A Bioinformatics Approach to Gene Expression and Gene Regulation Analysis in Cancer

13 April 2006

Obi L. Griffith

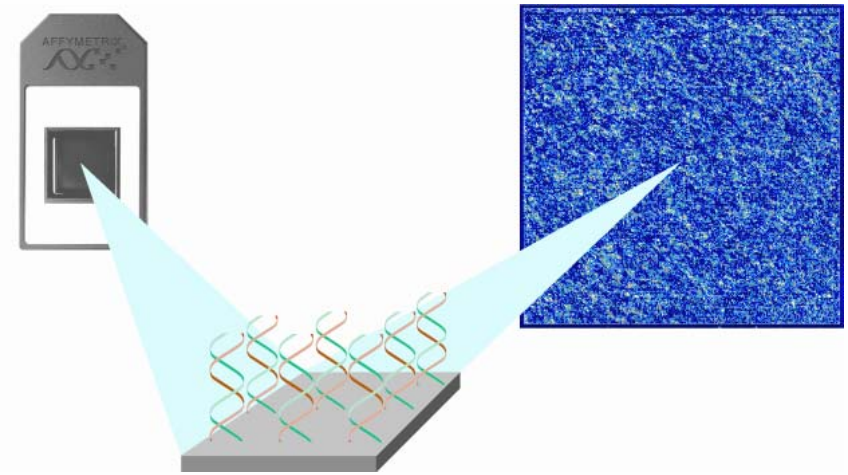
VanBUG student presentation

Supervisor: Steven Jones

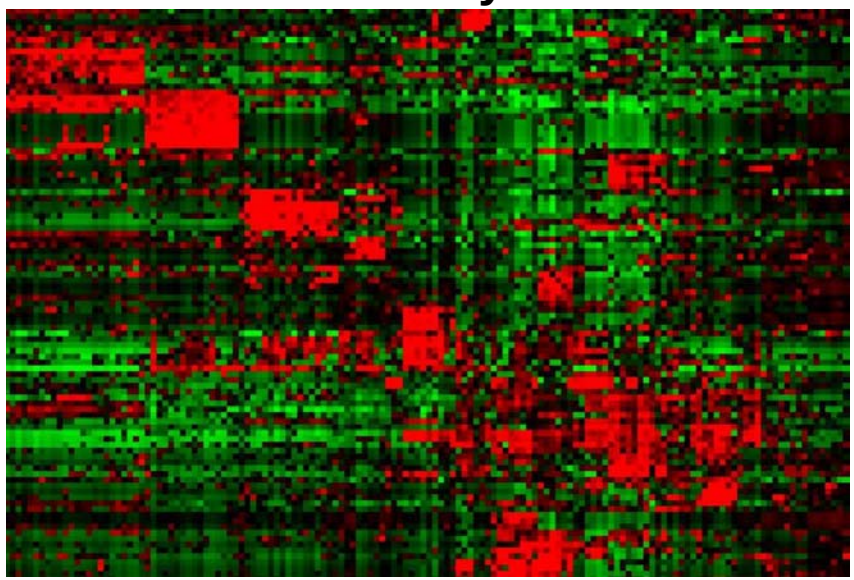
How can we use gene expression data to investigate cancer?

I. Multi-platform Coexpression
II. Multi-platform differential expression – Thyroid cancer
III. Differential Coexpression – Prostate cancer
IV. ORegAnno – Open Regulatory Annotation

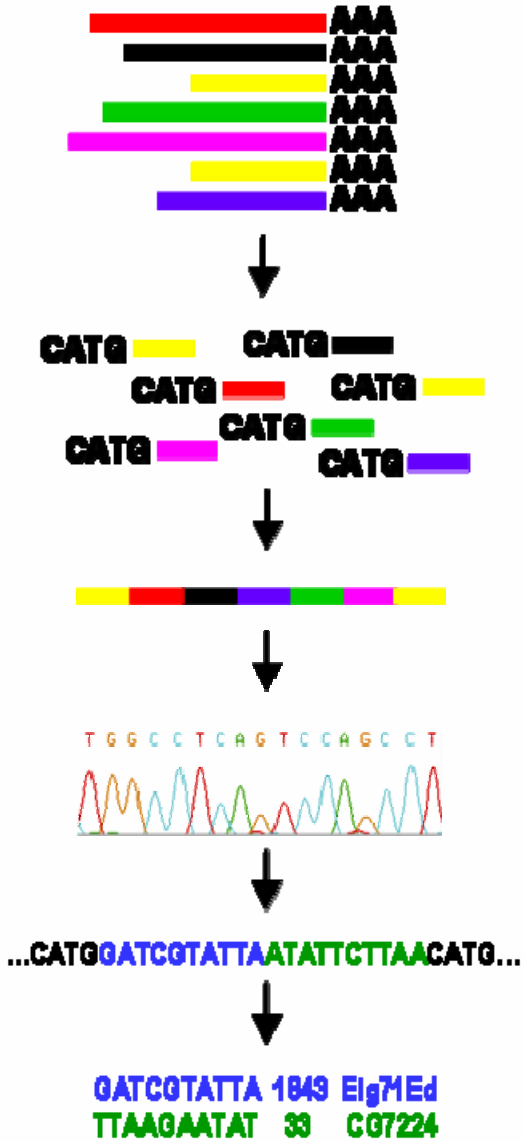
A. Oligonucleotide arrays



B. cDNA microarrays



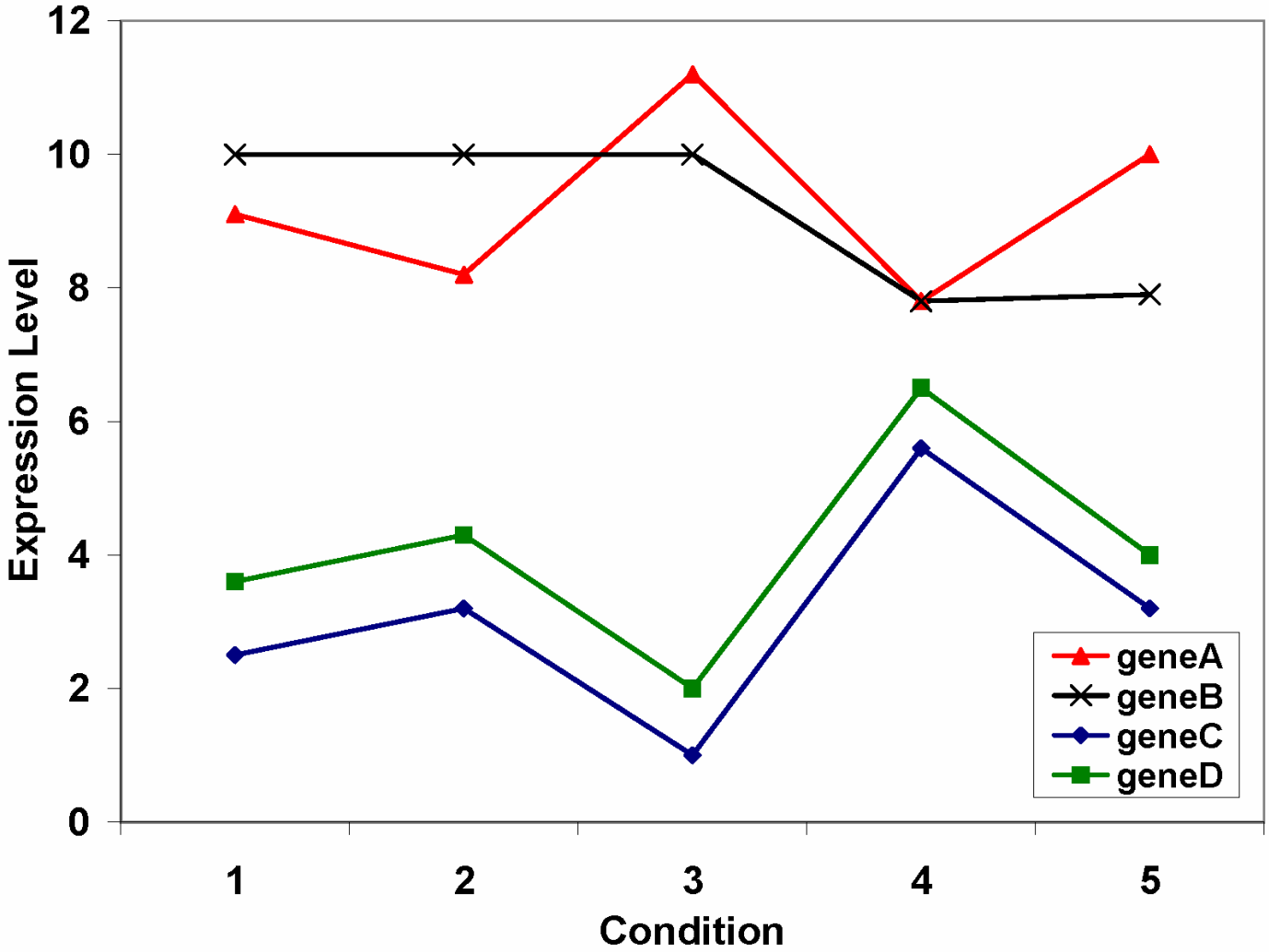
C. SAGE



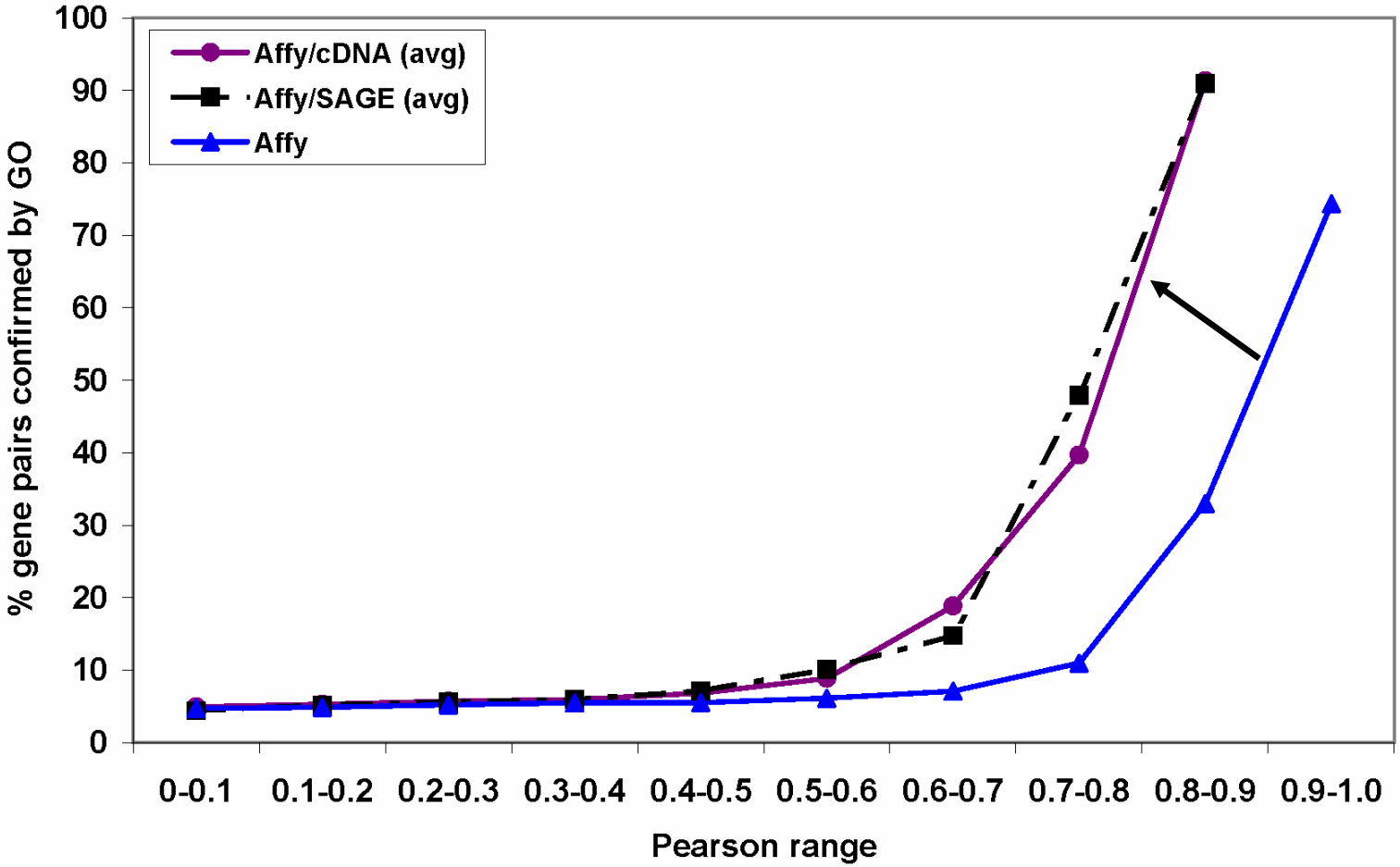
Which expression data?

Species	Platform	Experiments	Unique genes
H. sapiens	SAGE (short)	272	20,312
	Oligo. Array	1,640	12,452
	cDNA microarray	2,852	13,111
M. musculus	SAGE (long)	85	12,715
	Oligo. Array	1,802	8,164
	cDNA microarray	366	8,102
C. elegans	SAGE (long/short)	26	15,867
	cDNA microarray	1,059	16,682
	Total	**8,102**	**50,797**

I) What is Coexpression? Why do we care?

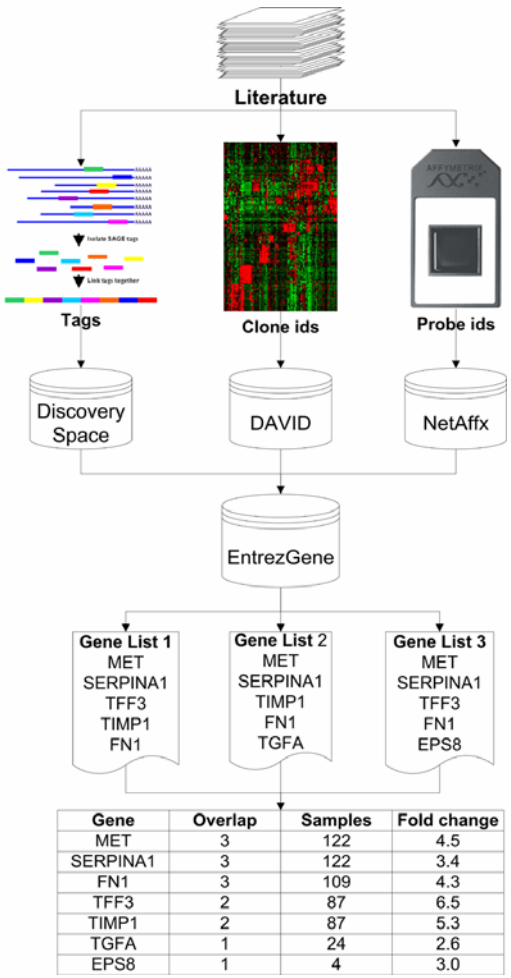


Coexpression methods that combine different platforms or datasets improve 'quality' of predictions (according to GO)



OL Griffith, ED Pleasance, DL Fulton, M Oveisi, M Ester, AS Siddiqui, SJM Jones. 2005. Assessment and Integration of Publicly Available SAGE, cDNA Microarray, and Oligonucleotide Microarray Expression Data for Global Coexpression Analyses. Genomics. 86:476-488

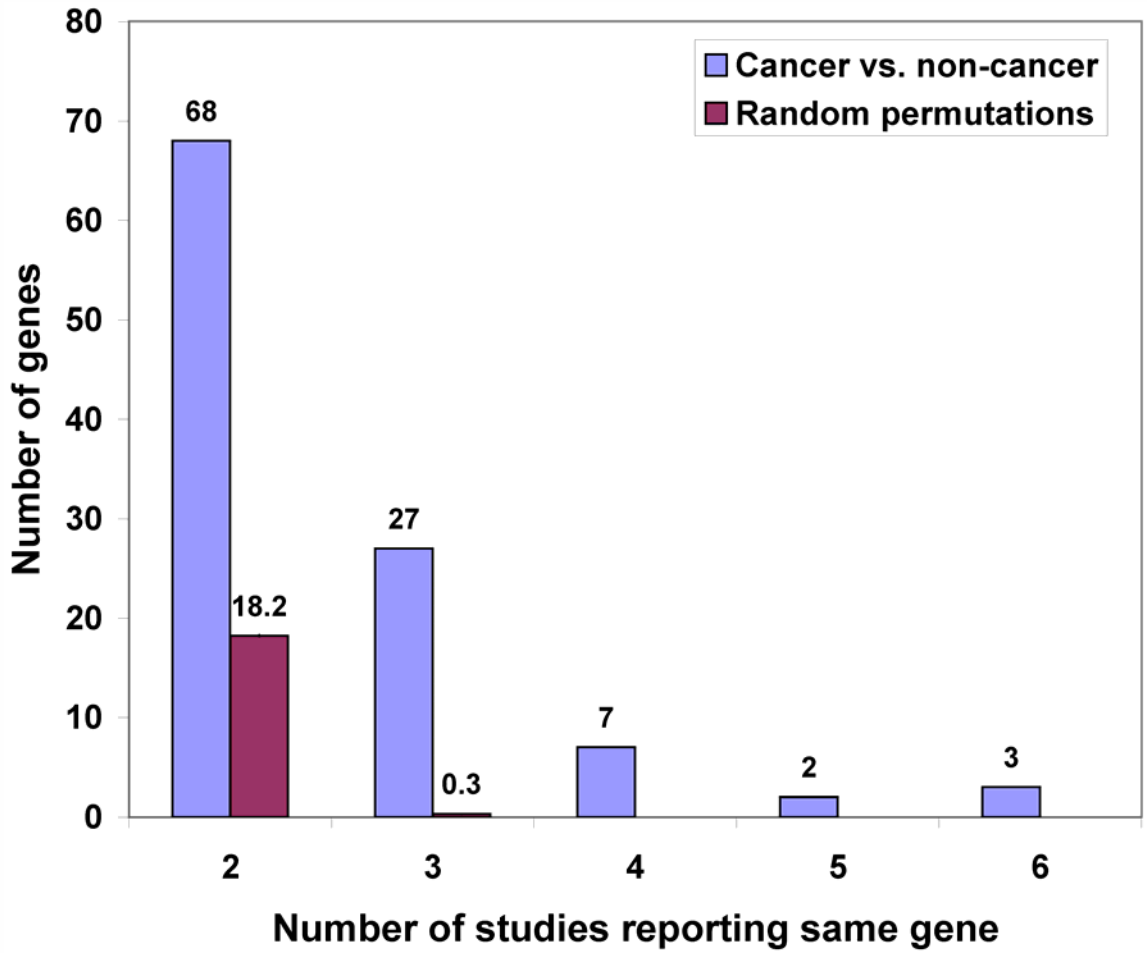
II) Multi-platform approach for differential expression in thyroid cancer



Gene	Overlap	Samples	Fold change
MET	3	122	4.5
SERPINA1	3	122	3.4
FN1	3	109	4.3
TFF3	2	87	6.5
TIMP1	2	87	5.3
TGFA	1	24	2.6
EPS8	1	4	3.0

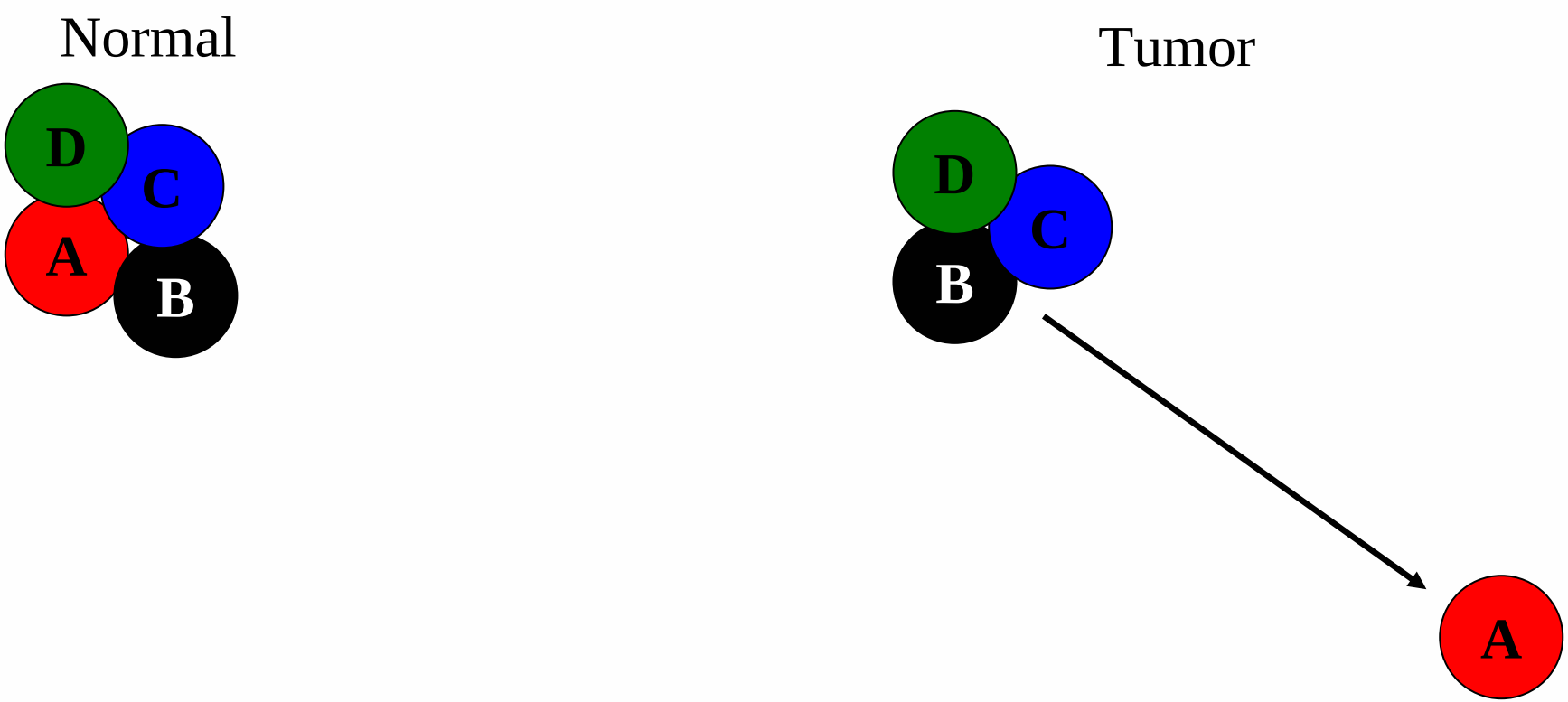
- Collaboration with Dr. Wiseman
- Collect and curate data from over 20 studies
- Analyze datasets for overlap
- Rank genes according to amount of overlap, size of studies, and fold change

A significant number of genes are consistently reported as differentially expressed from multiple independent studies

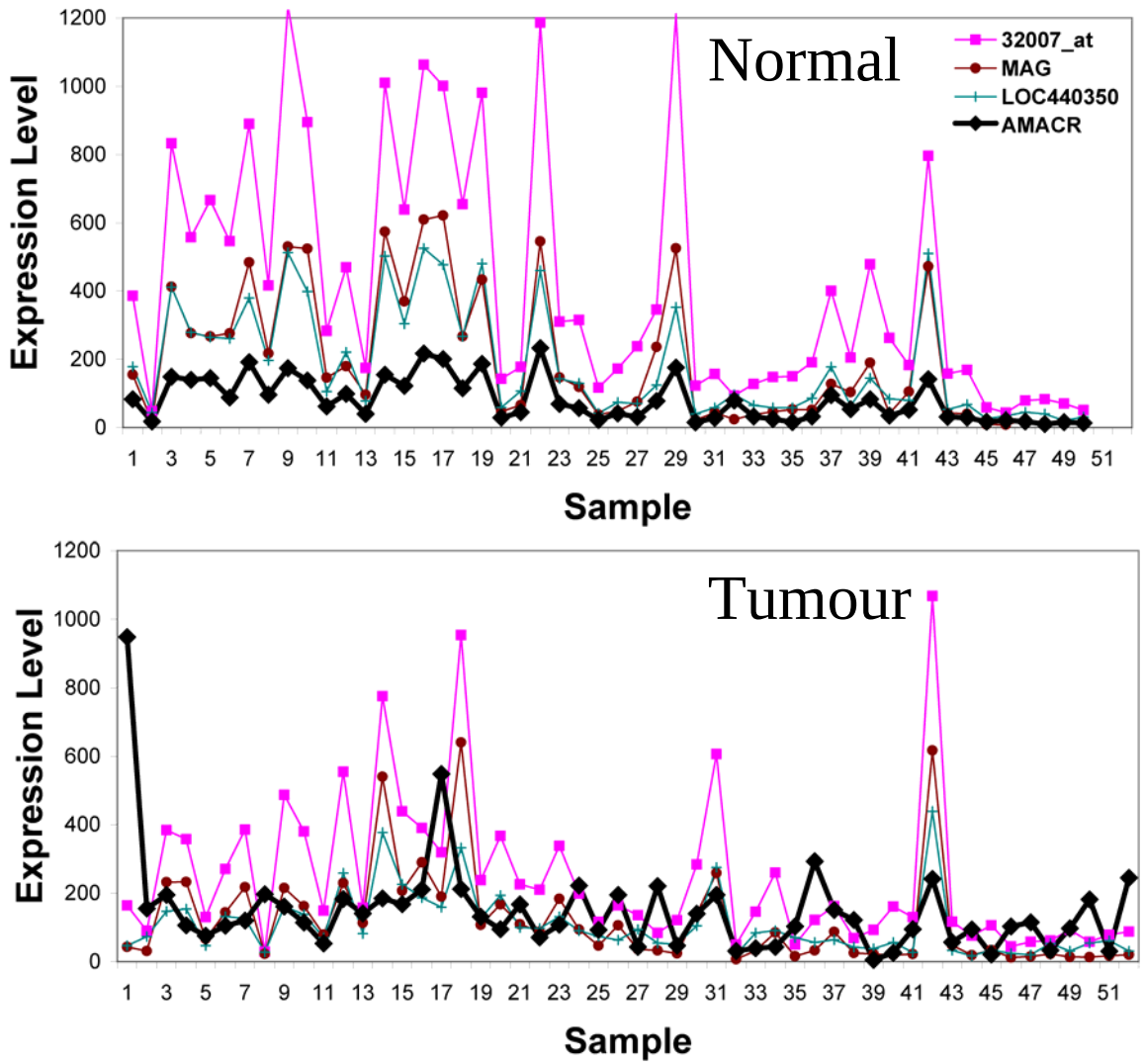


OL Griffith, A Melck, SJ.M Jones and SM Wiseman. A Meta-analysis and Meta-review of Thyroid Cancer Gene Expression Profiling Studies Identifies Important Diagnostic Biomarkers. Under review. JCO.

III) Differential coexpression in cancer



An example of differential coexpression in prostate cancer (AMACR)



IV) ORegAnno - www.oreganno.org

SB Montgomery*, OL Griffith*, MC Sleumer, CM Bergman, M Bilenky, ED Pleasance, Y Prychyna, X Zhang, SJM Jones. 2006. ORegAnno: An open access database and curation system for literature-derived promoters, transcription factor binding sites, and regulatory variation. Bioinformatics. 22(5):637-40.

Summary

- Coexpression of large public expression datasets can identify genes with common biological processes
- Using multi-platform methods improves results
- Multi-platform methods also useful to identify consistent differential expression
- Differential coexpression represents a novel alternative to differential expression
- ORegAnno – an open resource for regulatory annotation

Acknowledgements

Supervisor
Dr. Steven Jones

Coexpression analysis
Yuliya Prychyna
Maggie Zhang
Yan Jia Pan
Erin Pleasance
Debra Fulton

Thyroid meta-analysis
Sam Wiseman
Adrienne Melck

Subspace clustering
Byron Gao
Martin Ester

Differential coexpression
Erin Pleasance
Malachi Griffith

ORegAnno
Stephen Montgomery
Yuliya Prychyna
Maggie Zhang
Misha Bilenky
Monica Sleumer
Casey Bergman

Michael Smith Foundation for
Health Research

GENOME Sciences Centre

BC Cancer Agency
CARE & RESEARCH

Thanks!

Questions?

Supplemental slides

Difference in Mean correlation

Norm	Exp1	Exp2	Exp3	Exp4	Exp5	…
geneA	1.2	1.3	-1.4	0.1	2.2	…
geneB	1.3	1.3	-0.9	0.1	2.3	…
geneC	-1.2	1.0	0.1	0.5	1.4	…
…	…	…	…	…	…	…

Tumor	Exp1	Exp2	Exp3	Exp4	Exp5	…
geneA	11	35	2	4	50	…
geneB	12	35	0	3	47	…
geneC	0	10	4	15	20	…
…	…	…	…	…	…	…

Calculate all correlations for each gene

Norm	geneA	geneB	geneC	geneD	…
geneA	NA	0.91	0.01	0.99	…
geneB	0.91	NA	-0.03	0.87	…

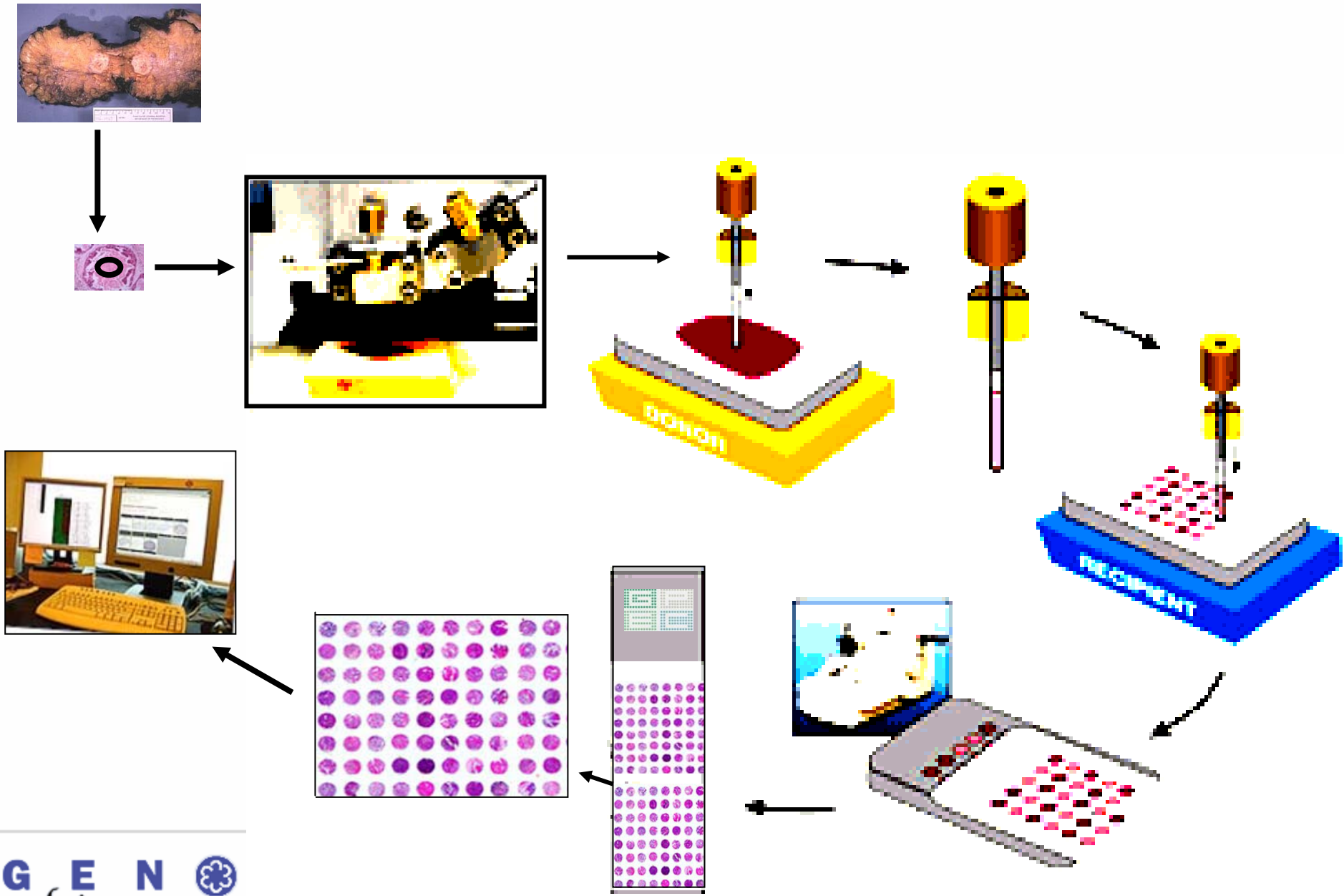
Tumor	geneA	geneB	geneC	geneD	…
geneA	NA	0.31	0.01	0.23	…
geneB	0.31	NA	-0.03	0.90	…

Find n nearest genes in normal and compare to tumor

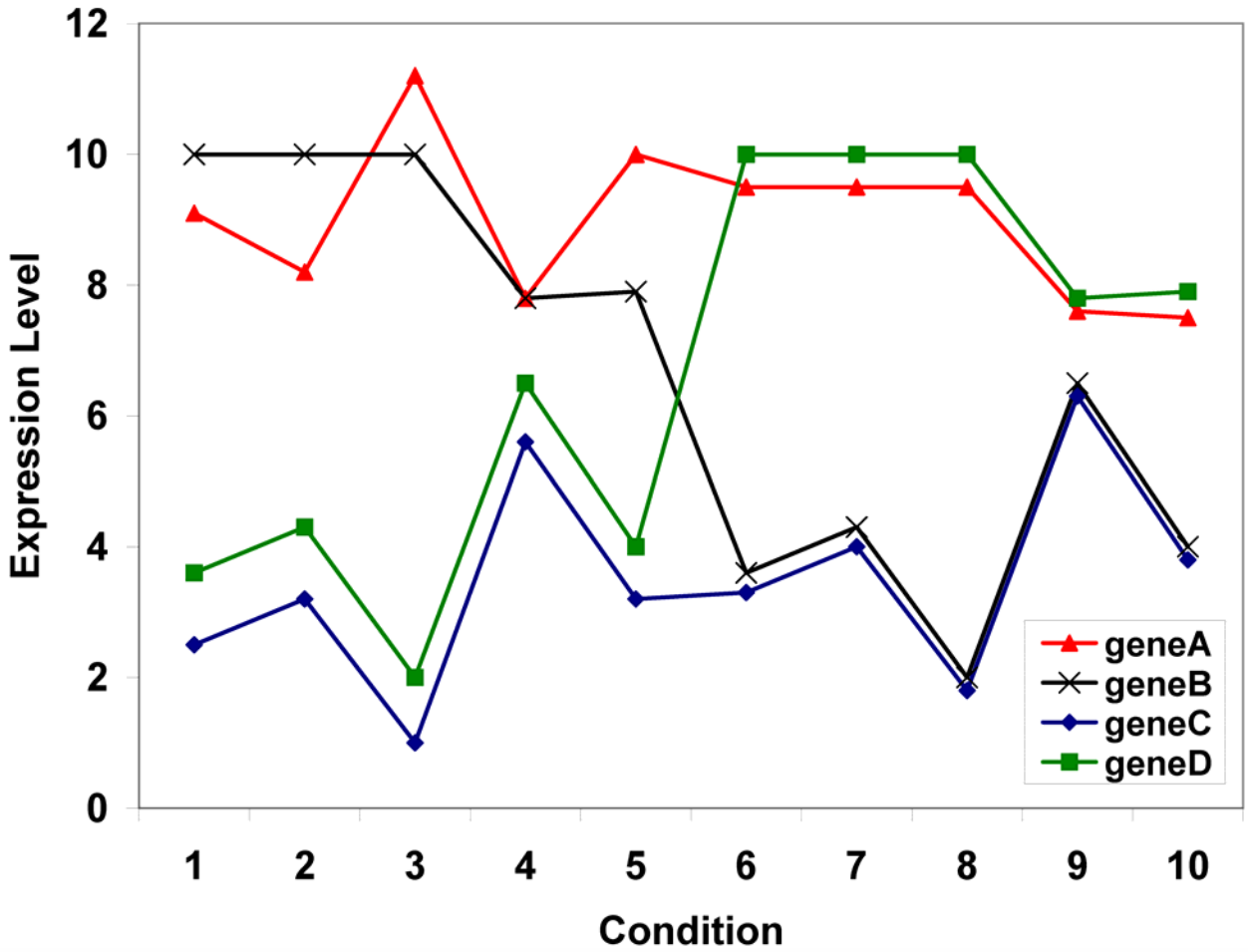
Norm	geneD	geneB	geneX	geneY	…
geneA	0.99	0.91	0.90	0.89	…

Tumor	geneD	geneB	geneX	geneY	…
geneA	0.23	0.31	0.18	0.01	…

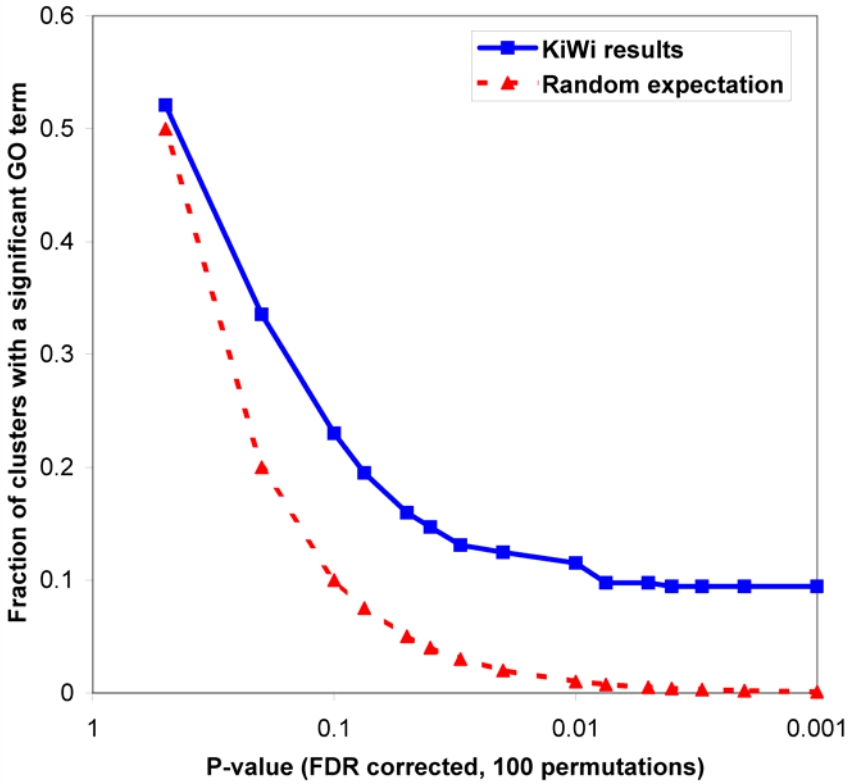
Methods: Tissue Array Construction

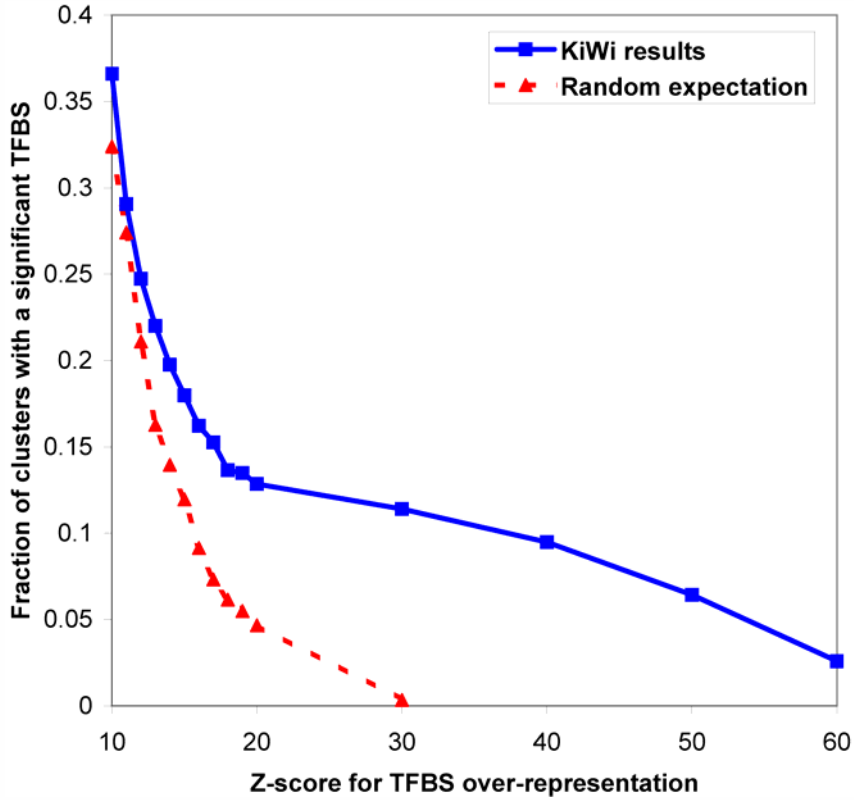


Subspace clustering – what is it?



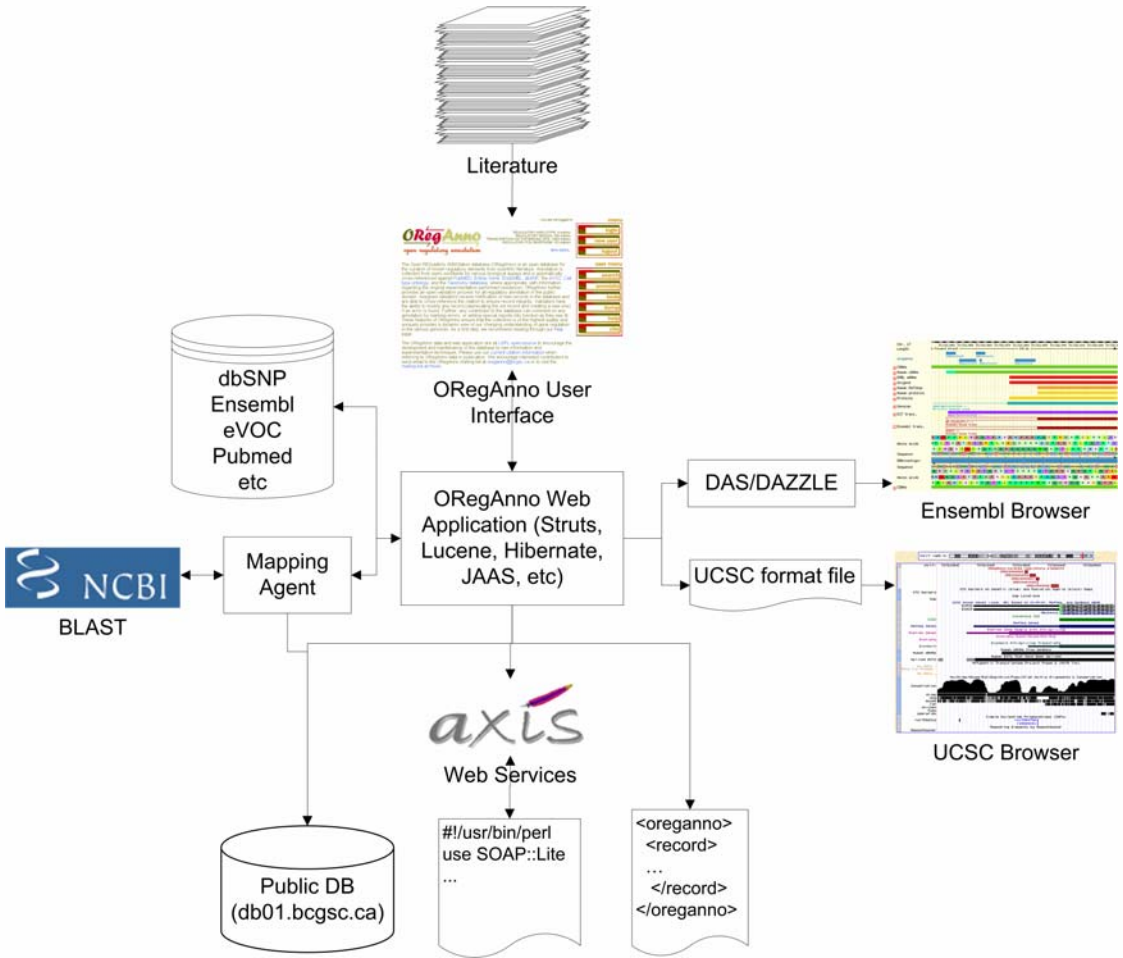
Subspace clustering – does it work?





BJ Gao, OL Griffith, M Ester, SJM Jones. Discovering significant OPSM subspace clusters in massive gene expression data. Under Review. KDD. 10 March, 2006.

www.oreganno.org



SB Montgomery*, OL Griffith*, MC Sleumer, CM Bergman, M Bilenky, ED Pleasance, Y Prychyna, X Zhang, SJM Jones. 2006. ORegAnno: An open access database and curation system for literature-derived promoters, transcription factor binding sites, and regulatory variation. Bioinformatics. 22(5):637-40.