Gene co-expression networks in the mouse, monkey, and human brain
July 16, 2013

Jeremy Miller
Scientist I
jeremym@alleninstitute.org
Outline

1. Brief introduction to previous WGCNA studies in brain

2. Brief introduction to Allen Institute resources

3. Co-expression networks in the adult human brain

4. Identifying signatures of neurogenesis in the hippocampal subgranular zone in rodents and primates

5. Laminar and areal specification of the developing human neocortex
Previous results using WGCNA

1. Identification of a novel regulator of glioblastoma
   • ASPM was a hub in a cell cycle module with several known cancer genes

2. Gene co-expression differences between human and chimp
   • Weaker module conservation in cortex than subcortex networks
   • Differential connectivity related to protein sequence divergence

3. Identification of conserved modules in Alzheimer's and normal aging
   • Decreased expression in energy metabolism & synaptic plasticity modules
   • Potential role for oligodendrocyte dysfunction via PSEN1

4. Characterization of the human brain transcriptome
   • WGCNA can identify cell type markers from heterogeneous tissue
   • Modules for neurons, oligodendrocytes, astrocytes, microglia, and a class of cells in the subventricular zone neurogenic niche

All references available at: labs.genetics.ucla.edu/horvath/CoexpressionNetwork/
Things to consider when running WGCNA (1)

1. Which probes should be used in the analysis?
   – Typically I use one probe for each gene on the array \textit{(collapseRows)}
   – Genes can also be filtered beforehand, depending on analysis goals

2. Which samples should be used in the analysis?
   – This depends on the question being asked

3. How do we choose the parameters?
   – I find that linear space typically works better than log2
   – WGCNA robust to parameter choices--often the defaults are fine.

4. How should modules be selected from the dendrogram?
   – Many small modules = high confidence of co-expression relationships
   – Few large modules = better ability to annotate modules
   – With few samples, it is often better to use few modules
Things to consider when running WGCNA (2)

5. What can the modules tell us about biology?
   – What is known about the genes?
     • GO, IPA, userListEnrichment, etc.
   – How do the patterns (module eigengene) relate to biology?
     • Correlation with phenotype, etc.

6. Which parts of the network are preserved in other data sets?
   – Are expression patterns of hub genes changed between conditions?
   – Preservation can be summarized using modulePreservation
   – Differences can be identified using differential connectivity

7. Visualizations are important!
   – Displaying modules and networks sensibly can drastically improve your ability to understand the biology.
     • Network depictions (VisANT), WGCNA plotting functions, custom plots, etc.
Allen Brain Atlas Data portal – brain-map.org

Mouse data often used to compare or contrast with primate

- >3000 arrays in 6 adult brains – focused on breadth of coverage
- >1000 arrays in 4 prenatal brains (15-21pcw) – focused on cortical layers
- >500 arrays in five brain regions spanning macaque development

Several other resources that I won’t discuss
Experimental Set-Up: One Brain, Many Samples

Given a lot of samples (500-1000) across (initially) one human brain, what kinds of biological questions can we ask...

1. Are certain genes specific to specific parts of the brain?
2. Which genes show similar expression patterns across the brain?
3. Which brain regions show similar gene signatures?
4. How are known markers for cell types distributed across the brain?
5. Do genes show different patterns at a global scale (whole brain) and a local scale (within specific brain areas)?

http://human.brain-map.org/explorer.html
Experimental Set-Up: Two Brains, Many Samples

Given a second brain with comparable brain regions assayed, how consistent are gene expression patterns across brains?

(The analysis of all six brains is in progress)

WGCNA can address many of these questions!

http://human.brain-map.org/explorer.html
The whole brain: Global Analysis using WGCNA

- Network created using 911 samples for one brain
- Genes cluster into 13 distinct modules
- A few modules are for known cell types
- Good agreement with second brain
- What can we learn about these modules?
The whole brain: Global Analysis using WGCNA

- Brain region enrichment found using gene expression (bar graphs)
- Gene ontology enrichment found using DAVID / EASE
- Cell type enrichment found by comparing with known markers (userListEnrichment)
- Hub genes can confirm module characterization and suggest novel genes associated with biology

- Note dramatic differences in expression patterns of neurons and glia!
Global Analysis – What we did not learn

- **WGCNA tends to find the most prevalent patterns in the data, so to find local marker genes, look only at the relevant subset of samples.**

- We do **not** find modules associated with smaller brain areas (i.e., dentate gyrus, individual midbrain nuclei, etc.), even though we know these markers exist.

- Can we find different clusters if we use a small set of related samples?

Example:
Local Analysis – Hippocampus only (66 arrays)

- Modules do still tend to have good corroboration between brains
- But modules do not seem to group by cell type as much
  - So what do they represent?
• Most modules in this analysis represent *different patterns of expression within hippocampus, both within and between subregions*

• Similar results were found for other brain regions…
Section Summary

- We have used WGCNA to address the following features of the adult human brain transcriptome:
  1. Across the whole brain, genes group based on broad cell types, basic cellular functions, and distinct brain regions
  2. Across hippocampus genes enriched in compartments and/or with rostral to caudal patterning group together
  3. Neocortical regions tend to be enriched for neuronal markers, while certain subcortical regions are enriched for glial markers.
  4. Do genes show different patterns at a global scale (whole brain) and a local scale (within specific brain areas)?
    - While co-expression on both scales tends to be similar, there are enough differences that it is worth doing the analysis at a global and local scale.

- This type of analysis should be effective for any large-scale project (i.e., cancer databases, comparisons between cell lines, etc.)
Allen Brain Atlas Data portal – brain-map.org

Mouse data used to compare & contrast with primate

>500 arrays in five brain regions spanning macaque development
Introduction

• Neurogenesis was originally thought to occur only during early development—which is true for most of the brain.

• In at least two locations neurogenesis continues throughout life:
  
  – **Subventricular zone (SVZ)** - cells generated in lateral ventricle wall migrate to olfactory bulb and differentiate into interneurons
  
  – **Subgranular zone (SGZ) of the hippocampus** – cells generated here become dentate granule cells.

• These new neurons make functional connections.

• The environment (“neurogenic niche”) is critical:
  
  – SGZ/SVZ precursors transplanted elsewhere show limited neurogenesis
  
  – Neural stem cells transplanted to SGZ/SVZ develop into appropriate neurons
SGZ niche is complex!

While much of the process and some of the players are known, our understanding of this process is far from complete.
Overview of the study

1. SGZ vs. GCL in mouse

2. PubMed AGEA NeuroBlast (etc.) to find more genes


4. Characterize SGZ genes in macaque using developmental time course and WGCNA

5. Experimental validation of neurogenesis genes
Robust SGZ-enriched genes identified in mouse

- 367 SGZ-enriched genes were found
- A large subset agree between ISH and microarray

- These genes mark many distinct cell types in this very small band!
  - Some genes assigned to cell types based on anatomy
  - Canonical markers for neurogenesis and several cell types also found using enrichment analysis.
Distinct cell types based on gene expression

- Astrocytes
  - Aldoc
  - Cst3

- Dividing Cells
  - Mki67
  - Top2a

- GABAergic interneurons
  - Gad2
  - Erbb4

- Oligodendrocytes
  - alveus
  - Mbpa
  - Mog
  - fl

- Vascular endothelial and blood cells
  - Pltp
  - Clcn5
Other highly SGZ-enriched genes

Many genes did not obviously mark a specific cell type, but...
Other highly SGZ-enriched genes

... many of these already had known roles in neurogenesis.

Can we identify (more) cross-species markers for neurogenesis using a primate model?
Anatomy and gene expression in monkey dentate gyrus

- Massive decrease in SGZ with time (almost gone by 48 mo.)
- Extensive polymorphic layer compared with mouse
- Transcription (MDS) consisitant with anatomy
- Do we find common signatures in monkey and mouse?
Analysis strategy in non-human primate

**Strategy**
1. Cluster genes using WGCNA on all SGZ and GCL samples
2. Identify modules with SGZ > GCL
3. Characterize modules using enrichment analysis
Macaque SGZ genes differ primarily at T=0

- Four modules show SGZ enrichment.
- These modules contain most of the mouse SGZ genes.
- Glia module up with time, neurogenesis down with time.
- Intermediate modules likely interneurons, radial glia, etc.

Focus on neurogenesis module.
Tan module expression correlates w/ proliferating cells

- The module eigengene for the tan module almost perfectly correlates with number of proliferating cells in macaque DG.
- SOX11 and SOX4 (hubs) are canonical markers for neurogenesis—required for neuronal differentiation.
- **What happens if we knock them out?**

Quantitative analysis of postnatal neurogenesis and neuron number in the macaque monkey dentate gyrus

Adeline Jobin,1 Pamela Banta-Lavereen,3 David G. Amaral3 and Pierre Lavereen1

1Laboratory of Brain and Cognitive Development, Department of Medicine, University of Fribourg, Switzerland
2Department of Psychiatry and Behavioral Sciences, Center for Neuroscience, California National Primate Research Center, The M.I.N.D. Institute, UC Davis, USA
Tan module expression correlates w/ proliferating cells

- Conditional knockout of SOX4 and SOX11 in cortex (and hippocampus) produces a mouse with no hippocampus
  - *These two genes are critical for hippocampal neurogenesis*
  - Neither single knockout has an obvious phenotype suggesting these genes have redundant function
- ISH for 46 genes were run in the hippocampus as part of the NHP Atlas.
- Two of these were in the tan module and showed the expected SGZ enrichment and decrease in expression with time.
Several genes show expected temporal pattern in mouse

- We next used the Allen Developing Mouse Brain Atlas to assess gene expression of these tan module genes in mouse.
- At least six genes had expression patterns enriched in SGZ and decreasing with time.

Interestingly, P14 in mouse seems to match with birth in macaque.
Section Summary

• The SGZ neurogenic niche contains a complex combination of cell types including radial glia / progenitors, dividing cells, immature neurons, astrocytes, vasculature, & interneurons

• There is a high level of transcriptional similarity in the SGZ of mouse and macaque

• The makeup of the neurogenic niche changes with development
  – In particular, expression of a group of genes is highly correlated with the number of proliferating cells
  – Two hub genes in this module (SOX4 and SOX11) have a functional role in hippocampal neurogenesis.
Mouse data used to compare & contrast with primate

>1000 arrays in 4 prenatal brains (15-21pcw) – focused on cortical layers
BrainSpan – Prenatal LMD Microarray

4 brains
(15, 16, 21, 21 pcw)

~25 neocortical regions / brain

9 layers / region

~500 total arrays in analysis
Transient layers during early prenatal development

Inside-out generation of neurons destined for successive cortical layers
What makes the developing human(/primate) neocortex unique?

1) Large secondary neurogenic zone, the outer subventricular zone

2) Large transient subplate zone, which is generated over a long period of time

3) Potentially some local generation of GABAergic interneurons, although finding is controversial
Samples cluster by layer and region

- Unbiased clustering of samples (MDS using all genes) groups samples by layer and location in neocortex.
- Layers with primarily dividing (germinal) cells separate from layers with postmitotic cells (i.e., neurons).
  - Do we see these patterns using WGCNA?
  - Do we find other patterns using WGCNA?
WGCNA identifies distinct cell populations

Modules potentially relevant in primate-specific development:

- **Germinal cells**: can we distinguish different types?
- **Cortical neurons**: markers for Autism in these layers?
- **Subplate neurons**: are there different markers in mouse brain?
- **Interneurons**: primate-specific expression in VZ?
Gene expression differences in human and mouse subplate

Differences in gene expression may help to explain expansion of subplate in primate compared with mouse.

WGCNA module used as starting point for targeted search of Allen Developing Mouse Brain Atlas
Focused network analysis distinguished different progenitor cell types

Radial glia enriched in VZ

Intermediate progenitors in SZi

Most SZo modules enriched in outer layers (neurons passing through?)
A small but significant number of genes with areal patterning were not identified by WGCNA.

- A few genes were enriched in rostral (front) or caudal (back) of cortex.
- Most of these genes are specific to one or two layers, and not necessarily the layers with the highest expression.
  - There are real results that we can confirm in mouse brain.
  - Rostral genes may underlie the expansion of frontal cortex in human.

Directed analyses are useful if you want to answer a specific question!
Section Summary

• Predominant gene expression variation due to age and layer
  • In particular, postmitotic vs. germinal layers

• Genes cluster based on expression in layers and major cell classes

• We learn valuable information by focusing on specific layers
  • Several distinct transcriptional signatures in each germinal layer
  • Human and mouse subplate have some transcriptional differences

• Several genes show layer-specific expression gradients
  • Too few genes/samples involved for robust identification using WGCNA, so more directed methods are also useful.
Overall Summary and Concluding Remarks

• WGCNA in a useful method for identifying patterns of co-expressed genes and for reducing the dimensionality of the data.

• We have applied this method for several Allen Institute atlas projects:
  - Co-expression networks in adult human brain
  - Molecular signatures of hippocampal subgranular zone
  - Laminar and cell type signatures in prenatal human brain

• Be aware of the scope of the data set when using WGCNA
  - Signatures of local phenomena can by masked by larger signatures (such as cell types, large regional differences, etc.)
  - Large data sets with many replicates (i.e. 100 samples from a single region) can produce networks very different from large data sets with many unique samples (i.e., one sample from each of 100 regions)
Acknowledgements

We wish to thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support.

Allen Institute
Mike Hawrylycz
Ed Lein
CK Lee
Vilas Menon
Susan Sunkin
Elaine Shen

UCLA
Steve Horvath
Peter Langfelder
Dan Geschwind
Mike Oldham

Any questions?