UQ-Brisbane SISG

Module 9: Gene Expression and Epigenetic Profiling

SISG Brisbane 2017

The University Of Queensland AUSTRALIA | IMB

Tuesday February 14, 2017

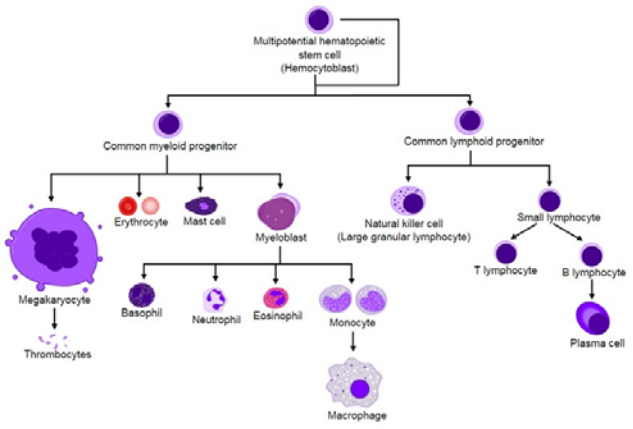
"Immunogenomics"

Greg Gibson

Georgia Institute of Technology

greg.gibson@biology.gatech.com

Why Blood Gene Expression has such a high correlation structure



1. Because there are 3 common and dozens of rare blood cell types, and any cell-type biased gene expression correlates with abundance of the cell-type.
2. Because the environment, including disease status, modulates the expression of up to thousands of genes in a coordinated manner
3. The genetic component of most individual transcript abundance is regulated in trans, which also tends to lead to covariance – eg *Stat1* mediates the interferon response

Wikipedia: White Bllood Cells

Blood Cell-type De-convolution

Some software:	CellMix:	https://web.cbio.uct.ac.za/~renaud/CRAN/web/CellMix/
	csSAM	https://cran.r-project.org/web/packages/csSAM/
	DeconSeq	https://www.bioconductor.org/packages/release/bioc/html/DeconRNASeq.html
	CIBERSORT	http://cibersort.stanford.edu/

General Approach:

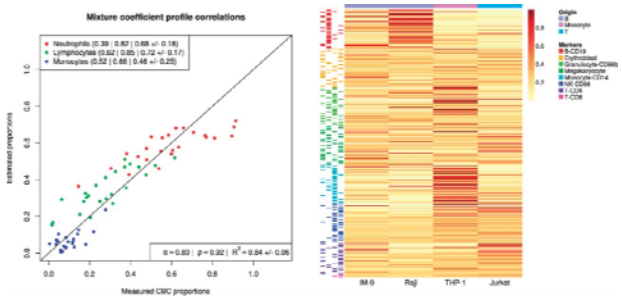
1. Identify cell-type specific genes in comparisons of purified cells
2. Generate the PC of those genes in the mixed blood sample
3. Estimate abundance of each cell type from the PC measures benchmarked against known or artificial mixtures

One issue is cross-platform comparability of gene lists – Illumina, Affymetrix, and RNASeq have different probes

CellMix

Chooses algorithmically from 7 different deconvolution methods

Uses 8 marker gene lists compiled from 11 public databases



Works even though these are cell lines, not primary cells from blood

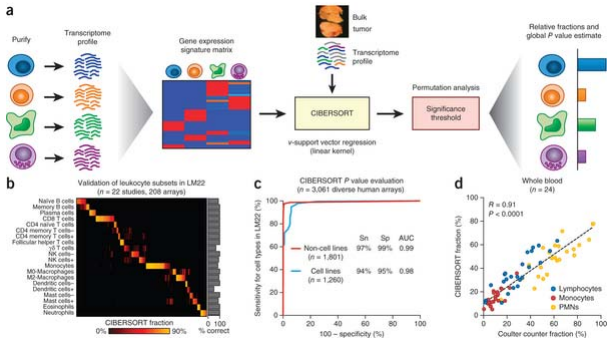
Note that Neutrophils 40%-80%; Lymphocytes 20%-50%; Monocytes 5%-20% are fairly typical ranges

Gaujoux and Seoighe (2013) *Bioinformatics* **29**: 2211-12 "CellMix: a comprehensive toolbox for gene expression deconvolution"

CIBERSORT

Existing deconvolution methods perform accurately on distinct cell subsets in mixtures with well-defined composition (for example, blood), but are considerably less effective for discriminating closely related cell types (for example, naïve vs. memory B cells).

Input = reference gene expression signatures and unknown profile
Algorithm= linear support vector regression (SVR) – a machine learning approach robust to noise
Output = estimated abundances and p-value for the deconvolution

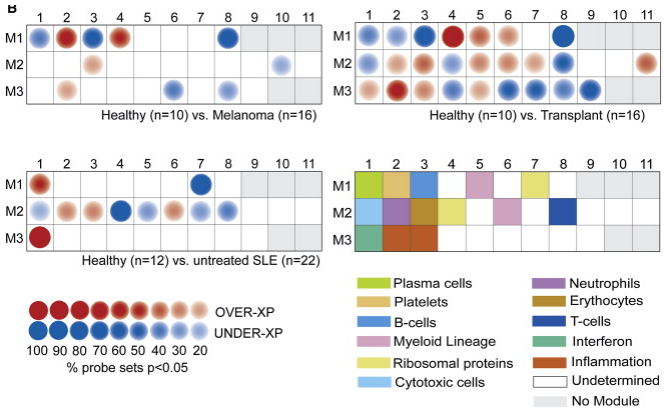


Newman et al (2015) *Nature Methods* **12**: 453-457 "Robust enumeration of cell subsets from tissue expression profiles"

Chaussabel Modules

Used k-means clustering to search for conserved modules of genes that are differentially expressed in 8 diseases, namely 239 samples for SLE, JIA, T1D, melanoma, 2 types of bacteremia, influenza, or liver transplantation

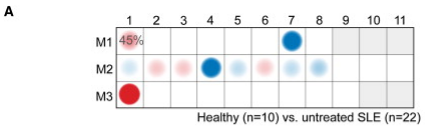
Identified 28 modules involving 4742 transcripts (average of 170 per module)



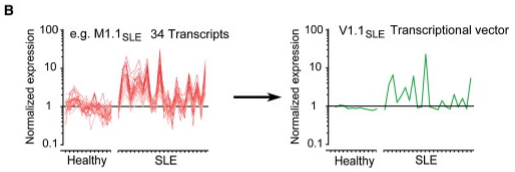
Chaussabel et al (2008) *Immunity* **29**: 150-164 "A modular analysis framework for blood genomics studies: application to SLE"

Application to study of Pediatric Lupus

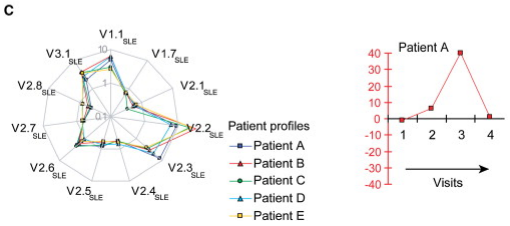
4 Modules are particularly strongly associated with SLE. The dots are proportional to the fraction of transcripts differentially expressed (DE) between cases and controls.



This plot shows the average expression of the DE genes in the 3.1 module in 10 controls and 22 patients. Averaging these yields the vector for the module across patients.

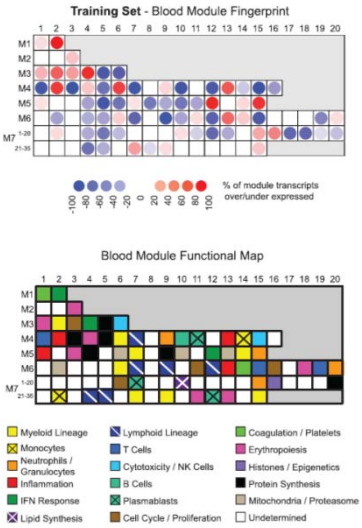


The radar plot shows for five patient the average value for each module. Similarly you can also plot changes over time. This analysis confirms elevation of 1.1 and 3.1, but suppression of 1.7 and 2.4 in the patients.



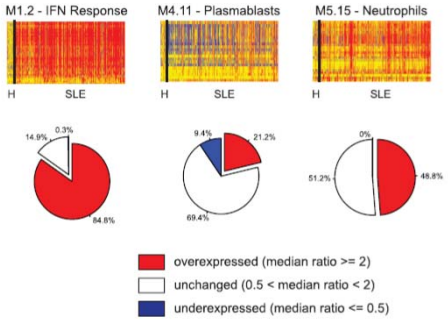
Chaussabel et al (2008) *Immunity* **29**: 150-164 "A modular analysis framework for blood genomics studies: application to SLE"

Update to 95 modules in 2016



158 Pediatric SLE patients
924 longitudinal PB profiles (avg ~ 6 per patient)

First asked how modules correlate with disease, and how many patients show the effect



Banchereau et al (2016) *Cell* **165**: 551-565 "Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients"

Patients show individualized patterns of correlation between modules and disease

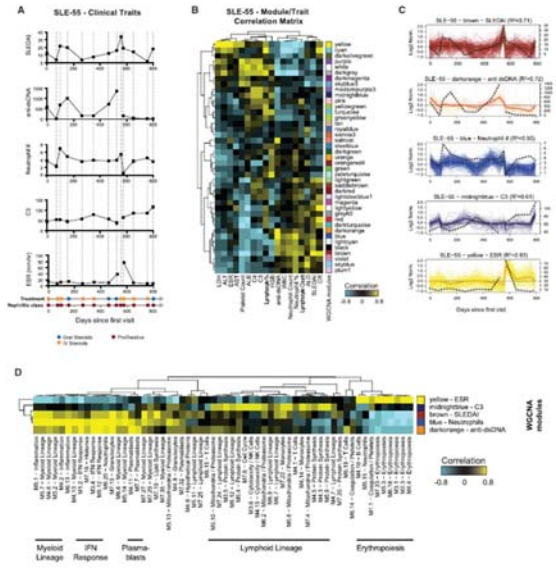
A. Trajectory of SLEDAI (disease activity) and 4 other clinical measures over 2 years

B. Correlation between18 disease measures and 40 WGCNA modules

C. Trajectories or module genes with traits

D. WGCNA modules map onto the 95 blood modules, which correlate with the 5 traits

E. There are ~8 classes of patients based on the SLEDAI correlations, impling heterogeneity of disease pathology

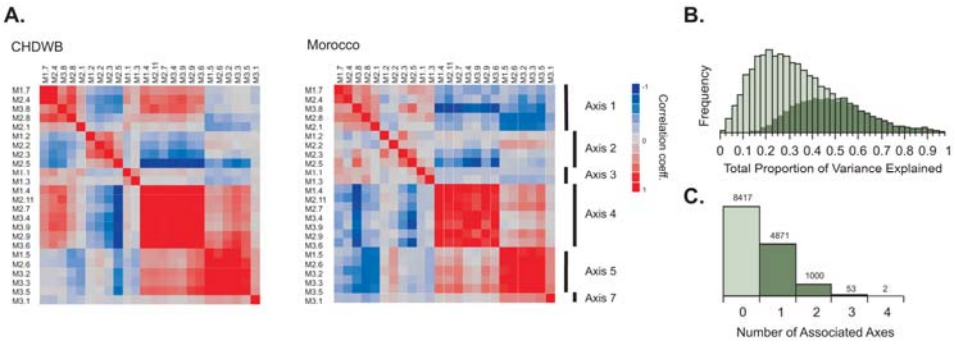


Banchereau et al (2016) *Cell* **165**: 551-565 "Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients"

Axes of Variation and Blood Informative Transcripts

Comparing across multiple cohort studies of healthy adults, the initial 28 modules reduce to 7 Axes

We have since identified 1 additional Axis, and suspect there are a few more in some datasets

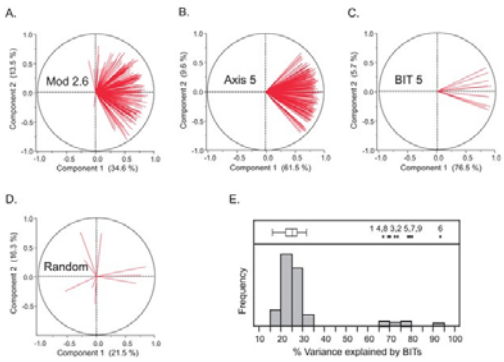


Preininger et al (2013) *PLoS Genet.* **9**: e1003362 "Blood Informative Transcripts define nine common axes of peripheral blood gene expression"

Eight Axes capture basic Immune functions

We define the Axis scores at the first Principal Component of the positively correlated genes

Each Axis corresponds to an identified aspect of immune function, but they explain much more of the variance than the corresponding cell counts. The covariance is due to both cell abundance and transcription.

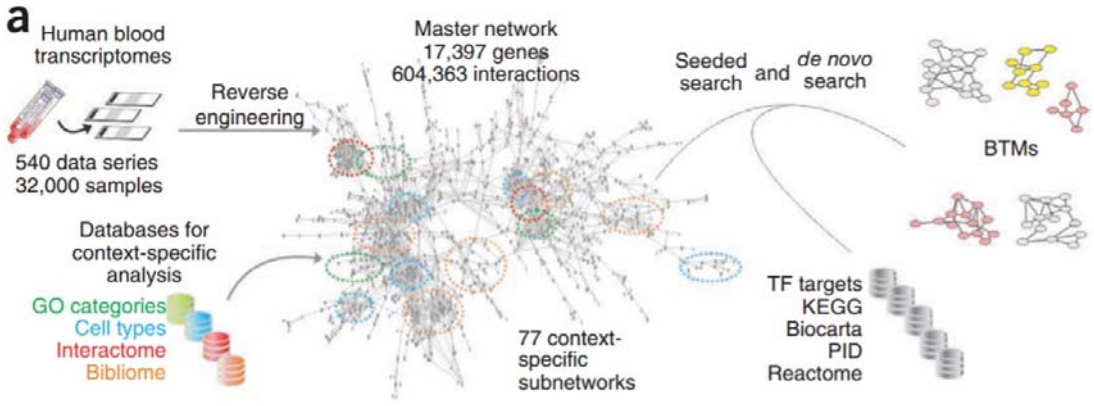


Axis 1	T-cell signaling
Axis 2	Reticulocyte development
Axis 3	B-cell signaling
Axis 4	Housekeeping functions
Axis 5	Neutrophils and TLR signaling
Axis 6	Antibody response ?
Axis 7	Interferon response
Axis 10	Mitosis / cell cycle

BIT are 10 Blood Informative Transcripts that define each Axis.

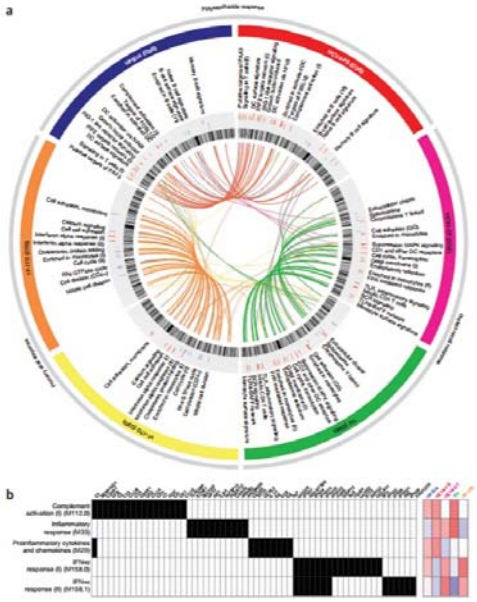
Preininger et al (2013) *PLoS Genet.* **9**: e1003362 "Blood Informative Transcripts define nine common axes of peripheral blood gene expression"

Identifying Blood Transcript Modules



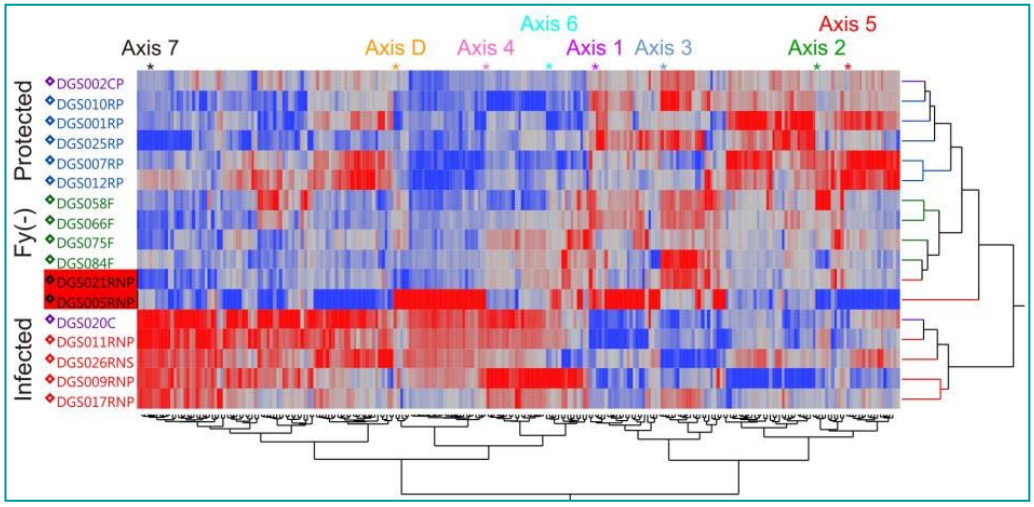
Li et al (2014) *Nat Immunology* **15**: 195-204 "Molecular signatures of antibody responses derived from a systems biological study of 5 human vaccines"

Blood Transcript Modules identify common vaccine response mechanisms



Li et al (2014) *Nat Immunology* **15**: 195-204 "Molecular signatures of antibody responses derived from a systems biological study of 5 human vaccines"

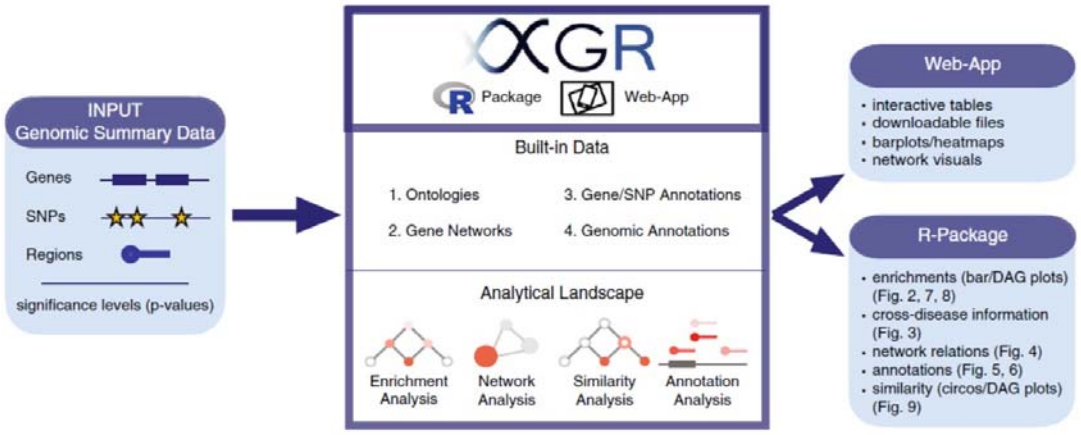
BTM meet BIT



Each of the 8 Axes embed with the 250 BTM, and help explain a malaria vaccine response

Monica Rojas-Peña, Socrates and Myriam Herrera

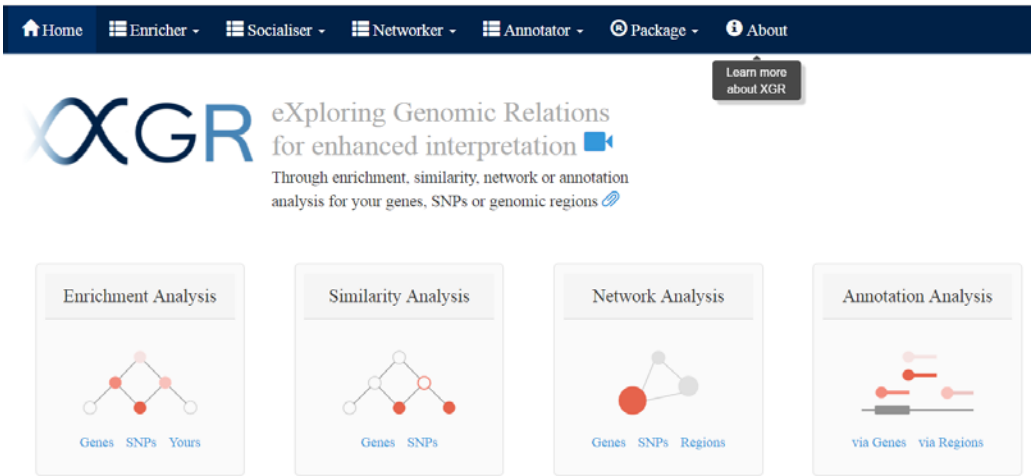
XGR Software: eXploring Genomic Relations



Fang et al (2016) *Genom Med.* **8**: 129 "XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits"

XGR User-friendly Website

http://galahad.well.ox.ac.uk/XGR



A sample of what XGR Software can do