“Genomics and Disease”

Systems Biology & Genomics Lab
Harry Perkins Institute of Medical Research, UWA, Australia
&
RIKEN Center for Life Science Technologies,
Division of Genomic Technologies, Yokohama, Japan

Alistair Forrest

FMDHS Research Week 2015, Perth
Genomics and disease

• Systems biology and networks

• Cell type specificity and disease

• Genetic diseases as diseases of cell types (or broken networks)

• Beyond exomes (SNPs in regulatory regions)

• Cell-to-cell communication networks (ligand-receptor mediated)
Traditional biology looks at one part in detail
Systems biology looks at all the parts.

How do they work together?
We are complex multicellular organisms composed of at least 400 distinct cell types, that have all arisen through sequential specialization from a single totipotent cell.

Acknowledgements: Images from Kyoto Human Embryo Visualization Project
Cell specialization

A: Hippocampal neuron
B: Alveolar macrophage
C: Pulmonary Artery Endothelial cells
D: Pancreatic islet
E: Inner ear hair cells
F: Rods and cones

Acknowledgements: D Douglas Melton, Harvard Stem Cell Institute, E Donja Pyott, University of North Carolina, A, B, F (pending, google images) or C (wiki commons)
Cell specific regulatory networks

- Shared core components

CD4+ T cell  CD8+ T cell  melanocyte
Cell specific regulatory networks

- Shared core components
- Common lineage components

CD4+ T cell

CD8+ T cell

melanocyte
Cell specific regulatory networks

- Shared core components
- Common lineage components
- Cell specific components

CD4+ T cell
CD8+ T cell
melanocyte
Cell specific regulatory networks

- Shared core components
- Common lineage components
- Cell specific components

FANTOM5 focuses on identifying these

CD4+ T cell
CD8+ T cell
melanocyte
Cell specific regulatory networks

- Shared core components
- Common lineage components
- Cell specific components

Where are the disease associated genes?

CD4+ T cell  CD8+ T cell  melanocyte
Single Molecule Cap Analysis Gene Expression (CAGE) on Helicos

- CAGE captures the 5' end of the mRNA as a short cDNA tag

A promoter-level mammalian expression atlas

The FANTOM Consortium and the RIKEN PMI and CLST (DGT)*

Regulated transcription controls the diversity, developmental pathways and spatial organization of the hundreds of cell types that make up a mammal. Using single-molecule cDNA sequencing, we mapped transcription start sites (TSSs) and their usage in human and mouse primary cells, cell lines and tissues to produce a comprehensive overview of mammalian gene expression across the human body. We find that few genes are truly ‘housekeeping’, whereas many mammalian promoters are composite entities composed of several closely separated TSSs, with independent cell-type-specific expression profiles. TSSs specific to different cell types evolve at different rates, whereas promoters of broadly expressed genes are the most conserved. Promoter-based expression analysis reveals key transcription factors defining cell states and links them to binding-site motifs. The functions of identified novel transcripts can be predicted by coexpression and sample ontology enrichment analyses. The functional annotation of the mammalian genome 5 (FANTOM5) project provides comprehensive expression profiles and functional annotation of mammalian cell-type-specific transcriptomes with wide applications in biomedical research.

http://fantom.gsc.riken.jp/5/
Cap Analysis of Gene Expression on Helicos single molecule sequencer – median 4 million uniquely mapping tags per sample (>=Q20)

All samples have been annotated using cell ontology (CL), disease ontology (DOID) and the uberon anatomical ontology (UBERON) – (Ack. to Terry Meehan, Chris Mungall)
Visualizing CAGE data – expression and genomic localization
ZENBU (全部) data integration and visualization system


http://fantom.gsc.riken.jp/zenbu/
Even for this canonical ‘housekeeper’ gene, there is variation in expression levels.
Distribution of maximum and median TPM values for 184,827 peaks identified from 988 human samples
Region 1 genes

House keeping genes \([\text{max} \sim \text{median}]\)

Gene Ontology enrichment: (Gostat)
- cell part \(n=3810/11717 \ p=2.83\times10^{-11}\)
- organelle part \(n=1447/2935 \ p=0\)
- macromolecule metabolic process \(n=2289/5669 \ p=2.84\times10^{-7}\)
- nucleic acid binding \(n=1252/2812 \ p=1.48\times10^{-62}\)
- protein complex \(n=729/1613 \ p=5.09\times10^{-36}\)
- cellular metabolic process \(n=2523/6483 \ p=1.2\times10^{-70}\)

Examples: mRNA transport, ligase, helicase and transferase activities, mitochondrion, ribosome, ACTB, GAPDH
Region 2 genes

Tissue/cell type specific genes [detected in <50% of samples]

Gene Ontology enrichment: (Gostat)
- plasma membrane (n=1477/2357 p= 0)
- cellular metabolic process (n=1972/6482 p=0)
- organelle part (n=664/2935 p=0)
- primary metabolic process (n=1994/6522 p=0)
- multicellular organismal process (n=1712/2820 p=0)
- receptor activity (n=799/1132 p=0)
- integral to membrane (n=2131/3862 p=0)
- extracellular part (n=524/705 p=6.76e-75)

Examples: GPCRs, solute channels, transcription factors, serine proteases, rhodoposins, extracellular matrix and muscle proteins,
Region 3 genes

Hybrid housekeeper/tissue specific genes [max > median]

Gene Ontology enrichment: (Gostat)
- regulation of cellular process (n=1246/3542 p=7.05e-23)
- intracellular (n=2720/8690 p=1.28e-21)
- cell cycle (n=285/632 p=6.11e-19)
- signal transduction (n=872/2790 p=2.71e-05)

Examples: protein kinases, GTPases, cell cycle regulators, Chromatin proteins

*Tendency for promoters of these genes to have broader more complex CAGE peak patterns*
Relationship between tissue/cell-type specific expression and observed phenotype.

1. Are mutations in genes that have restricted expression in specific tissues/cell-types good candidates for causative genes for diseases of those tissues?

2. For genes associated with known tissue specific phenotypes where are they expressed?

3. Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?
Relationship between tissue/cell-type specific expression and observed phenotype.

1. Are mutations in genes that have restricted expression in specific tissues/cell-types good candidates for causative genes for diseases of those tissues?

2. For genes associated with known tissue specific phenotypes where are they expressed?

3. Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?
OMIM: Alexander disease is a rare disorder of the central nervous system. Infants with Alexander disease develop a leukoencephalopathy with macrocephaly, seizures, and psychomotor retardation, leading to death usually within the first decade.
**Transcription factors**

Interestingly we detect expression of the majority of human transcription factors in the atlas (1679; 96%) at levels (median of maximal level 78TPM) similar to other coding genes.
Cell type enriched transcription factors

- Common lineage components
- Cell specific components

Enrichment = \( \log_{10}(\text{expr}(\text{cellX}) + 1) - \log_{10}(\text{expr}(\text{median}) + 1) \)
MGI phenotype report: **IRF8** - Homozygotes for a targeted null mutation exhibit increased incidence of viral infections, shortened life span, deregulated hematopoiesis, and hematological neoplasias. Heterozygotes show similar, but milder, phenotypes.

MGI phenotype report: **MYB** - Mice homozygous for deficient alleles of this gene display severe hematopoietic abnormalities. Red and white blood cells and platelets are all affected.

MGI phenotype report: **IKZF1** - Homozygous mutants have a variety of T, B, and hematopoietic cell maturation defects. Heterozygotes for one allele exhibit dominant negative effects and mice develop lymphoproliferative disorders.

MGI phenotype report: **NFE2** - Homozygotes for a targeted null mutation lack platelets and most die as neonates from internal bleeding. Survivors exhibit hypochromia, reticulocytosis, and splenomegaly.

http://fantom.gsc.riken.jp/5/sstar/
**MGI phenotype report: Creb3l1**
Mice homozygous for a knock-out allele exhibit postnatal growth retardation, fragile skeleton, and decreased bone density, cortical and trabecular thickness, and osteoblast maturation.

**MGI phenotype report: Prrx1**
Homozygotes for targeted null mutations exhibit skeletal defects affecting mandible, limbs, and vertebrae, vascular abnormalities, and neonatal lethality.

**MGI phenotype report: Six1**
Homozygotes for targeted null mutations exhibit rib defects, muscle hypoplasia, absence of kidneys and thymus, craniofacial defects, and ear and nose defects. Heterozygotes show partial hearing loss.

The short stature homeobox gene **SHOX** is involved in skeletal abnormalities in Turner syndrome

Mark Clement-Jones, Simone Schiller¹, Ercole Rao¹, Rüdiger J. Blaschke¹, Aimee Zuniga², Rolf Zeller², Stephen C. Robson³, Gerhard Binder⁴, Ian Glass⁵, Tom Strachan, Susan Lindsay+ and Gudrun A. Rappold¹++

http://fantom.gsc.riken.jp/5/sstar/
**Inner ear hair cells**

### Species
- Mouse (Mus musculus)

### Genomic View
- ensembl, UCSC

---

**Enrichment Score**

<table>
<thead>
<tr>
<th>Enrichment Score</th>
<th>TPM</th>
<th>TF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.39</td>
<td>245.35</td>
<td>Pou3f4</td>
</tr>
<tr>
<td>2.22</td>
<td>2672.44</td>
<td>Fosb</td>
</tr>
<tr>
<td>2.15</td>
<td>190.10</td>
<td>Fosb</td>
</tr>
<tr>
<td>2.05</td>
<td>153.26</td>
<td>Fosb</td>
</tr>
<tr>
<td>1.97</td>
<td>198.26</td>
<td>Tbx19</td>
</tr>
<tr>
<td>1.93</td>
<td>64.08</td>
<td>Pou3f4</td>
</tr>
<tr>
<td>1.92</td>
<td>172.90</td>
<td>Six1</td>
</tr>
<tr>
<td>1.91</td>
<td>71.27</td>
<td>Pou4f3</td>
</tr>
<tr>
<td>1.81</td>
<td>78.79</td>
<td>Pou3f3</td>
</tr>
<tr>
<td>1.59</td>
<td>47.88</td>
<td>Zic2</td>
</tr>
<tr>
<td>1.56</td>
<td>1604.37</td>
<td>Egr1</td>
</tr>
<tr>
<td>1.55</td>
<td>66.32</td>
<td>Fosf2</td>
</tr>
<tr>
<td>1.54</td>
<td>55.29</td>
<td>Egr2</td>
</tr>
</tbody>
</table>

---

**Association Between X-Linked Mixed Deafness and Mutations in the POU Domain Gene POU4F4**

Yvette J. M. de Kok,* Silvère M. van der Maarel,* Maria Bitner-Glindzicz, Irene Huber, Anthony P. Monaco, Susan Malcolm, Marcus E. Pembrey, Hans-Hilger Ropers, Frans P. M. Cremerst†

Deafness with fixation of the stapes (DFN3) is the most frequent X-linked form of hearing impairment. Here, it is reported that a candidate gene for this disorder, Brain 4 (POU4F4), which encodes a transcription factor with a POU domain, maps to the same interval. In five unrelated patients with DFN3 but not in 50 normal controls, small mutations were found that result in truncation of the predicted protein or in nonconservative amino acid substitutions. These findings indicate that POU4F4 mutations are a molecular cause of DFN3.

---

**Mutation in Transcription Factor POU4F3 Associated with Inherited Progressive Hearing Loss in Humans**


The molecular basis for autosomal dominant progressive nonsyndromic hearing loss in an Israeli Jewish family, Family H, has been determined. Linkage analysis placed this deafness locus, DFN15, on chromosome 5q31. The human homolog of mouse Pou4f3, a member of the POU-domain family of transcription factors whose targeted inactivation causes profound deafness in mice, was physically mapped to the 25-centimorgan DFN15-linked region. An 8-base pair deletion in the POU homeodomain of human POU4F3 was identified in Family H. A truncated protein presumably impairs high-affinity binding of this transcription factor in a dominant negative fashion, leading to progressive hearing loss.

---

**The role of Six1 in mammalian auditory system development**

Weiming Zheng, Li Huang, Zhu-Bo Wei, Derek Silvius, Bihui Tang andPin-Xian Xu*

McLaughlin Research Institute for Biomedical Sciences, 1520 23rd Street South, Great Falls, MT 59005, USA

Development 139, 1309-4000 © 2003 The Company of Biologists Ltd

doi: 10.1242/dev.1391149031

---

**Deafness in mice lacking the T-box transcription factor Tbx18 in otic fibrocytes**

Mark-Oliver Trowe†, Hannes Maier‡, Michaela Schweizer‡ and Andreas Kissert‡*

In the cochlea, fibrocytes play important physiological roles, including the maintenance of the ionic composition of the endolymph. Human deafness upon fibrocyte alterations witnesses their crucial role for hearing. We demonstrate that differentiation of otic fibrocytes requires the T-box transcription factor gene Tbx18. Tbx18 expression during inner ear development is restricted to the sub-region of otic mesenchyme that is fated to differentiate into fibrocytes. We rescued the somitic defect that underlies the perinatal lethality of Tbx18 mutant mice by a transgenic approach, and measured auditory brainstem responses. Adult Tbx18 deficient mice showed profound deafness and a complete disruption of the endocochlear potential that is essential for the transduction of sound by sensory hair cells. The differentiation of otic fibrocytes of the spiral ligament was severely compromised. Tissue architecture of the stria vasculosa of the lateral wall was disrupted, exhibiting an almost complete absence of the basal cell layer, and a reduction of intermediate and marginal cells, respectively. Stria vasculosa defects resulted from the failure of Tbx18 mutant otic fibrocytes to generate the basal cell layer by a mesenchymal-epithelial transition. Defects in otic fibrocyte differentiation may be subordinate to a primary role of Tbx18 in early compartmentalization of the otic mesenchyme, as lineage restriction and boundary formation between otic fibrocytes and the surrounding otic capsule were severely affected in the mutant. We show that the genetic control of patterning and differentiation of the otic mesenchyme, uncovered distinct steps of stria vasculosa formation and illuminates the importance of non-epithelially-derived otic cell types for normal hearing and the etiology of deafness.

---

**Inner ear cells**: Albert Edge & Judith Kempfle

Harvard

http://fantom.gsc.riken.jp/5/sstar/
Relationship between tissue/cell-type specific expression and observed phenotype.

1. Are mutations in genes that have restricted expression in specific tissues/cell-types good candidates for causative genes for diseases of those tissues?

2. For genes associated with known tissue specific phenotypes where are they expressed?

3. Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?
Relationship between tissue/cell-type specific expression and observed phenotype.

1. Are mutations in genes that have restricted expression in specific tissues/cell-types good candidates for causative genes for diseases of those tissues?

2. For genes associated with known tissue specific phenotypes where are they expressed?

3. Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?
OMIM: Abnormal expansion of a polyglutamine tract in the N terminus of huntingtin causes Huntington disease, a devastating autosomal dominant neurodegenerative disease characterized by motor, psychiatric, and cognitive dysfunction (summary by Futter et al., 2009).
OMIM: Rett syndrome is a progressive neurologic developmental disorder and one of the most common causes of mental retardation in females. Because RTT occurs almost exclusively in females, it had been proposed that RTT is caused by an X-linked dominant mutation with lethality in hemizygous males.
Human BRCA1 – breast and ovarian cancer

UBERON:0000310

chr17:41277372..41277418,−

Number of samples

breast
other

0 889

Human BRCA1 – breast and ovarian cancer 889
Parkinson’s disease genes

- **PARK2 – Parkinson’s disease**
- **PINK1 – Parkinson’s disease**
- **LRRK2 – Parkinson’s disease**
- **SNCA – Parkinson’s disease**

UBERON:0001016

chr6:163148780..163148813,−
nervous system other

Number of samples p1@PARK2

chr6:163148780..163148813,−
0 5 10 15 20 25

UBE:0001016

chr1:8021713..8021788,+ nervous system other

Number of samples p1@PARK7

chr1:20959943..20959992,+ nervous system other

Number of samples p1@PINK1

chr4:90758258..90758372,− nervous system other

Number of samples p1@SNCA

chr12:40618764..40618841,+ nervous system other

Number of samples p1@LRRK2

chr6:163148780..163148813,− nervous system other

Number of samples p1@PARK2

chr1:8021713..8021788,+ nervous system other

Number of samples p1@PARK7
Relationship between tissue/cell-type specific expression and observed phenotype.

1. Are mutations in genes that have restricted expression in specific tissues/cell-types good candidates for causative genes for diseases of those tissues?

2. For genes associated with known tissue specific phenotypes where are they expressed?

3. Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?
Relationship between tissue/cell-type specific expression and observed phenotype.

1. Are mutations in genes that have restricted expression in specific tissues/cell-types good candidates for causative genes for diseases of those tissues?

2. For genes associated with known tissue specific phenotypes where are they expressed?

3. Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?
Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?

Susceptible vs Resistant tissues?

Undiagnosed multi-system diseases
Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?

Susceptible vs Resistant tissues?

• Tissue specific/cell type specific interactions (think synthetic lethal or cell type specific modifiers) e.g. MECP2 and mCA

Undiagnosed multi-system diseases
Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?

Susceptible vs Resistant tissues?

- Tissue specific/cell type specific interactions (think synthetic lethal or cell type specific modifiers) e.g. MECP2 and mCA

- Tissue specific environmental modifiers (e.g. UV exposure and skin cancer, drug use and neurological disorders)

Undiagnosed multi-system diseases
Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?

Susceptible vs Resistant tissues?

• Tissue specific/cell type specific interactions (think synthetic lethal or cell type specific modifiers) e.g. MECP2 and mCA

• Tissue specific environmental modifiers (e.g. UV exposure and skin cancer, drug use and neurological disorders)

Undiagnosed multi-system diseases

• Disease presentation = most obvious phenotype
Why do mutations in some broadly expressed genes result in tissue specific phenotypes? Are they actually multi-system diseases?

Susceptible vs Resistant tissues?

• Tissue specific/cell type specific interactions (think synthetic lethal or cell type specific modifiers) e.g. MECP2 and mCA

• Tissue specific environmental modifiers (e.g. UV exposure and skin cancer, drug use and neurological disorders)

Undiagnosed multi-system diseases

• Disease presentation = most obvious phenotype

• Disease presentation = first to manifest
Relationship between tissue/cell-type specific expression and observed phenotype.

<table>
<thead>
<tr>
<th>Disease presentation</th>
<th>Specific</th>
<th>Broad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>Cell autonomous: GFAP, TAU, SNCA</td>
<td>HTT, PARK7, BRCA1, MECP2</td>
</tr>
<tr>
<td></td>
<td>Cell non-autonomous: LRRK2??</td>
<td></td>
</tr>
<tr>
<td>Broad</td>
<td>INS, GH1</td>
<td>TP53</td>
</tr>
</tbody>
</table>
Beyond exomes:
• Only 4.9% of GWAS SNPs lie in coding sequence! (Maurano et al. Science 2012)

Regulatory regions:
• Promoters
• Enhancers!!!
What are enhancers?

1. Activator proteins bind to enhancer, triggering DNA bending.

2. Activators interact with coactivators to stimulate chromatin remodeling and histone acetylation.

3. Activators bind to Mediator, triggering assembly of RNA polymerase and general transcription factors at the promoter site.
An atlas of active enhancers across human cell types and tissues

Robin Andersson1*, Claudia Gebhard2,3*, Irene Miguel-Escalada4, Ilka Hoof4, Jette Bornholdt4, Mette Boyd1, Yun Chen1, Xiaobei Zhao1,5, Christian Schmid6, Takahiro Suzuki6,7, Evgenia Ntini8, Erik Arner6,7, Eivind Valen1,9, Kang Li1, Lucia Schwarzsicher2, Dagmar Glatz2, Johanna Raithel2, Berit Lilje1, Nicolas Rapin1,10, Frederik Otzen Bagger1,10, Mette Jørgensen1, Peter Reising Andersen8, Nicolas Bertin6,7, Owen Rackham6,7, A. Maxwell Burroughs6,7, Kenneth Baillie11, Yuri Ishizu6,7, Yuri Shimizu6,7, Erina Furuhata6,7, Shiuri Maeda6,7, Yutaka Negishi6,7, Christopher J. Mungall12, Terrence F. Meehan13, Timo Lassmann6,7, Masayoshi Itoh6,7,14, Hideya Kawaji6,14, Naoto Kondo6,14, Jun Kawai6,14, Andreas Lennartsson15, Carsten O. Daub6,7,15, Peter Heutink16, David A. Hume11, Torben Heick Jensen8, Harukazu Suzuki6,7, Yoshihide Hayashizaki6,14, Ferenc Müller4, The FANTOM Consortium†, Alistair R. R. Forrest6,7, Piero Carninci6,7, Michael Rehli2,3 & Albin Sandelin1

Enhancers control the correct temporal and cell-type-specific activation of gene expression in multicellular eukaryotes. Knowing their properties, regulatory activity and targets is crucial to understand the regulation of differentiation and homeostasis. Here we use the FANTOM5 panel of samples, covering the majority of human tissues and cell types, to produce an atlas of active, in vivo-transcribed enhancers. We show that enhancers share properties with CpG-poor messenger RNA promoters but produce bidirectional, exosome-sensitive, relatively short unspliced RNAs, the generation of which is strongly related to enhancer activity. The atlas is used to compare regulatory programs between different cells at unprecedented depth, to identify disease-associated regulatory single nucleotide polymorphisms, and to classify cell-type-specific and ubiquitous enhancers. We further explore the utility of enhancer redundancy, which explains gene expression strength rather than expression patterns. The online FANTOM5 enhancer atlas represents a unique resource for studies on cell-type-specific enhancers and gene regulation.
Active enhancers are marked by bi-directional CAGE

Bi-directional CAGE signal on Enhancers Identified by H3K27Ac and H3K4me1 centred on P300

Acknowledgements
Robin Andersson
Albin Sandelin
Claudia Gebhard
Michael Rehli

Acknowledgements
Robin Andersson
Albin Sandelin
Claudia Gebhard
Michael Rehli
CAGE expression identifies cell-type-specific enhancer usage.

Dendrogram resulting from agglomerative hierarchical clustering of tissue samples based on their enhancer expression. Each leaf represents one CAGE tissue sample.

Acknowledgements
Robin Andersson
Albin Sandelin
Claudia Gebhard
Michael Rehli

Disease-associated SNPs are enriched in enhancers

Using the NHGRI genome-wide association studies (GWAS) catalogue and extending the compilation of lead SNPs with proxy SNPs in strong linkage disequilibrium we identified diseases/traits whose associated SNPs overlapped enhancers, promoters, exons and random regions significantly more than expected by chance (Fisher’s exact test \( P < 0.01 \), Supplementary Table 16).

Disease-associated SNPs were over-represented in regulatory regions to a greater extent than in exons (Fig. 6c).
Linking enhancers to TSSs and disease-associated SNPs.

eRNAs associated with disease SNPs are expressed in pathologically relevant samples
For many traits where enriched disease-associated SNPs were within enhancers, enhancer activity was detected in pathologically relevant cell types (Fig. 6d and Supplementary Figs 31 and 32).
- Graves’ disease-associated SNPs enriched in enhancers that are expressed predominantly in thyroid tissue,
- lymphocytes for chronic lymphocytic leukaemia.
Summary

• Comprehensive sequencing based expression atlas (non-coding and coding).

• The majority of our genes have cell-type-restricted expression patterns

• Key transcription factors defining cell states are identified

• Tissue specific disease associated genes can be expressed specifically or broadly

• Enhancer RNA expression profiles link causative SNPs to cell type of relevance
Reminder: Cell specialization

A. Hippocampal neuron
B. Alveolar macrophage
C. Pulmonary Artery Endothelial cells
D. Pancreatic islet
E. Inner ear hair cells
F. Rods and Cones

Acknowledgements: D Douglas Melton, Harvard Stem Cell Institute, E Donja Pyott, University of North Carolina, A, B, F (pending, google images) or C (wiki commons)
Cell-to-cell communication

• Cell specialization allows division of labour in multicellular organisms.
• To coordinate multicellular processes, cells need to communicate.
• Which cells are talking to each other and via what molecules?

Postdoctoral researcher
- Jordan Ramilowski

Scientific programmer
- Jayson Harshbarger
1. Curated Peptide Ligand (L) - Receptor (R) pairs

2. FANTOM5 CAGE expression data across broad collection of samples to construct cell – cell interactions based on the LR binding
Csf1r: Csf1 receptor

Expressed in which primary cells?

<table>
<thead>
<tr>
<th>Average TPM</th>
<th>cell type</th>
</tr>
</thead>
<tbody>
<tr>
<td>418.9</td>
<td>Langerhans cells, immature</td>
</tr>
<tr>
<td>266.7</td>
<td>Dendritic Cells - monocyte immature derived</td>
</tr>
<tr>
<td>256.8</td>
<td>Monocytes</td>
</tr>
<tr>
<td>234.3</td>
<td>Macrophage</td>
</tr>
<tr>
<td>110.7</td>
<td>Basophils</td>
</tr>
</tbody>
</table>
Csf1: Csf1 ligand
Expressed in which primary cells?

<table>
<thead>
<tr>
<th>Average TPM</th>
<th>cell type</th>
</tr>
</thead>
<tbody>
<tr>
<td>815.5</td>
<td>Mast cell</td>
</tr>
<tr>
<td>226.6</td>
<td>Macrophage</td>
</tr>
<tr>
<td>94.2</td>
<td>Synoviocyte</td>
</tr>
<tr>
<td>82.4</td>
<td>Hair Follicle Dermal Papilla Cells</td>
</tr>
<tr>
<td>77.1</td>
<td>Basophil</td>
</tr>
<tr>
<td>75.9</td>
<td>Dendritic Cells - monocyte immature derived</td>
</tr>
<tr>
<td>72.7</td>
<td>Meosthelial cells</td>
</tr>
<tr>
<td>60.3</td>
<td>Memory, regulatory T-cells</td>
</tr>
<tr>
<td>59.5</td>
<td>Pericyte</td>
</tr>
<tr>
<td>54.2</td>
<td>Adipocyte</td>
</tr>
<tr>
<td>53.1</td>
<td>Tenocyte</td>
</tr>
<tr>
<td>49.4</td>
<td>Trabecular Meshwork Cells</td>
</tr>
<tr>
<td>45.9</td>
<td>Chondrocyte</td>
</tr>
<tr>
<td>40.3</td>
<td>Mesenchymal stem cell</td>
</tr>
<tr>
<td>40.3</td>
<td>Myoblast</td>
</tr>
<tr>
<td>39.6</td>
<td>Preadipocyte</td>
</tr>
<tr>
<td>36.1</td>
<td>Endothelial Cells - Microvascular</td>
</tr>
<tr>
<td>35.7</td>
<td>Prostate stromal cells</td>
</tr>
<tr>
<td>33.9</td>
<td>Hair Follicle Outer Root Sheath Cells</td>
</tr>
<tr>
<td>32.7</td>
<td>Monocytes</td>
</tr>
<tr>
<td>31.9</td>
<td>Schwann cells</td>
</tr>
<tr>
<td>31.2</td>
<td>Astrocyte, cerebellar</td>
</tr>
<tr>
<td>30.6</td>
<td>Fibroblast</td>
</tr>
<tr>
<td>29.9</td>
<td>Smooth muscle cells</td>
</tr>
<tr>
<td>28.4</td>
<td>Hepatic Stellate Cells (lipocyte)</td>
</tr>
<tr>
<td>28.2</td>
<td>Cardiomyocyte</td>
</tr>
<tr>
<td>27.6</td>
<td>Keratocyte</td>
</tr>
<tr>
<td>27.6</td>
<td>CD34 Hematopoietic Stem Cells</td>
</tr>
<tr>
<td>27.3</td>
<td>Neutrophil</td>
</tr>
<tr>
<td>24.3</td>
<td>Perineurial cells</td>
</tr>
<tr>
<td>24.2</td>
<td>Meningeal cell</td>
</tr>
<tr>
<td>21.5</td>
<td>Nucleus Pulposus Cell</td>
</tr>
<tr>
<td>19.6</td>
<td>Osteocyte</td>
</tr>
<tr>
<td>19.5</td>
<td>Hepatic Sinusoidal Endothelial Cells</td>
</tr>
<tr>
<td>16.1</td>
<td>Anulus Pulposus Cell</td>
</tr>
<tr>
<td>16.1</td>
<td>Skeletal muscle cells</td>
</tr>
<tr>
<td>14.3</td>
<td>Eosinophil</td>
</tr>
<tr>
<td>12.0</td>
<td>Endothelial Cells - Lymphatic</td>
</tr>
</tbody>
</table>
Cell-to-cell web resource

http://forrest-lab.github.io/connectome

- Given a cell type of interest – what are the top/most specific receptors and ligands it expresses?
- Given a pair of cells – what are the top/most specific signalling paths between them?
- Given a ligand and receptor pair – which cells use that pair the most?

**Fig1: Csf1-Csf1r cell-to-cell signaling network.**
Primary cells from FANTOM5 expressing at least 25% of the maximum expression value for either CSF1 ligand or CSF1R receptor are shown. Arrow thickness scaled to [[ligand expression on transmitting cell] x [receptor expression on receiving cell]]. imoDC: immature monocyte derived dendritic cell.
How many ligands and receptors does each cell type express?

Average cell
• ~100-170 receptors and 60-200 ligands at appreciable levels (10TPM)
What about autocrine signalling?
There are 10s to 100s of pathways between any 2 cells.
Max-signalling pairs \( L_{\text{max}} - R_{\text{max}} \)

- **Mesenchymal**
  - Multicellular organismal development: 9.93E -03
  - Organ development: 2.11E -02
  - System development: 4.18E -02
  - Cell-matrix adhesion: 4.42E -02

- **Nervous system**
  - Organ development: 3.46E -02
  - Neuron development: 9.43E -03
  - Neuron differentiation: 5.79E -02

- **Epithelial**
  - Multi-cellular organismal development: 9.93E -03
  - System development: 4.18E -02

- **Endothelial**
  - Organ development: 1.12E -07
  - System development: 1.66E -04
  - System development: 2.25E -03

- **Hematopoietic**
  - Multi-cellular organismal development: 9.93E -03
  - Organ development: 5.79E -02
  - System development: 7.90E -02

- **Other**
  - Multi-cellular organismal development: 9.93E -03
  - Multi-cellular organismal development: 7.90E -02
  - System development: 7.90E -02

- **Signalling Pairs**
  - Mesenchymal > Epithelial: 6.02E -02
  - Epithelial > Hematopoietic: 5.79E -02
  - Epithelial > Neurovascular: 4.82E -02
  - Hematopoietic > Nervous System: 2.35E -03
  - Hematopoietic > Endothelial: 5.34E -03
  - Endothelial > Mesenchymal: 3.39E -02

**Pathways**
- GO:0007275: Multicellular organismal development
- GO:0048856: Anatomical structure development
- GO:0007160: Cell-matrix adhesion
- GO:0048731: System development
- GO:0032502: Developmental process
- GO:0006935: Chemotaxis
- GO:0001501: Skeletal system development
- GO:0007399: Nervous system development
- GO:0007275: Multicellular organismal development
- GO:0048513: Organ development
- GO:0007275: Multicellular organismal development
- GO:0048856: Anatomical structure development
- GO:0007167: Enzyme linked receptor protein signaling pathway
- GO:0048666: Neuron development
- GO:0022008: Neurogenesis
- GO:0030182: Neuron differentiation
- GO:0007167: Enzyme linked receptor protein signaling pathway
- GO:0048646: Anatomical structure formation involved in morphogenesis
- GO:0001525: Angiogenesis
- GO:0007178: Transmembrane receptor protein serine/threonine kinase signaling pathway
- GO:0048513: Organ development
| GO:0007275 | multicellular organismal development | 9.93E-03 |
| GO:0048856 | anatomical structure development | 9.93E-03 |
| GO:0007160 | cell-matrix adhesion | 2.11E-02 |
| GO:0048731 | system development | 4.18E-02 |
| GO:0032502 | developmental process | 4.42E-02 |
| GO:0007160 | cell-matrix adhesion | 2.11E-02 |
| GO:0048731 | system development | 4.18E-02 |
| GO:0032502 | developmental process | 4.42E-02 |

**Mesenchymal > Hematopoietic Signalling**

- GO:0006955 immune response
  - GO:0006952 defense response
  - GO:0006954 inflammatory response
  - GO:0006953 immune system process
  - GO:0006951 immune response
  - GO:0006950 immune system process

**Epithelial > Hematopoietic Signalling**

- GO:0002376 immune system process
  - GO:0002375 immune system process
  - GO:0002374 immune system process

**Hematopoietic > Epithelial Signalling**

- GO:0001664 G-protein coupled receptor binding
  - GO:0001663 G-protein coupled receptor binding
  - GO:0001662 G-protein coupled receptor binding

**Multicellular Processes**

- GO:0007275 multicellular organismal development
- GO:0048856 anatomical structure development
- GO:0007160 cell-matrix adhesion
- GO:0048731 system development
- GO:0032502 developmental process

**Endothelial > Hematopoietic Signalling**

- GO:0007155 cell adhesion
  - GO:0007389 pattern specification process
  - GO:0003002 regionalization

**Endothelial > Epithelial Signalling**

- GO:0007389 pattern specification process
  - GO:0007388 pattern specification process
  - GO:0007387 pattern specification process

**Endothelial > Mesenchymal Signalling**

- GO:0007160 cell-matrix adhesion
  - GO:0007155 cell adhesion

**Nervous System > Mesenchymal Signalling**

- GO:0007176 neurological system process
  - GO:0007175 neurological system process

**Mesenchymal > Endothelial Signalling**

- GO:0007167 enzyme linked receptor protein signaling pathway
  - GO:0007155 cell adhesion
  - GO:0007389 pattern specification process
  - GO:0003002 regionalization

**Epithelial > Endothelial Signalling**

- GO:0007389 pattern specification process
  - GO:0007388 pattern specification process
  - GO:0007387 pattern specification process

**Hematopoietic > Mesenchymal Signalling**

- GO:0007160 cell-matrix adhesion
  - GO:0007155 cell adhesion

**Hematopoietic > Epithelial Signalling**

- GO:0007242 intracellular signal transduction
  - GO:0007241 intracellular signal transduction
  - GO:0007240 intracellular signal transduction

**Hematopoietic > Nervous System Signalling**

- GO:0007155 cell adhesion
  - GO:0007389 pattern specification process
  - GO:0003002 regionalization

**Epithelial > Nervous System Signalling**

- GO:0007176 neurological system process
  - GO:0007175 neurological system process
  - GO:0007174 neurological system process

**Nervous System > Endothelial Signalling**

- GO:0007167 enzyme linked receptor protein signaling pathway
  - GO:0007155 cell adhesion
  - GO:0007389 pattern specification process
  - GO:0003002 regionalization

**Hematopoietic > Nervous System Signalling**

- GO:0007155 cell adhesion
  - GO:0007389 pattern specification process
  - GO:0003002 regionalization
Mesenchymal > mesenchymal signalling
- GO:0007275 multicellular organismal development 9.93E-03
- GO:0048856 anatomical structure development 9.93E-03
- GO:0007160 cell-matrix adhesion 2.11E-02
- GO:0004731 system development 4.18E-02
- GO:00032502 development 4.42E-02

Epithelial > hematopoietic signalling
- GO:0006952 defense response 2.35E-03
- GO:0009605 response to external stimulus 5.34E-03
- GO:0006954 inflammatory response 5.34E-03
- GO:00030005 cellular di-, tri-valent inorganic cation homeostasis 5.34E-03
- GO:0045087 innate immune response 5.34E-03

Mesenchymal > hematopoietic signalling
- GO:0006935 chemotaxis 4.82E-02

Mesenchymal > epithelial signalling
- GO:0048513 organ development 2.11E-02
- GO:0007275 multicellular organismal development 3.42E-04
- GO:0007160 cell-matrix adhesion 1.72E-03
- GO:0007167 enzyme linked receptor protein signaling pathway 2.04E-03
- GO:0001501 skeletal system development 5.19E-03

Mesenchymal > nervous system signalling
- GO:0007399 nervous system development 9.43E-03
- GO:0007275 multicellular organismal development 5.79E-02
- GO:00048666 neuron development 6.02E-02
- GO:00022008 neurogenesis 7.90E-02
- GO:00031482 neuron differentiation 7.90E-02

Mesenchymal > endothelial signalling
- GO:0007167 enzyme linked receptor protein signaling pathway 1.12E-07
- GO:00048666 anatomical structure formation involved in 1.66E-04
- GO:0001525 angiogenesis 2.25E-03
- GO:0007178 transmembrane receptor protein serine/threonine kinase signaling pathway 2.25E-03
- GO:0048513 organ development 2.25E-03

Epithelial > hematopoietic signalling
- GO:0006952 defense response 2.35E-03
- GO:0009605 response to external stimulus 5.34E-03
- GO:0006954 inflammatory response 5.34E-03
- GO:00030005 cellular di-, tri-valent inorganic cation homeostasis 5.34E-03
- GO:0045087 innate immune response 5.34E-03

Epithelial > endothelial signalling
- GO:0006952 defense response 4.96E-04
- GO:0009605 response to external stimulus 1.85E-03
- GO:0006954 inflammatory response 1.85E-03
- GO:00050878 regulation of body fluid levels 4.52E-03
- GO:0009653 anatomical structure morphogenesis 8.32E-03

Endothelial > hematopoietic signalling
- GO:0007155 cell adhesion 2.60E-02
- GO:0007389 pattern specification process 4.86E-02
- GO:0003002 regionalization 4.86E-02

Endothelial > epithelial signalling
- GO:0007389 pattern specification process 1.20E-02
- GO:00048869 cellular developmental process 1.20E-02
- GO:0001501 skeletal system development 1.20E-02
- GO:0006507 biological regulation 2.43E-02
- GO:0048513 organ development 2.94E-02

Endothelial > mesenchymal signalling
- GO:0007167 cell-matrix adhesion 3.39E-02

Endothelial > nervous system signalling
- GO:00050877 neurological system process 3.46E-02

Nervous system > mesenchymal signalling
Cell-to-cell communication applications

• Identifying growth factors

**Fig3:** CCL2 promotes expression of pluripotency genes in stem cells

**Fig4:** BMP2 ligand promotes survival of human mast cells

Yuki Hasegawa – RIKEN, Japan

Magda Babina – Charite, Germany
Cell-to-cell communication in cancer

PhD student – Riti Roy

Hanahan and Weinberg, 2011
What messages are sent between these cell types?

PhD student – Riti Roy
Cell-cell communication in cancer

• Which cells are talking to each other?

• Cancer cells talk to:
  – Themselves (positive feedback loop)
  – Blood vessels (please feed me)
  – Immune cells (please ignore me)
Potential collaborations

Disease-gene & Disease-enhancer associations

Cell-to-cell communication in disease (single cell data)

Transcriptome experiments & TRNs

Drug repurposing

Pan-cancer biomarkers
• 261 authors
• 120 international collaborators
• 19 countries

**Australian contributors**

Anthony G. Beckhouse\textsuperscript{21,22}, Antje Blumenthal\textsuperscript{22,25}, James Briggs\textsuperscript{21}, Geoffrey J. Faulkner\textsuperscript{52}, Kelly J. Hitchens\textsuperscript{21,22}, Tony J. Kenna\textsuperscript{25}, Levon M. Khachigian\textsuperscript{69}, S. Peter Klinken\textsuperscript{71}, Alan Mackaysim\textsuperscript{74}, Dmitry A. Ovchinnikov\textsuperscript{21}, Margaret Patrikakis\textsuperscript{69}, Dipti Vijayan\textsuperscript{21,22}, Christine A. Wells\textsuperscript{21,22}, Louise N. Winteringham\textsuperscript{71}, Ernst Wolvetang\textsuperscript{21}.

\textsuperscript{21}Australian Institute for Bioengineering and Nanotechnology (AIBN), UQ

\textsuperscript{22}Australian Infectious Diseases Research Centre (AID), UQ

\textsuperscript{25}Diamantina Institute, UQ

\textsuperscript{52}Cancer Biology Program, Mater Medical Research Institute, Raymond Tce, South Brisbane, QLD

\textsuperscript{74}National Centre for Adult Stem Cell Research, Eskitis Institute for Cell and Molecular Therapies, Griffith University, Brisbane, Queensland

\textsuperscript{69}Centre for Vascular Research, University of New South Wales, Sydney NSW.

\textsuperscript{71}Harry Perkins Institute of Medical Research, and the Centre for Medical Research, University of Western Australia. Perth, WA
RIKEN acknowledgements

Bioinformatics
Kawaji Hideya
Michiel de Hoon
Timo Lassmann
Marina Lizio
Erik Arner
Max Burroughs
Jessica Severin
Nicolas Bertin
Jayson Harshburger
Hisashi Shimoji
Andrew Kwon
Jordan Ramilowski
Bogumil Kaczkowski

Sample handling
Yuri Ishizu
Hiromi Nishiyori
Sachiko Ishikawa

FANTOM5 HQ
Piero Carninci
Carsten Daub
Harukazu Suzuki
Jun Kawai

Yoshihide Hayashizaki

Library production
Masayoshi Itoh
Sayaka Sato
Miki Kojima
Sakai Mizuho
Mitsuyoshi Murata
Mutsumi Kanamori-Katayama

Funding:
MEXT JAPAN

Helicos sequencing service
GeNAS (Gene Network Analysis Service)
http://www.osc.riken.jp/english/technicalservice

Normal brain samples
Peter Heutink and Patricia Rizzu - VUMC
Brains were provided by the Dutch Brain Bank.

Tissue specificity of disease
Jessica Mar – Albert Einstein NY, USA
Alka Saxena - Guy's and St Thomas' NHS, UK

http://fantom.gsc.riken.jp/5/
In 2015 I started a new position in Perth Western Australia at the Harry Perkins Institute of Medical Research

Always looking for good students and postdocs

Any questions feel free to drop me a line: 
alistair.forrest@perkins.uwa.edu.au