LRpath analysis reveals common pathways dysregulated via DNA methylation across cancer types

Maureen A. Sartor

Computational Medicine and Bioinformatics

University of Michigan

Motivation

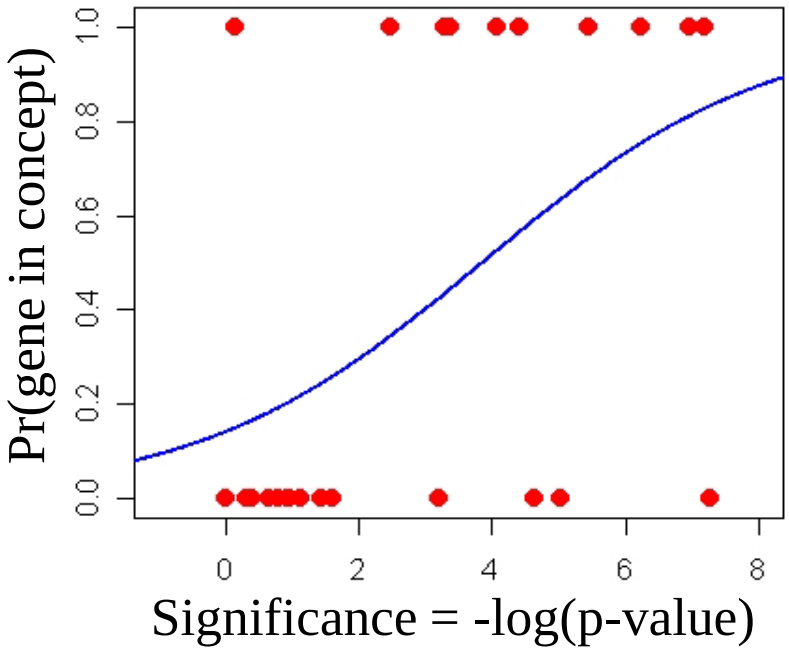
- The relative contribution of epigenetic mechanisms to carcinogenesis is not well understood - ***Do epigenetic mechanisms target similar genes and pathways as somatic mutations or different pathways?***
- Illumina HumanMethylation27 BeadChip platform assesses the percent methylation of over 27,000 CpG sites across the genome
- Several studies have been published testing for genes with aberrant methylation in their promoter regions. Interestingly, a majority of these publicly available datasets are studying cancer

Overview

- The time is ripe for an integrative analysis - used data from The Cancer Genome Atlas (TCGA) and NCBI's Gene Expression Omnibus (GEO).
- **Hypothesis**: During the pathogenesis of cancer, certain pathways or biological gene groups are commonly dysregulated via DNA methylation across cancer types.
- **Approach**: Employed *LRpath* and clustering analysis to unravel the commonly altered pathways and other biological concepts across 10 different cancer studies of DNA methylation data profiled using the Illumina Infinium HumanMethylation27 BeadChip.

LRpath method

Illustration of Logistic Regression Model Used



Repeat for all gene sets

Calculate:

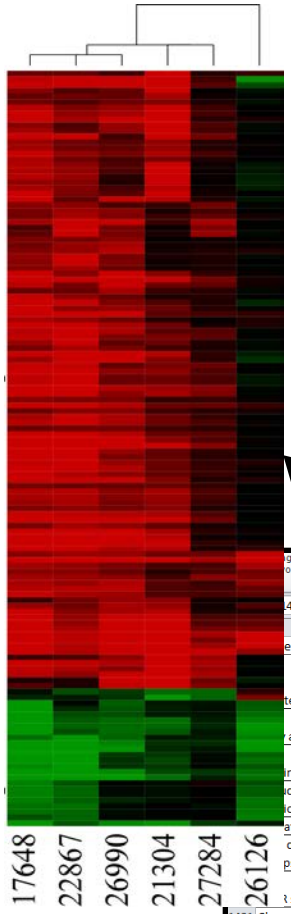
1. ***p*-value for H0: β = 0**
2. **Adjusted *p*-values for multiple comparisons (FDR)**
3. **Odds ratio that a gene belongs to the group for a significant vs. non-significant level**

Generate output

KEGG ID	Pathway Name	No. genes	Odds Ratio	LRpath FDR	DEG Entrez IDs (p<0.05)
hsa04610	Complement and coagulation cascades	54	3.34	0.0002	624, 629, 715, 716, 717, 720, 725, 730, 1361, 1380, 2152, 2155, 2157, 3075, 3426, 5054, 5104, 5328, 5624, 7056
hsa04510	Focal adhesion	137	2.31	0.0008	87, 88, 857, 858, 859, 894, 896, 1311, 1499, 2534, 3082, 3479, 3675, 3678, 3679, 3694, 3791, 3909, 3918, 4233, 5155, 5228, 6696, 7058, 7422, 7424, 9855, 10319, 56034
hsa04110	Cell cycle	71	2.54	0.0082	890, 891, 894, 896, 983, 1031, 1032, 1111, 1387, 4085, 4087, 4173, 4175, 5111, 5591, 7043, 10912, 11200, 85417
hsa04350	TGF-beta signaling pathway	50	2.69	0.0168	94, 268, 650, 652, 1311, 1387, 4052, 4086, 4087, 4091, 7043, 7058, 57154

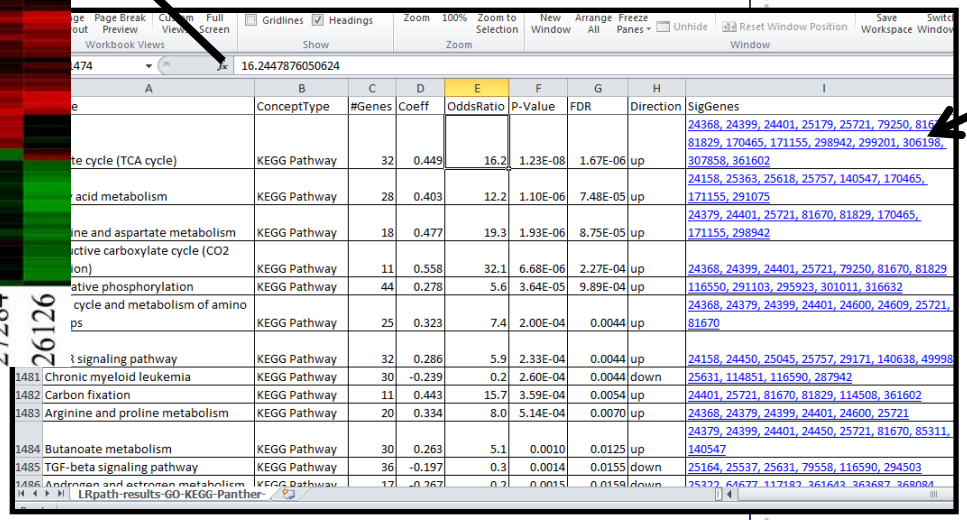
- **Sartor MA, et al. *Bioinformatics*. 2009; 25(2): 211-7.**

http://lrpath.ncibi.org



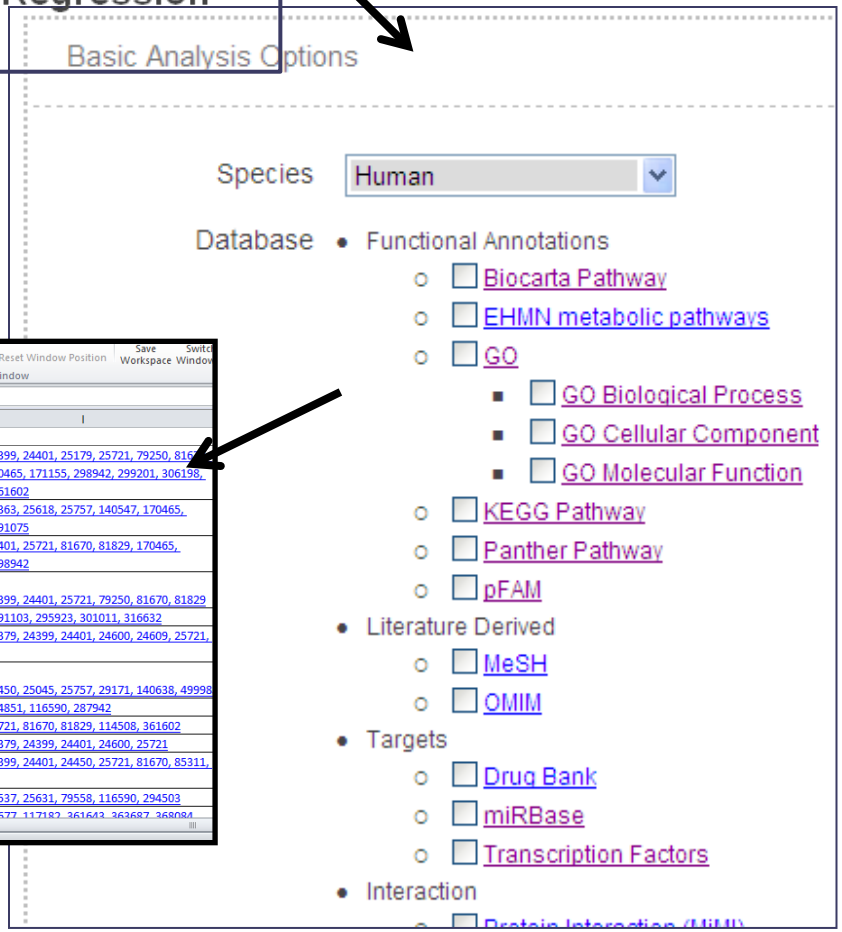
LRPath

Pathway Analysis using Logistic Regression



	A	B	C	D	E	F	G	H	I
	e	ConceptType	#Genes	Coeff	OddsRatio	P-Value	FDR	Direction	SigGenes
	te cycle (TCA cycle)	KEGG Pathway	32	0.449	16.2	1.23E-08	1.67E-06	up	24368, 24399, 24401, 25179, 25721, 79250, 81670, 81829, 170465, 171155, 298942, 299201, 306198, 307858, 361602
	acid metabolism	KEGG Pathway	28	0.403	12.2	1.10E-06	7.48E-05	up	24158, 25363, 25618, 25757, 140547, 170465, 171155, 291075
	ine and aspartate metabolism	KEGG Pathway	18	0.477	19.3	1.93E-06	8.75E-05	up	24379, 24401, 25721, 81670, 81829, 170465, 171155, 298942
	ion)	KEGG Pathway	11	0.558	32.1	6.68E-06	2.27E-04	up	24368, 24399, 24401, 25721, 79250, 81670, 81829
	ative phosphorylation	KEGG Pathway	44	0.278	5.6	3.64E-05	9.89E-04	up	116550, 291103, 295923, 301011, 316632
	cycle and metabolism of amino ps	KEGG Pathway	25	0.323	7.4	2.00E-04	0.0044	up	24368, 24379, 24399, 24401, 24600, 24609, 25721, 81670
	signaling pathway	KEGG Pathway	32	0.286	5.9	2.33E-04	0.0044	up	24158, 24450, 25045, 25757, 29171, 140638, 49998
1481	Chronic myeloid leukemia	KEGG Pathway	30	-0.239	0.2	2.60E-04	0.0044	down	25631, 114851, 116590, 287942
1482	Carbon fixation	KEGG Pathway	11	0.443	15.7	3.59E-04	0.0054	up	24401, 25721, 81670, 81829, 114508, 361602
1483	Arginine and proline metabolism	KEGG Pathway	20	0.334	8.0	5.14E-04	0.0070	up	24368, 24379, 24399, 24401, 24600, 25721
1484	Butanoate metabolism	KEGG Pathway	30	0.263	5.1	0.0010	0.0125	up	24379, 24399, 24401, 24450, 25721, 81670, 85311, 140547
1485	TGF-beta signaling pathway	KEGG Pathway	36	-0.197	0.3	0.0014	0.0155	down	25164, 25537, 25631, 79558, 116590, 294503
1486	Androgen and estrogen metabolism	KEGG Pathway	17	-0.267	0.2	0.0015	0.0159	down	25322, 64677, 117183, 361643, 363687, 368084

LRpath-results-GO-KEGG-Panther-



LRPath

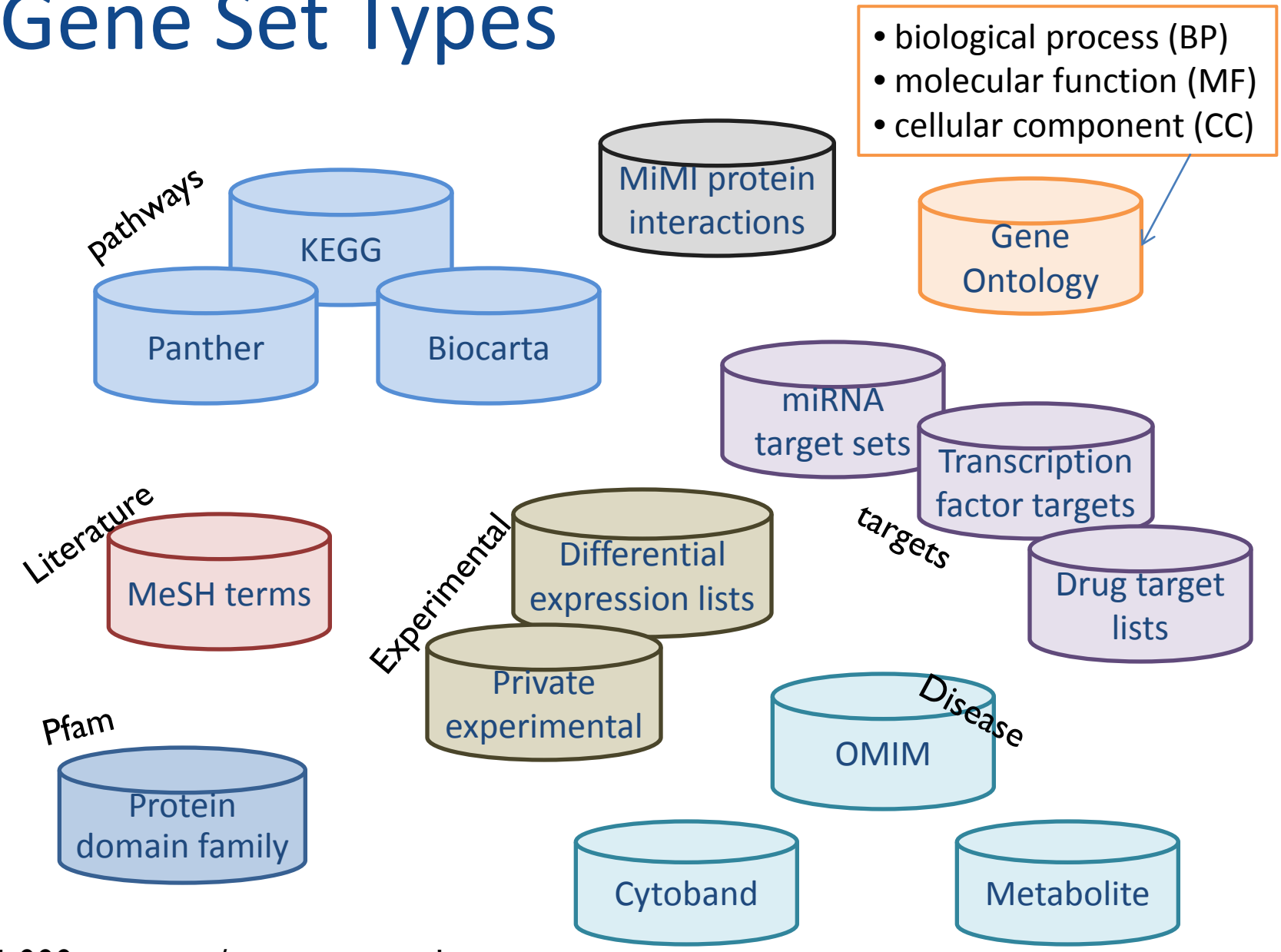
Pathway Analysis using Logistic Regression

http://lrpath.ncibi.org

Advantages of LRpath

- Strong performance for datasets with both large *and small* sample sizes
- Ability to test both 'directional' and 'non-directional' tests
- Random sets interpretation without the need for significance values to be "approximately normally distributed"
- Identical significance values for repeated runs (no dependence on permutations)
- Flat p-value distribution under the null (i.e. no significant sets)

Gene Set Types



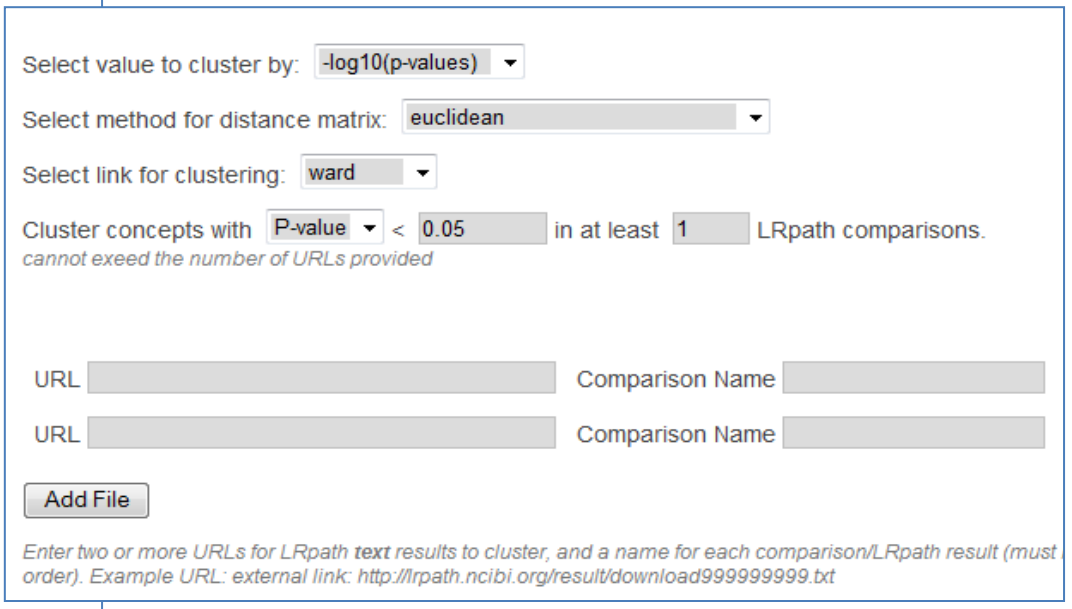
>21,000 concepts/gene sets total

LRPath

Pathway Analysis using Logistic Regression

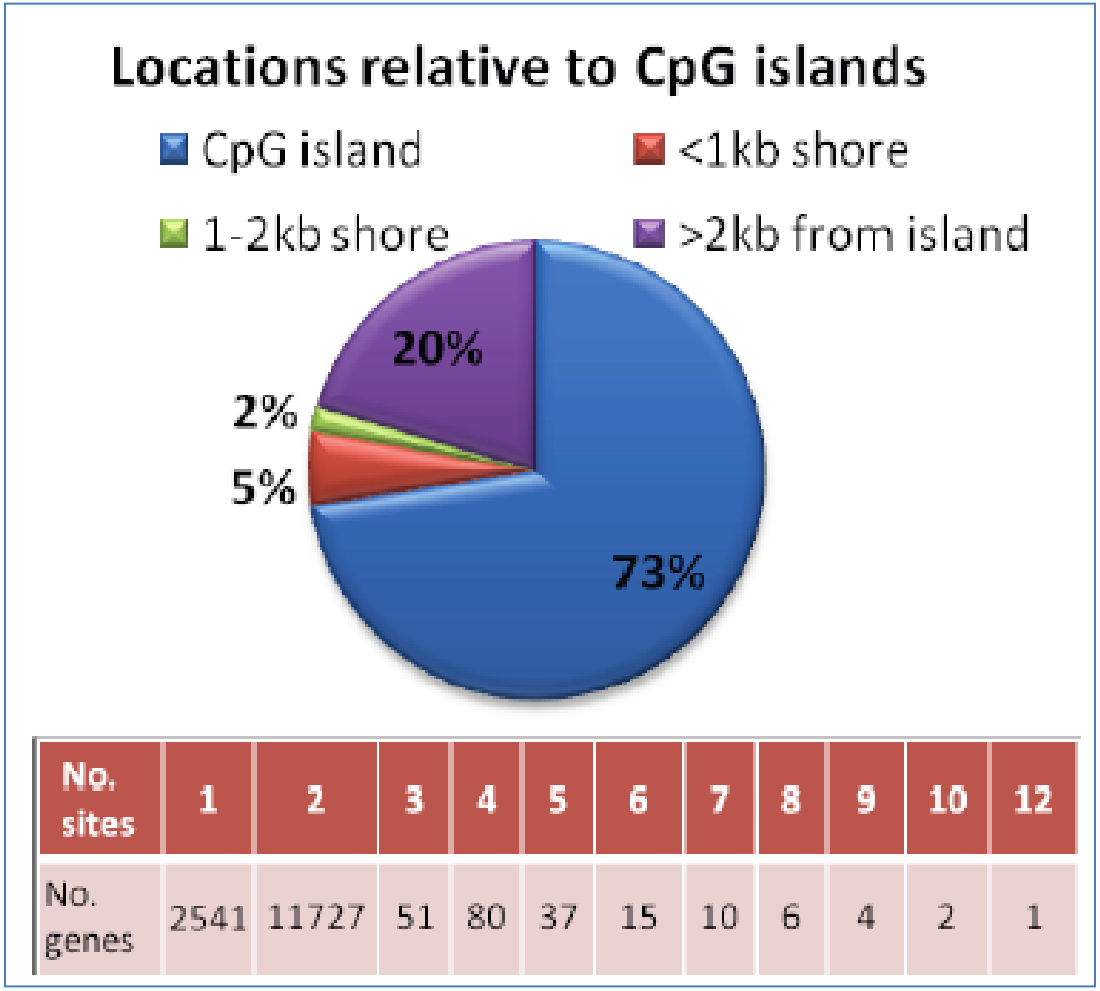
http://lrpath.ncibi.org

Clustering Options



Illumina Infinium HumanMethylation27 BeadChip

- Assesses percent methylation of >27,000 sites for >14,000 genes
- Most genes are represented by 1 or 2 sites on the array
- a small percent, (imprinted and cancer-related genes) are represented by up to a dozen sites.
- In addition, 110 miRNA promoters are covered by 254 sites.



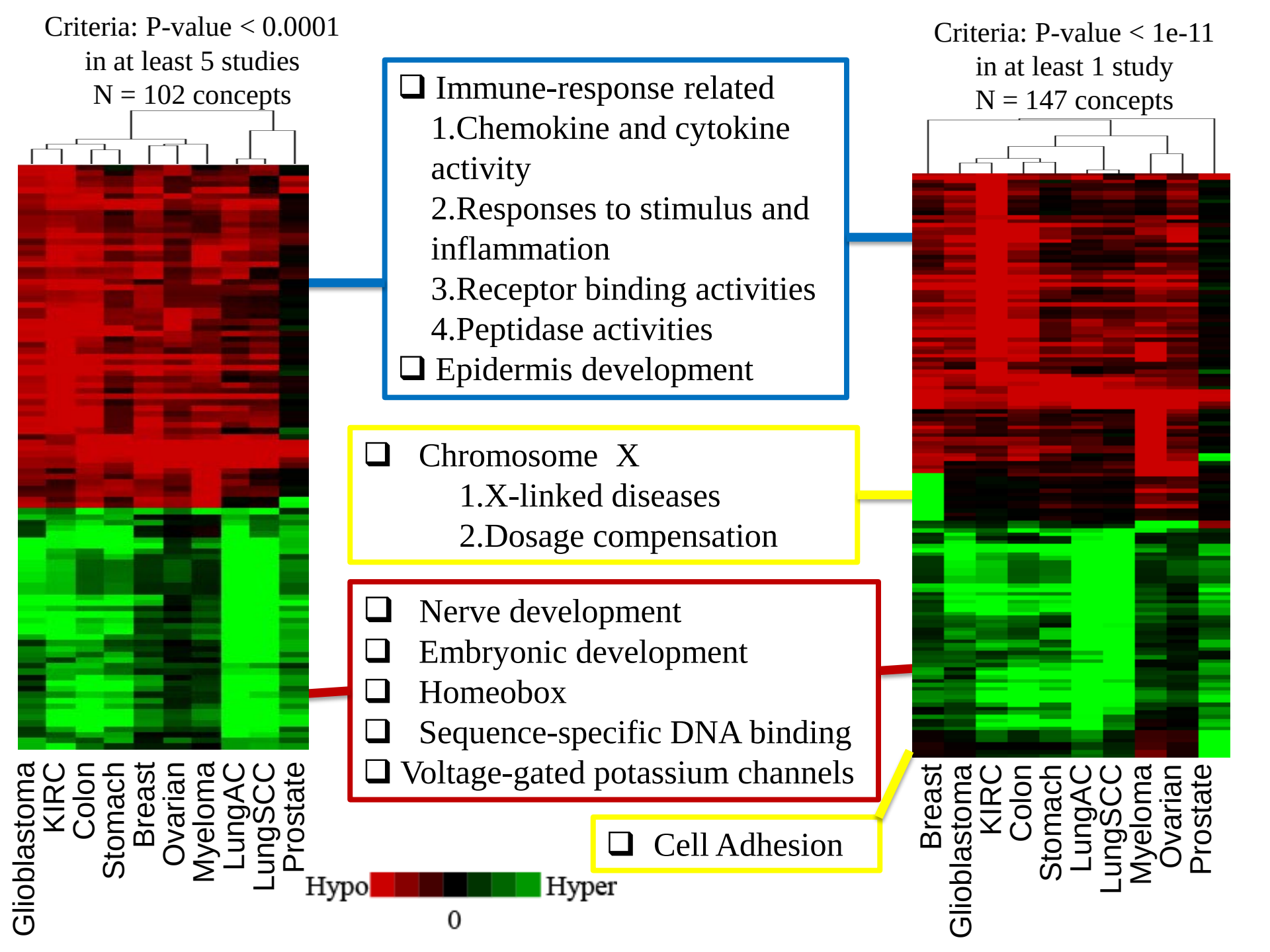
No. sites	1	2	3	4	5	6	7	8	9	10	12
No. genes	2541	11727	51	80	37	15	10	6	4	2	1

10 tumor vs. normal studies from TCGA and GEO

Source	GEO 17648	GEO 21304	GEO 22867	GEO 26126	GEO 26990	TCGA	TCGA	TCGA	TCGA	TCGA
Tumor Type	Colon	Multiple Myeloma	Glioblastoma	Prostate	Breast	Kidney	Lung AC	Lung SCC	Ovarian	Stomach
Normal Sample #	22	3	4	98	8	199	24	27	8	57
Cancer Sample #	22	161	77	95	47	199	24	27	39	57
P-value < 0.01	7922	4489	1403	9151	3699	10664	6419	6738	4376	8436
P-value < 0.01 and at least 10% change in average methylation	5642	4343	1179	3263	3039	2022	3847	3641	1900	3000

Genes harboring aberrant promoter methylation between normal and cancer samples were determined using an empirical Bayes method. GSE# IDs are provided for data from GEO.

Colon, Kidney, lung, and stomach cancers contained tumor/normal samples matched by patient.


Criteria: P-value < 0.0001
in at least 5 studies
N = 102 concepts
Criteria: P-value < 1e-11
in at least 1 study
N = 147 concepts
❑ Immune-response related
1.Chemokine and cytokine activity
2.Responses to stimulus and inflammation
3.Receptor binding activities
4.Peptidase activities
❑ Epidermis development
❑ Chromosome X
1.X-linked diseases
2.Dosage compensation
❑ Nerve development
❑ Embryonic development
❑ Homeobox
❑ Sequence-specific DNA binding
❑ Voltage-gated potassium channels
❑ Cell Adhesion
Glioblastoma
KIRC
Colon
Stomach
Breast
Ovarian
Myeloma
LungAC
LungSCC
Prostate
Breast
Glioblastoma
KIRC
Colon
Stomach
LungAC
LungSCC
Myeloma
Ovarian
Prostate
Hypo
Hyper
0

Significance of Overlap Between Pairs of Studies - hypomethylation

GO: Immune response

	Breast	Colon	Glioblastoma	KIRC	LungAC	LungSCC	Myeloma	Ovarian	Prostate	Stomach
Breast										
Colon	3.64E-16									
Glioblastoma	1.39E-09	7.84E-09								
KIRC	7.26E-15	2.90E-29	1.43E-17							
LungAC	2.40E-22	4.90E-22	1.72E-12	4.03E-21						
LungSCC	1.38E-20	1.97E-21	3.53E-10	3.94E-30	5.56E-49					
Myeloma	2.55E-14	7.28E-16	2.45E-08	1.01E-16	1.16E-24	7.26E-25				
Ovarian	1.15E-16	2.37898E-14	3.13E-03	8.41E-11	1.17E-16	5.76E-26	5.75E-11			
Prostate	2.48E-03	0.25	0.19	0.50	0.04	0.09	0.18	0.14		
Stomach	9.40E-09	4.20E-22	9.07E-04	9.31E-14	8.69E-14	1.36E-12	6.32E-08	1.77E-14	0.15	

GO: Epidermis development

	Breast	Colon	Glioblastoma	KIRC	LungAC	LungSCC	Myeloma	Ovarian	Prostate	Stomach
Breast										
Colon	3.38E-06									
Glioblastoma	4.46E-08	6.44E-04								
KIRC	2.46E-08	1.37E-07	1.96E-09							
LungAC	1.14E-08	2.46E-11	3.81E-06	7.39E-09						
LungSCC	1.46E-04	1.03E-03	3.99E-03	9.32E-08	8.61E-10					
Myeloma	4.73E-08	5.23E-04	3.31E-07	2.93E-12	1.63E-08	2.69E-08				
Ovarian	4.42E-06	3.01E-04	3.02E-06	6.76E-06	8.89E-08	2.84E-10	1.13E-05			
Prostate	0.09	0.16	0.51	0.45	0.14	4.72E-02	0.27	0.44		
Stomach	6.30E-08	3.31E-07	2.43E-02	5.28E-05	9.12E-06	1.75E-02	3.33E-03	8.84E-04	4.64E-02	

* Red indicates p-value less than 0.05

Significance of Overlap Between Pairs of Studies - hypermethylation

GO: Neurogenesis

	Breast	Colorectal	Glioblastoma	KIRC	LungAC	LungSCC	Myeloma	Ovarian	Prostate	Stomach
Breast										
Colorectal	6.59E-13									
Glioblastoma	1.27E-09	2.03E-03								
KIRC	1.66E-16	1.22E-08	2.66E-04							
LungAC	7.77E-22	2.08E-25	1.51E-03	1.82E-27						
LungSCC	1.78E-22	4.05E-17	1.16E-07	3.90E-24	4.96E-40					
Myeloma	0.21	0.21	2.35E-02	0.40	0.75	0.10				
Ovarian	0.45	3.15E-03	0.39	0.44	0.35	0.22	0.81			
Prostate	9.31E-17	8.50E-16	1.89E-03	5.25E-13	2.20E-29	2.59E-21	0.47	0.48		
Stomach	1.03E-13	1.95E-33	2.05E-04	2.14E-13	7.76E-30	1.49E-18	0.48	1.40E-02	1.25E-05	

* Red indicates p-value less than 0.05

Conclusions

- Pathways affected by differential methylation were surprisingly concordant across cancer types
  - Promoters of genes involved in voltage-gated potassium channels, which play a role in cell proliferation processes, tend to be hypermethylated.
  - Genes in developmental concepts such as homeobox, embryonic and nerve development tend to be hypermethylated (many PRC2 target genes)
  - Genes in epidermis development and keratinization are hypomethylated.
  - Immune response concepts identified by GO, KEGG pathways, and MeSH terms are hypomethylated (elevated immune response is a commonly affected mechanism across multiple cancer types.)
- For most tumor types, similar genes are affected by a change in CpG methylation in a pathway.
  - The same significant pathways could be affected by different sets of methylated genes across various cancer types. However, for tested biological concepts, they appear to be mostly the same genes, with a few exceptions.

Conclusions

- DNA repair, one of the most commonly affected pathways in cancer development, is *depleted* in differentially methylated genes.
  - We hypothesize that genes involved in DNA damage and cell cycle tend to be dysregulated by alternative mechanisms such as genomic aberrations.
- Performing an integrative analysis of biological concepts dysregulated via methylation across ten cancer types, we identified concepts affected in multiple cancer types that support biologically important findings.
- A subset of the known cancer pathways appears to be commonly dysregulated via DNA methylation across cancers

Acknowledgements

- National Center for Integrative Biomedical Informatics (NCIBI)
  - Brian Athey (PI)
  - Gilbert S. Omenn
  - Terry Weymouth
  - Vasu Mahavisno
  - Alla Karnovsky
- Collaborators
  - Laura Rozek
  - Dana Dolinoy
- Lab Members
  - Julie (Jung) Kim