1	7	7	35	4	20
CACNA1S	6	7	35	3	15
MST1R	6	7	35	1	5
IGF2R	5	7	35	2	10
EGF (includes EG13645)	5	7	35	2	10
HIST1H4A (includes others)	4	7	35	1	5
ALOX5	2	7	35	1	5
MTU81	8	6	30	1	5
LAMA2	7	6	30	2	10
COL4A3	6	6	30	0	0
GRIN3A	6	6	30	0	0
MERTK	6	6	30	2	10
PDE11A	6	6	30	1	5
PDE4B	5	6	30	1	5
MCC	5	6	30	1	5
LAMA4	5	6	30	1	5
PLEC	5	6	30	1	5
SPTBN1	5	6	30	0	0

Right Panel: TG details

TG: thyroglobulin

Bottom Navigation:

- 20120410 MIR FFP...xls
- 20120410 MIR FFP...xls
- Show all downloads...

Protein Complexes Level (groups of proteins):

Top ranked protein complex: Growth Factor Receptor (18genes, 38 Variants, 80% of Cases)

Welcome Yury Gusev | Logout FAQ

My Samples | My Analyses rg_test1 [x] Feedback

Filter Cascade

Variants	Genes
71949	8777

Common Variants (47882 variants, 8338 genes)

Predicted Deletions (17709 variants, 6309 genes)

Cancer Driver Variants (3569 variants, 1326 genes)

Genetic Analysis (1667 variants, 927 genes)

Biological Context (1638 variants, 913 genes)

Add Filter Export Find groups/complexes Top 100 groups/complexes

Legend [hide]

Function	Cell Quality
Loss	<20
Normal	20-30
Gain	>30

- Identical to Reference Genome
- Heterozygous Variant
- Heterozygous/Ambiguous
- Homozygous Variant
- Copy Number Gain/Heterozygous
- Copy Number Gain/Homozygous
- Homozygous
- Nullozygous
- Gene Fusion
- No genotype

growth factor receptor genes

Name	p-value ▲	#Genes	#Variants	#Cases	%Cases	#Controls	%Controls
growth factor receptor	1.688E-15	16	38	18	90	5	25
Integrin	8.142E-6	9	18	16	80	5	25
MLH1-MSH2-MSH6-PMS2	1.041E-5	4	7	6	30	1	5
EPHA1B	1.162E-5	6	14	10	50	2	10
OSN-PI3K-PIP2-Src	1.226E-5	7	25	15	75	4	20
KU-WRN-Artemis-DNA Ligase IV-DNA-PK-XRCC4	1.759E-5	5	10	10	50	1	5
MutLo-MutS α -Exo1-FEN1	1.759E-5	5	8	6	30	2	10
alcohol group acceptor phosphotransferase	1.793E-5	11	17	9	45	1	5
EphA8 dimer	1.800E-5	6	14	10	50	2	10
EGFR/PDGFR/GFR	3.029E-5	4	11	14	70	2	10
PI3K (complex)	3.905E-5	6	24	14	70	3	15
phosphatidylinositol-4,5-bisphosphate 3-kinase	5.185E-5	5	23	16	80	4	20
Gpcr	6.015E-5	30	49	19	95	7	35
Fgfr	6.852E-5	4	7	5	25	0	0
MutLo-MutS α	6.852E-5	4	7	6	30	1	5
MutLo-MutS β	6.852E-5	4	7	5	25	1	5
IR52-PI3K	7.635E-5	6	24	14	70	3	15
Integrin	7.635E-5	6	14	11	55	3	15
SHP2-PI3K-GAB2	1.034E-4	6	24	14	70	3	15
IL-1/RTLR	1.229E-4	5	13	10	50	1	5
caspase	1.229E-4	5	7	10	50	2	10
KU-WRN-DNAPK-Artemis	1.329E-4	4	9	9	45	1	5
Ephb	1.329E-4	4	9	6	30	0	0
Vla-4	1.329E-4	4	10	8	40	2	10
Basc	1.802E-4	6	9	7	35	2	10
Laminin2	2.184E-4	3	13	9	45	3	15
Ephb dimer	2.319E-4	4	9	6	30	0	0
Plk	3.748E-4	4	13	12	60	1	5

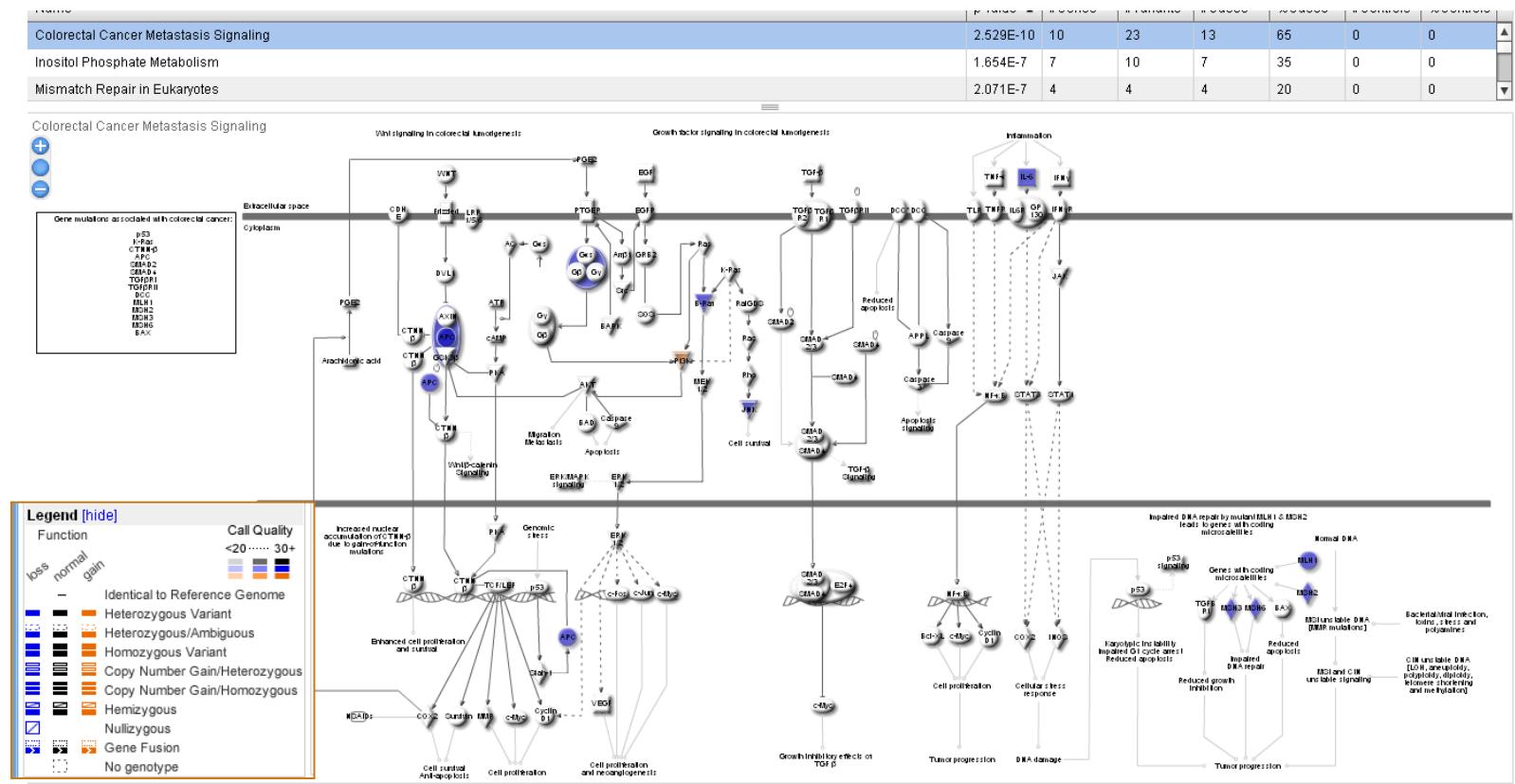
Contact Us | About Ingenuity Systems | Privacy Policy ©2012 Ingenuity Systems, Inc. All rights reserved

20120410 MIR FFP...xls 20120410 MIR FFP...xls Show all downloads... x

Pathway Level : 0% variants in control samples

Colorectal Cancer metastasis signaling pathway

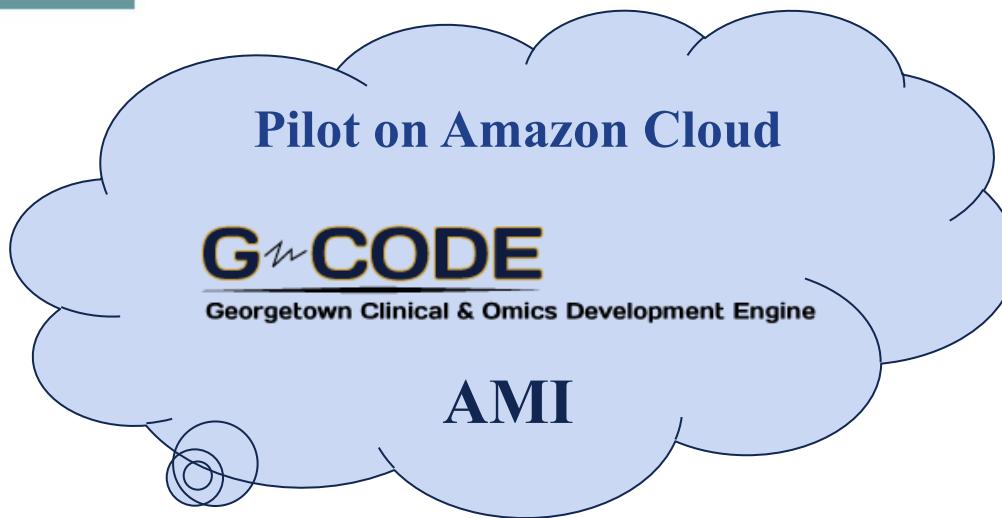
(color-coded genes showing type of mutation by color and quality of call by intensity)



Top Genes: GO enrichment With Ingenuity IPA

Name	p-value	# Molecules
Top Bio Functions		
• Infectious Disease	3.90E-04 - 2.81E-02	3
• Inflammatory Response	3.90E-04 - 3.76E-02	2
• Neurological Disease	3.90E-04 - 2.97E-02	3
• Cancer	1.17E-03 - 1.13E-02	1
• Gastrointestinal Disease	1.17E-03 - 1.70E-02	1
Top Molecular Functions		
• Antigen Presentation	3.90E-04 - 3.76E-02	1
• Cellular Movement	3.90E-04 - 3.76E-02	2
• Cell-To-Cell Signaling and Interaction	7.80E-04 - 2.93E-02	2
• Cellular Function and Maintenance	1.17E-03 - 3.42E-02	2
• Cell Death		

G-CODE : Clinical Omics Development Engine



- Cloud computing refers to the on-demand provision of computational resources via a computer network.
- Cloud computing is an attractive model for application deployment because it provides the following things
 - Agility
 - APIs
 - Reduced Cost
 - Device independent
 - Scalability
 - Performance
 - Security

Multidisciplinary Team

- Lombardi Comprehensive Cancer Center:
 - Dr. Louis Weiner's Lab, Dr Stephen Byers's Lab, Dr. John Marshall
 - Genomics, Cytogenetics and Metabolomics Shared Resources
- G-DOC Development Team:
 - Michael Harris, Andrew Shinohara, Kevin Rosso, Lavinia Carabet
- Analytical Group:
 - Georgetown - Yuriy Gusev, Krithika Bhuvaneshwar, Robinder Gauba, Lei Song
 - Virginia Tech: Joseph Wang
 - ISB: Ilya Shmulevich , Hector Rovira, Timo Erkkilä
- Funding
 - Georgetown Special Projects Initiative
 - NCI *In Silico* Research Centers of Excellence
 - NCI Center for Cancer Systems Biology (U54)
 - FDA Center for Regulatory Science

Madhavan et al. G-DOC®: A Systems Medicine Platform for Personalized Oncology. Neoplasia. 2011 Sep;13(9):771-83

ICBI: Discovery, Clinical Research, Clinical Care

- ‘omics’ data analysis
- EHR/PHR integration
- Biomarker discovery
- Disease Classification
- Genotype-Phenotype Correlation
- Biospecimen management and analysis
- Novel technology/method development

Clinical Research



Clinical Education



Basic/Traslational



Biospecimen

- MD/MS Systems Medicine
- Partnership
- Joint Grants
- Joint Appointments
- Contracted Services
- Shared Resources
- Medical Internships
- Joint Clinical Trials
- Shared IP
- Joint publications

email: sm696@georgetown.edu

G-DOC URL: <https://gdoc.georgetown.edu>

ICBI Website (coming soon): <http://icbi.georgetown.edu/>

Innovation Center for Biomedical Informatics (ICBI)

