Immunological Genome Project:

Does the big immunological/hematological database support good little science?

Yasuyuki Saito, M.D. , Ph.D.

Exp. Hematology, USZ

Overview

1. Immunological Genome Project
2. Workflow and data analysis
3. Recent findings using ImmGen database
4. ImmGen data browsers

Introduction

- Classical immunology studies are mainly focused on a particular protein or biological process of a particular type of the cells/organs.

- With technological advances many type of the immune/hematopoietic cells as well as their progenitors have been identified.

- It is feasible for individual laboratories to understand which signals / molecules are regulated in a particular cell lineage, whereas it's difficult to understand how molecular networks operate immune system for individual laboratories.

- ImmGen is a consortium aiming to generate gene expression profiles for primary immune / hematopoietic cells from different anatomical locations that were analyzed directly ex vivo at finely defined stages of differentiation.

The pros and cons of consortium biology

The pros

1. The only way of getting to the result for some large questions

2. Promotes a culture of openness, through data and material sharing

3. Promotes uniform standards, data and data formats that are compatible between labs, as well as the use of common nomenclature and shared reagents.

4. Results can diffuse broadly, extend, enrich the "little science" that follows

5. High-thoughputout approaches can be more "economical" than repetitive conventional approaches

The cons

1. Need to compete for limited funding

2. suppress scientific creativity by promoting "science by committee"

3. establishes self-perpetuating structures that tend to create projects to ensure their own survival, rather than for clear scientific need.

4. Results can diffuse broadly, extend, enrich the "little science" that follows

5. It is wasteful; "Spending money instead of thought"

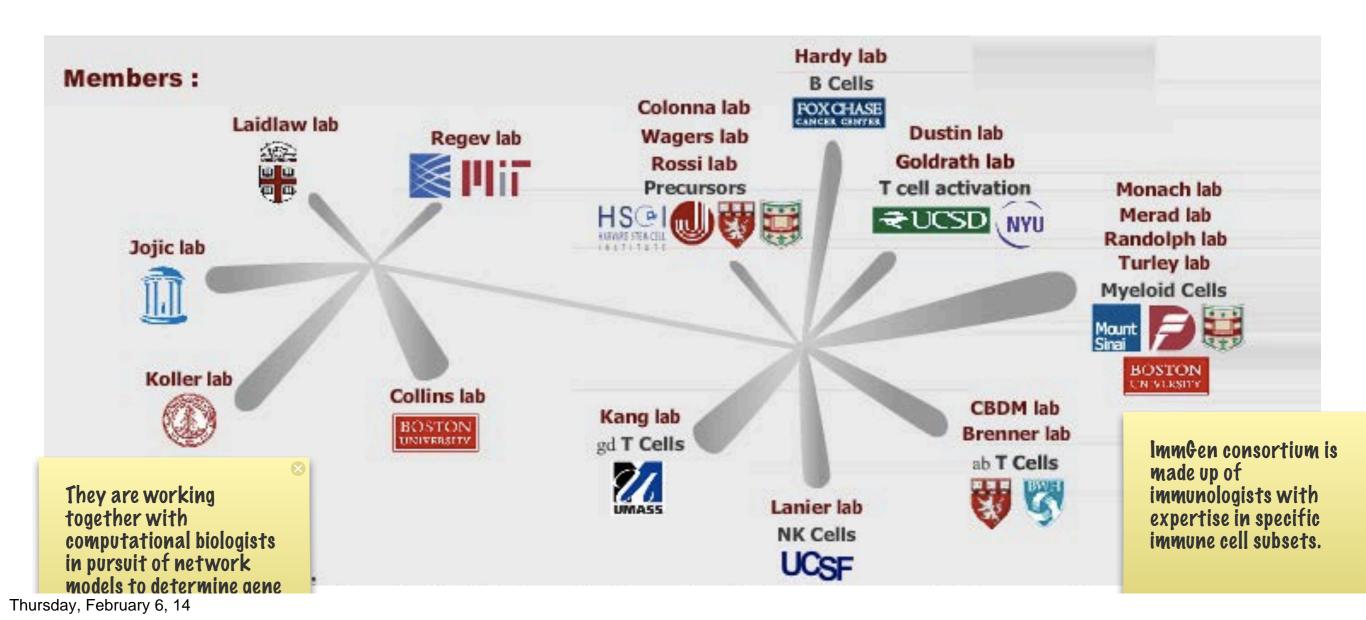Modified from Benoist et al., Nature Rev. Immunol, 2012 vol.12 734

Immunological Genome Project (ImmGen Project)

The Immunological Genome Project: networks of gene expression in immune cells
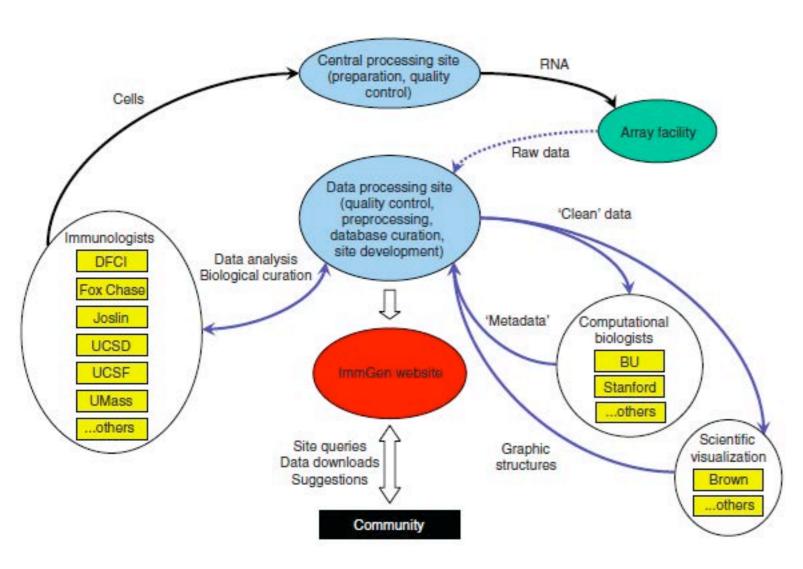
Tracy S P Heng, Michio W Painter & The Immunological Genome Project Consortium

Heng TSP et al., Nature Immunol, 2008 vol.9 1091-1094

- Initiated in 2007, released April 2012
- Aim: to establish a comprehensive, public compendium of gene networks
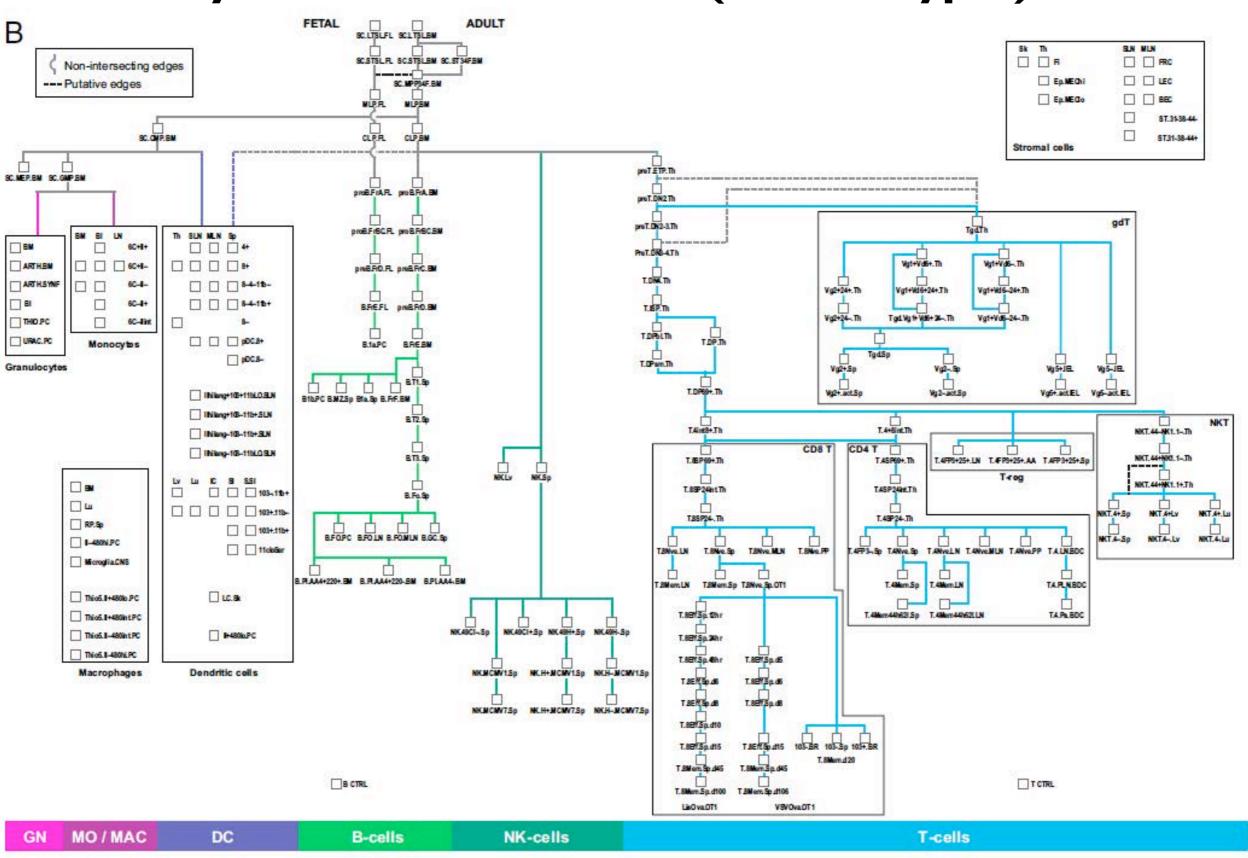- a collaborative group of 15 immunology and 5 computational biology laboratories
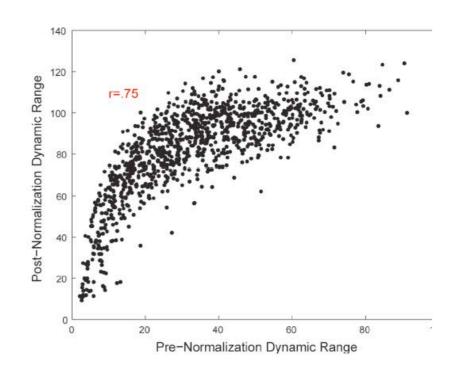
Project workflow

- All the mice used were genetically identical and sourced from single location (C57BL/6J from Jackson lab.)

- At least 3 mice were pooled, and tissue was harvested at a fixed time (between 8:30 and 9:30) to avoid circadian variation.

- The cells were double-sorted and $1\text{-}5\times10^4$ cells were collected.

- Homogeneity, RNA preparation, probe labeling and hybridization are done in a central processing site.
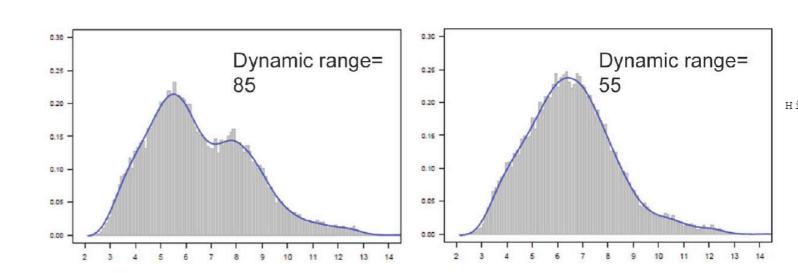
mouse hematopoietic/immune cell lineage tree profiled by ImmGen consortium (244 cell types)
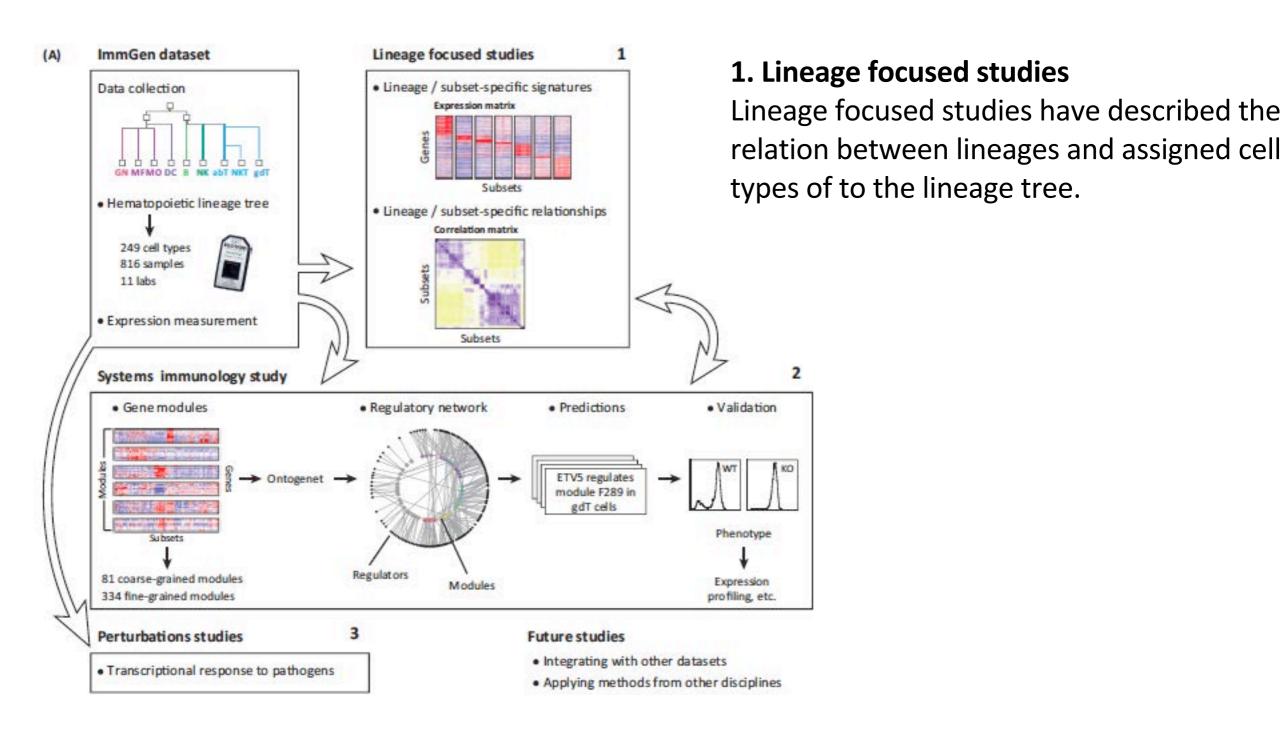
Data Processing

• Micoarray-based technique is applied to the initial analysis (phase I)

• The Affymetrix Mouse Gene ST1.0 Array (covered 26,166 transcripts in total) was used

• The Dynamic range was normalized, and if the DR was <40, sample was systemically dropped.

Data analysis

(A) ImmGen dataset

Data collection

GN MFMO DC B NK abT NKT gdT

- Hematopoietic lineage tree

249 cell types
816 samples
11 labs

- Expression measurement

Lineage focused studies 1

- Lineage / subset-specific signatures

Expression matrix

Genes / Subsets

- Lineage / subset-specific relationships

Correlation matrix

Subsets / Subsets

Systems immunology study 2

- Gene modules
- Regulatory network
- Predictions
- Validation

Modules / Genes / Subsets → Ontogenet → Regulators / Modules → ETV5 regulates module F289 in gdT cells → WT / KO / Phenotype → Expression profiling, etc.

81 coarse-grained modules
334 fine-grained modules

Perturbations studies 3

- Transcriptional response to pathogens

Future studies

- Integrating with other datasets
- Applying methods from other disciplines

1. Lineage focused studies

Lineage focused studies have described the relation between lineages and assigned cell types of to the lineage tree.

Gene expression profile across hematopoietic subsets

findings

Deciphering the transcriptional network of the dendritic cell lineage

Jennifer C Miller[1,2], Brian D Brown[1,3], Tal Shay[4], Emmanuel L Gautier[1,5,6], Vladimir Jojic[7,10], Ariella Cohain[3], Gaurav Pandey[3], Marylene Leboeuf[1,2], Kutlu G Elpek[8], Julie Helft[1,2], Daigo Hashimoto[1,2], Andrew Chow[1,2,9], Jeremy Price[1,2], Melanie Greter[1,2,7], Milena Bogunovic[1,2], Angelique Bellemare-Pelletier[8], Paul S Frenette[9], Gwendalyn J Randolph[1,5,6], Shannon J Turley[8], Miriam Merad[1,2] & the Immunological Genome Consortium[11]

Show gene signature of DC lienage cells and difference with macrophages at progenitor cell level

Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages

Emmanuel L Gautier[1–3], Tal Shay[4], Jennifer Miller[2,5], Melanie Greter[2,5], Claudia Jakubzick[1,2], Stoyan Ivanov[3], Julie Helft[2,5], Andrew Chow[2,5], Kutlu G Elpek[6,7], Simon Gordonov[8], Amin R Mazloom[8], Avi Ma'ayan[8], Wei-Jen Chua[3], Ted H Hansen[3], Shannon J Turley[6,7], Miriam Merad[2,5], Gwendalyn J Randolph[1–3] & the Immunological Genome Consortium[9]

Gautier, E. et al. (2012). *Nature Immunology*, *13*(11), 1118–1128.

Identify Macrophage specific TF (TCEF3, CEBPa, Bach1 and CREG-1) together with specific marker (MerTK , CD64)

Intrathymic programming of effector fates in three molecularly distinct γδ T cell subtypes

Kavitha Narayan[1,5], Katelyn E Sylvia[1,5], Nidhi Malhotra[1], Catherine C Yin[1], Gregory Martens[2], Therese Vallerskog[2], Hardy Kornfeld[2], Na Xiong[3], Nadia R Cohen[4], Michael B Brenner[4], Leslie J Berg[1], Joonsoo Kang[1] & The Immunological Genome Project Consortium[6]

Identify three genetically distinct γδT cell subtypes (naive, IL-17+ effector, IFNg+ effector) in thymus

Narayan, K., et al. (2012). *Nature Immunology*, *13*(5), 511–518.

Shared and distinct transcriptional programs underlie the hybrid nature of *i*NKT cells

Nadia R Cohen[1,5], Patrick J Brennan[1,5], Tal Shay[2], Gerald F Watts[1], Manfred Brigl[1,3], Joonsoo Kang[4], Michael B Brenner[1] & The Immunological Genome Project Consortium[6]

Invariant NKT cells shared genetical feature of NK, αβT cells

Cohen, N. R., et al. (2013). *Nature Immunology*, *14*(1), 90–99.

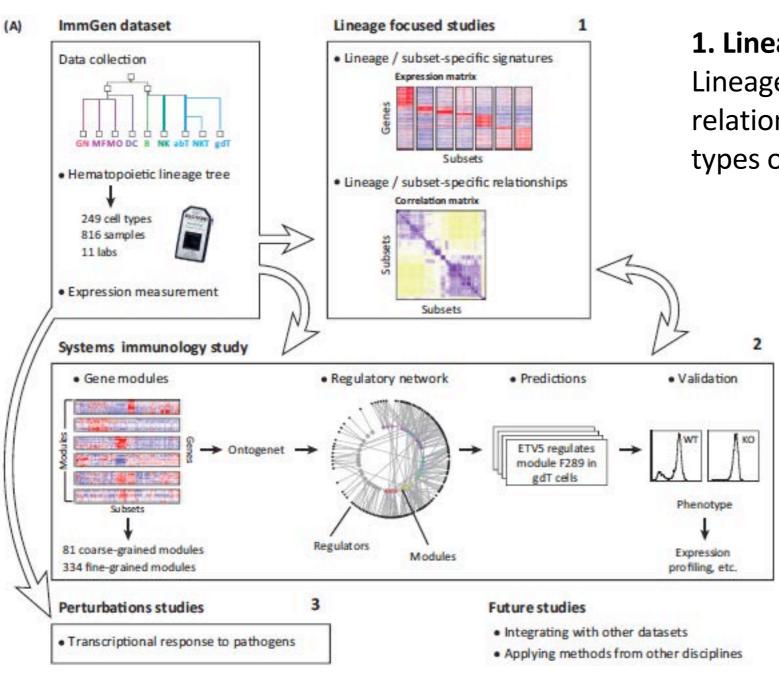## Transcriptional profiling of stroma from inflamed and resting lymph nodes defines immunological hallmarks

Deepali Malhotra[1,2,8], Anne L Fletcher[1,8], Jillian Astarita[1,2], Veronika Lukacs-Kornek[1], Prakriti Tayalia[3], Santiago F Gonzalez[4], Kutlu G Elpek[1], Sook Kyung Chang[5], Konstantin Knoblich[1], Martin E Hemler[1], Michael B Brenner[5,6], Michael C Carroll[4], David J Mooney[3], Shannon J Turley[1,7] & the Immunological Genome Project Consortium[9]

Lymph node stromal cells, especially FRCs express genes relevant to cytokine signaling.

Malhotra, D., et al. (2012). *Nature Immunology*, *13*(5), 499–510.

Data analysis

1. Lineage focused studies

Lineage focused studies have described the relation between lineages and assigned cell types of to the lineage tree.

2. Systems immunology study

System studies that have defined modules of coexpressed genes across the entire dataset and reconstructed the regulatory program of the modules using Ontogenet

Ontogenet: reconstructing lineage-sensitive regulation

Ontogenet: a new ALGORITHM to delineate the regulatory circuits that drive hematopoietic cell differentiation

Criterion 1: the expression of each module of genes is determined by a combination of activating and repressing transcription factors

Criterion 2: the activity of those transcription factors may change in different cell types
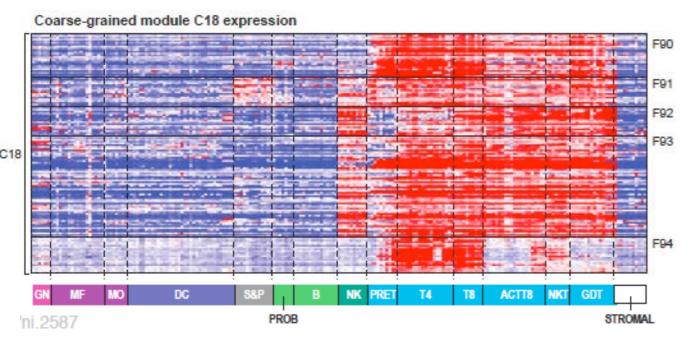
Criterion 3: the identity and activity of the transcription factors that regulate one module are more similar in cells that are close to each other in the lineage tree
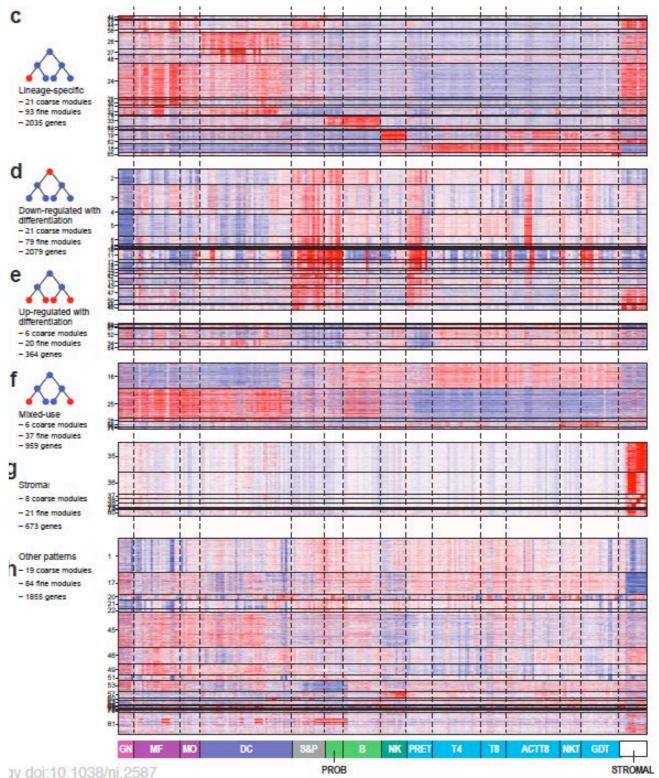
Criterion4: master regulators of a lineage are active across the sublineages, nut the subtypes can also have additional, more specific regulators (GATA-3 for whole T cells vs Foxp3 for Treg cells)

mouse hematopoietic/immune cell lineage tree profiled by ImmGen consortium (244 cell types)
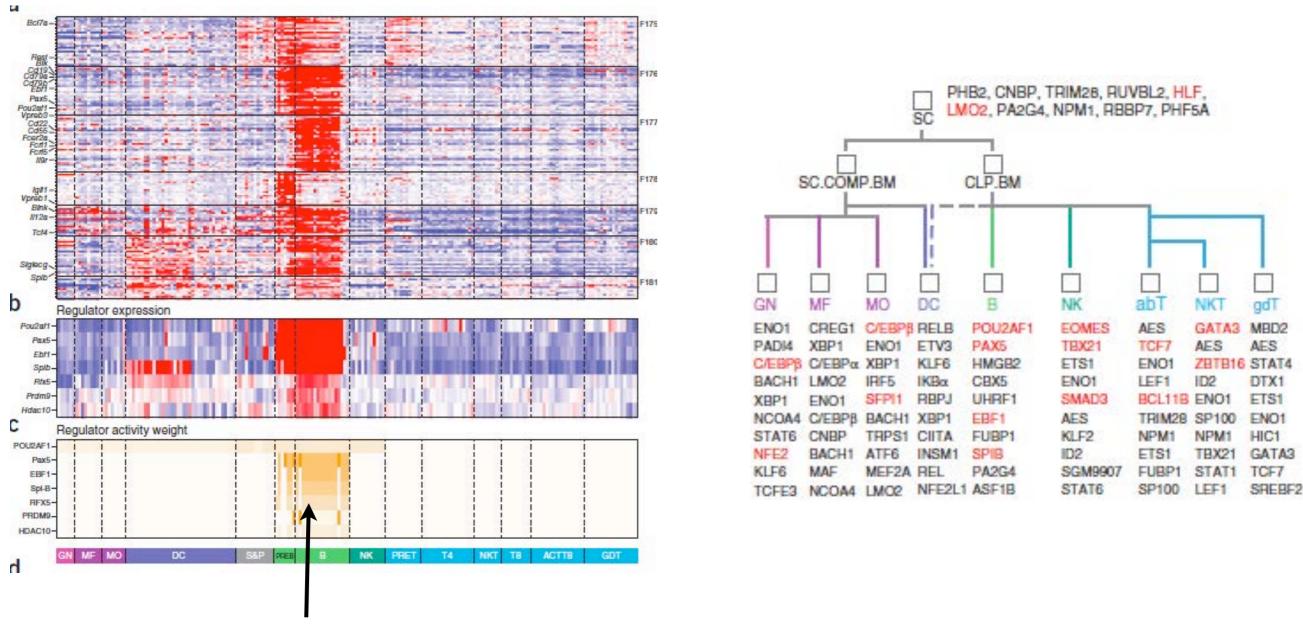
Coarse- and fine- grained module

Coarse-grained module C18 expression

Coarse grained module: total 81
Fine grained module: total 334

Module C33: early B molecule, relevant to B cell development

Regulator of C33 module calculated by ontoget algorithm

Integrating ImmGen with other datasets
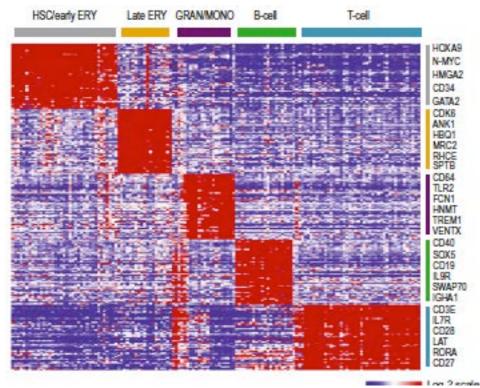
- -38 type of distinct cell populations from 4 to 7 donors (CB or PB)
- -Affymetrix HG_U133AAofAv2 microarrays (22,944 probes)

Signature genes characterizing the lineages
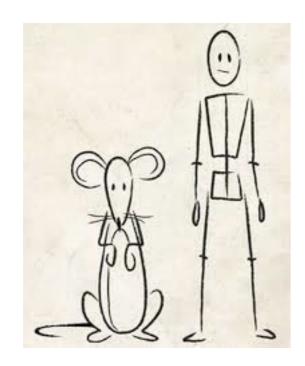
Lineage-specific regulation of TF expression
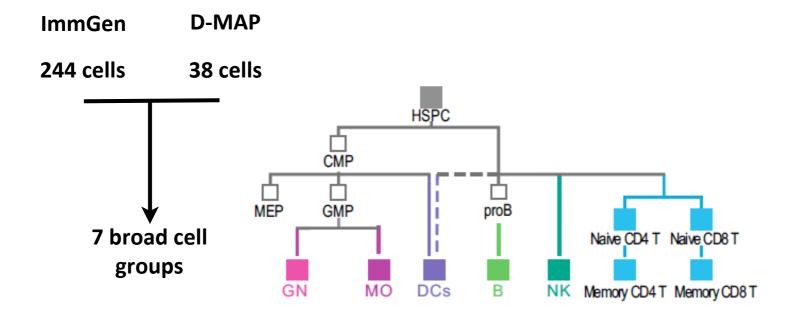
Conservation and divergence in the transcriptional programs of the human and mouse immune systems

Tal Shay[a,1], Vladimir Jojic[b,1], Or Zuk[a], Katherine Rothamel[c], David Puyraimond-Zemmour[c], Ting Feng[c], Ei Wakamatsu[c], Christophe Benoist[c,2], Daphne Koller[b], Aviv Regev[a,d], and the ImmGen Consortium[3]
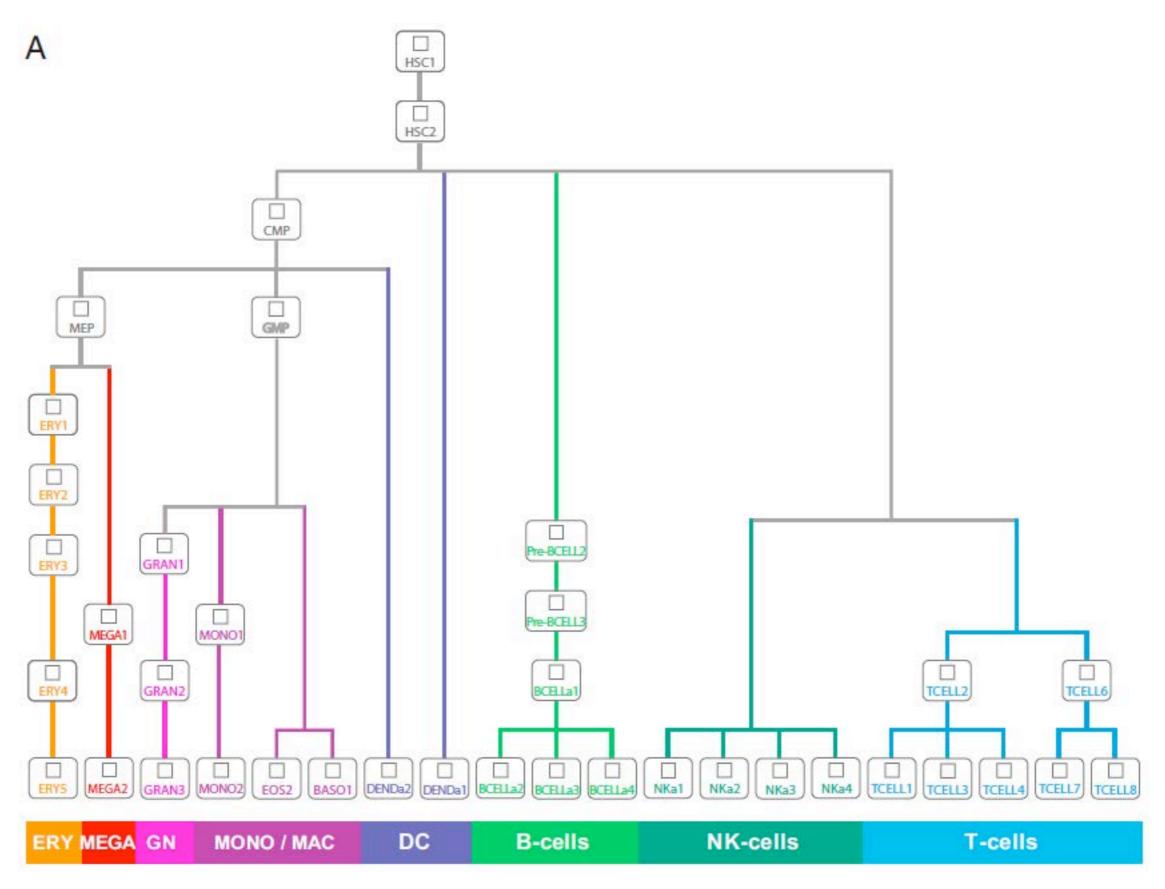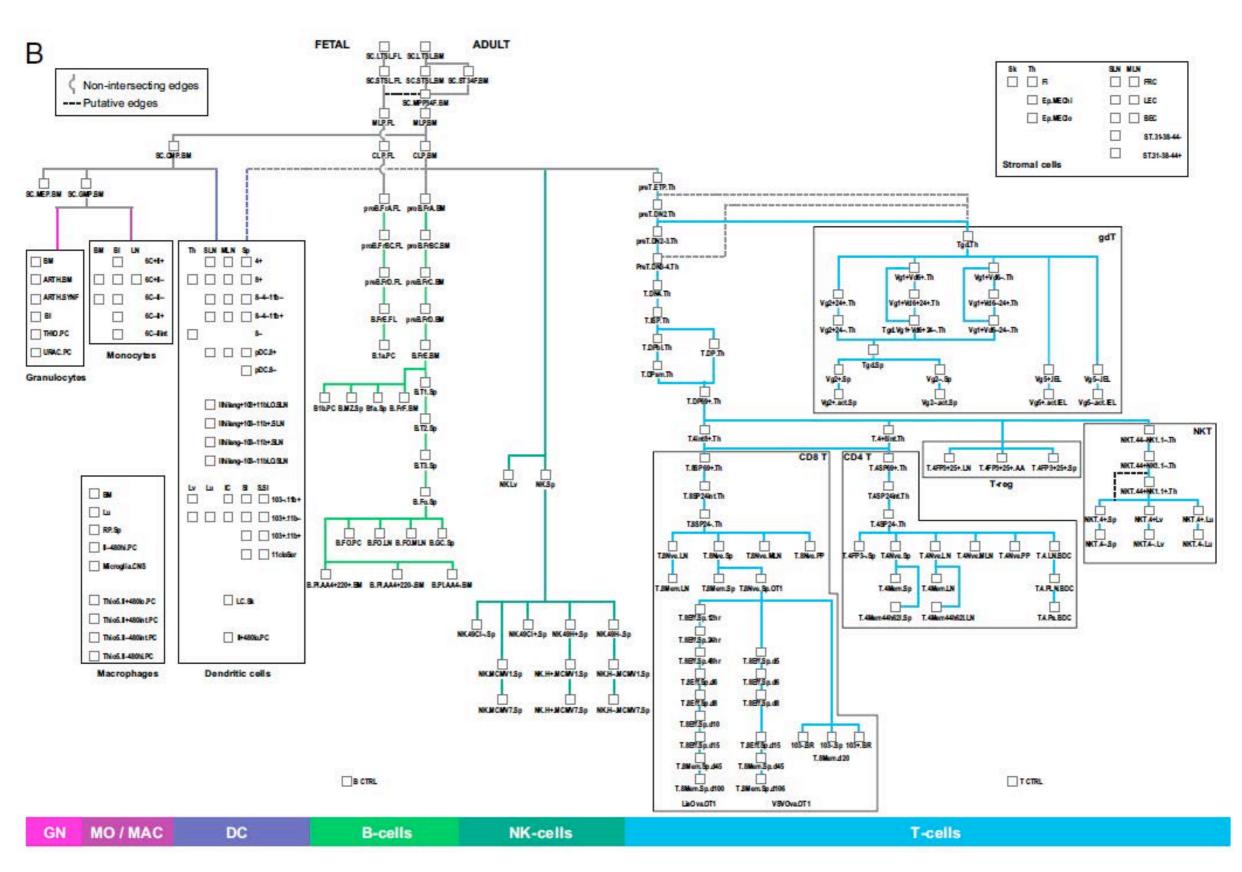
Using the Ensembl COMPARA database, 10,248 one-to-one orthologs between the two species are measured.
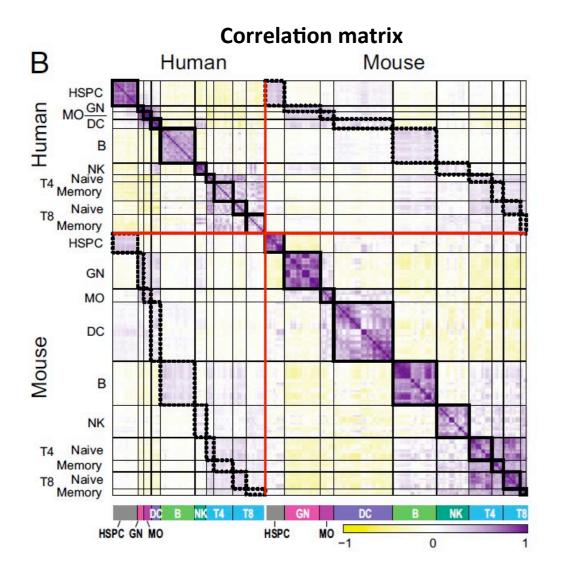

ImmGen   D-MAP

244 cells   38 cells

7 broad cell groups

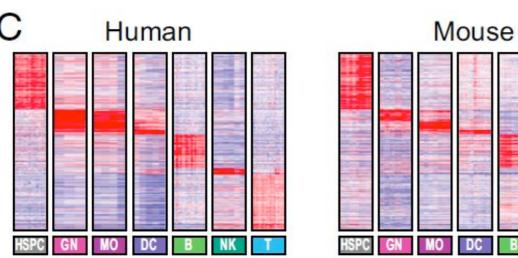# human hematopoietic/immune cell lineage tree profiled by DMAP consortium (38 cell types)

mouse hematopoietic/immune cell lineage tree profiled by ImmGen consortium (244 cell types)

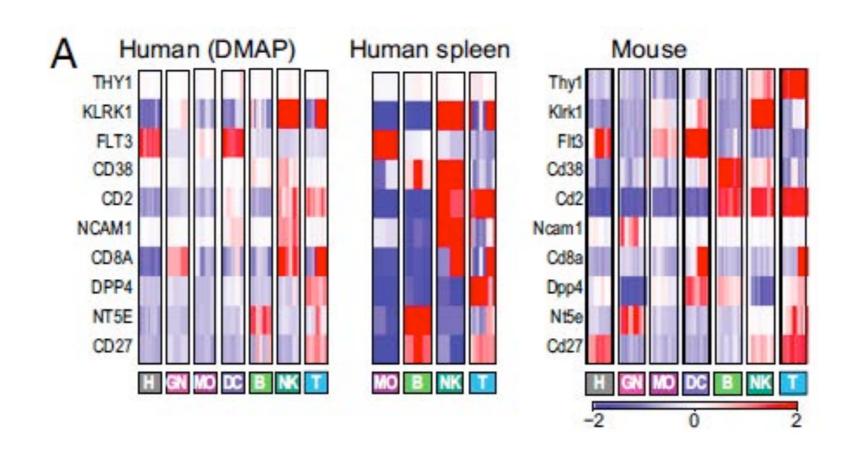# The grobal expression profiles of matching lineages between human and mouse

Correlation matrix

-The global expression profiles of matching lineages are similar between the two species for both individual cell types and average group profiles
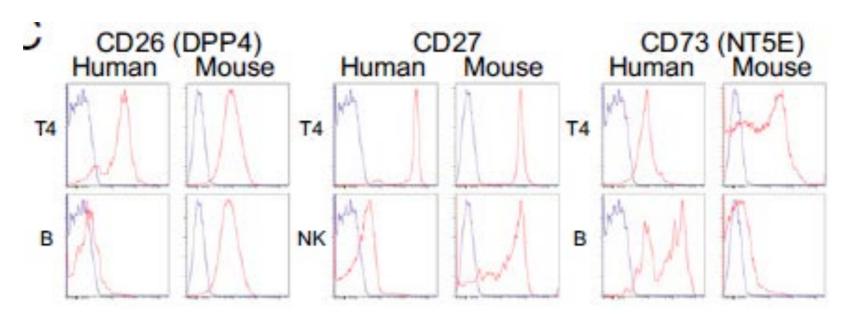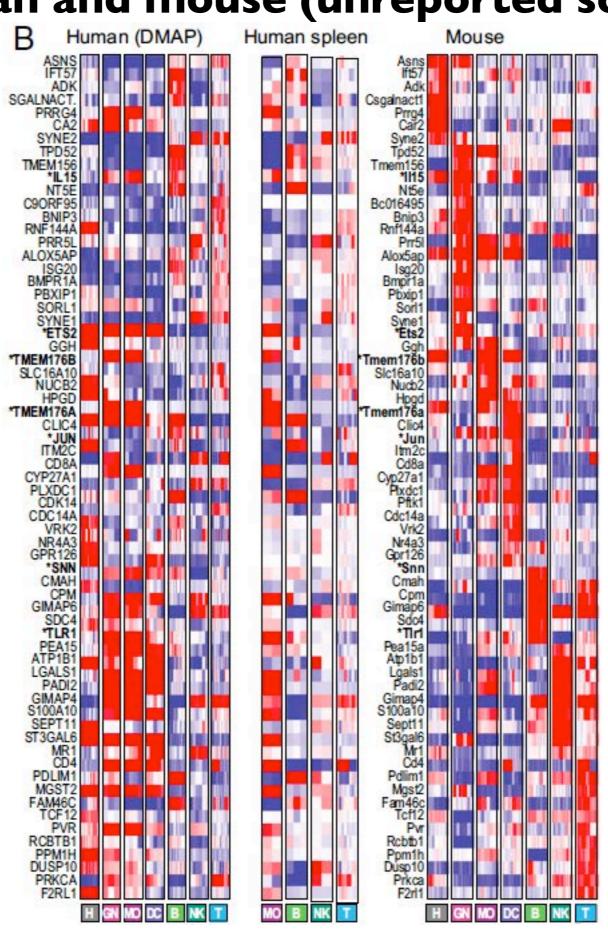
-Mean-centered expression values of the genes shared between matching lineage signatures are also similar between two species.

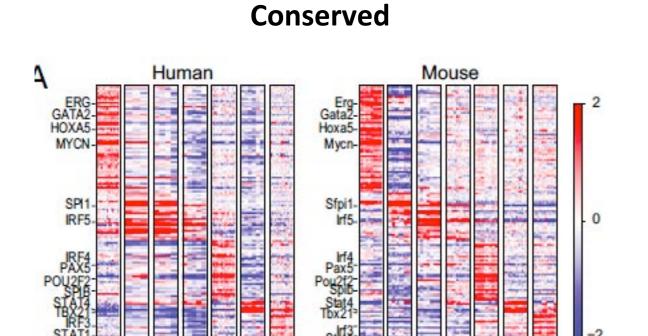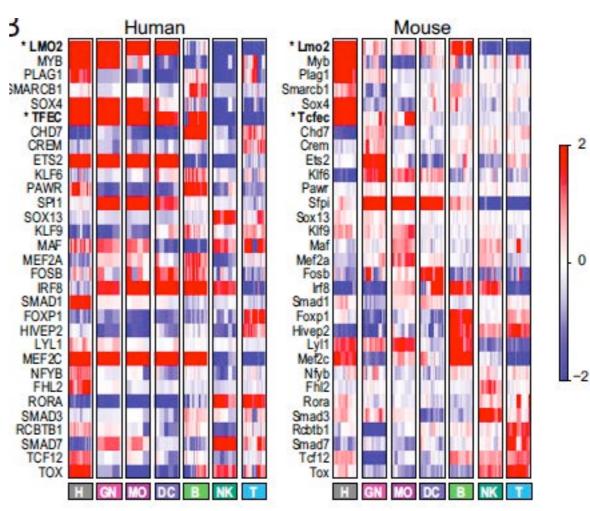# Differentially expressed genes between human and mouse (previously reported)

Differentially expressed genes between human and mouse (unreported so far)

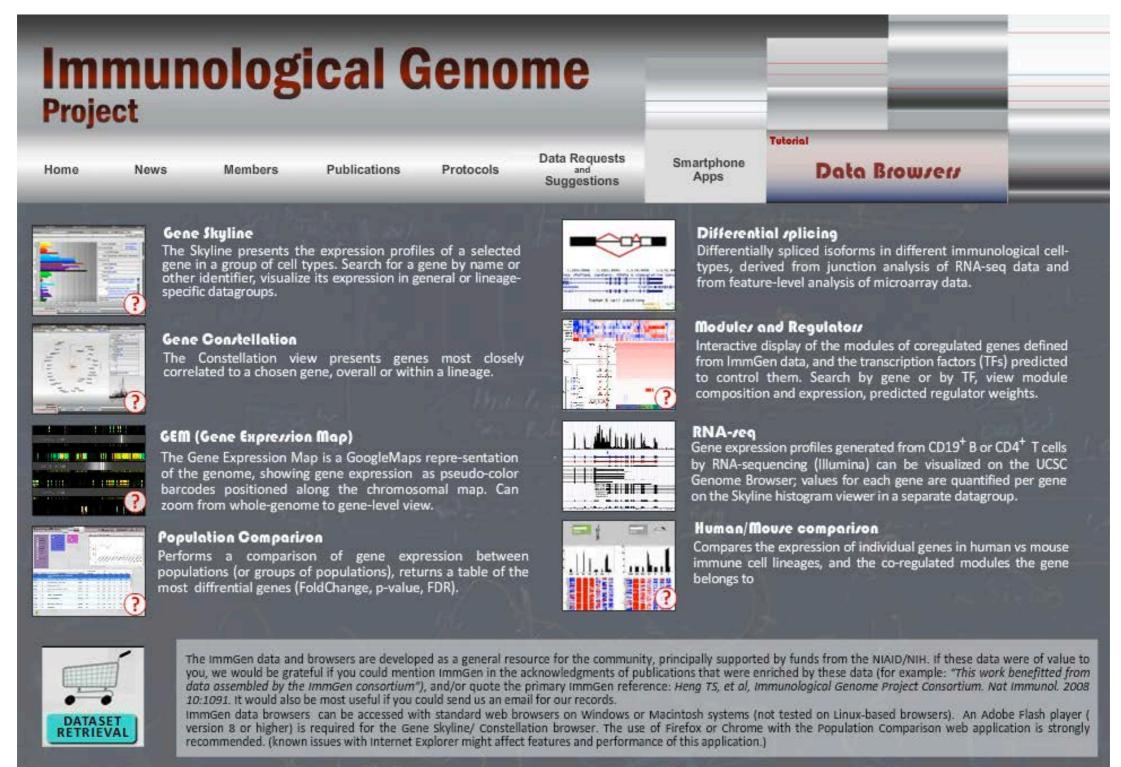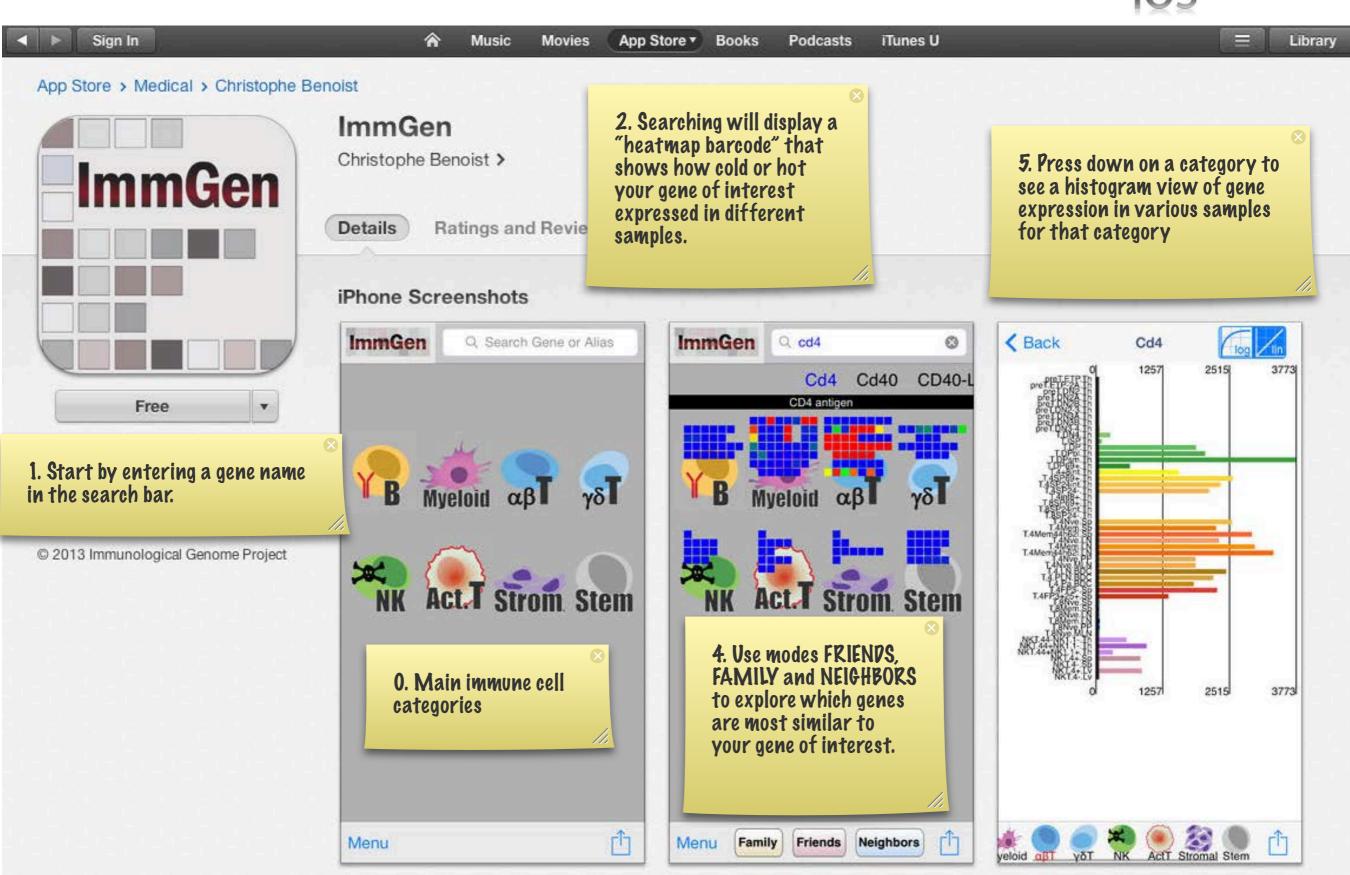# Conserved and Divergent lineage-specific expression of regulators

Divergent

Conserved

ImmGen data browser

Summary

1. Immunological Genome Project, is a part of consortium biology in immunology aiming to achieve comprehensive gene network in mouse immune/hematopoietic system

2. The powerful dataset have provided many evidences regarding i) gene expression profiles across hematopoietic subsets, ii) developmental intermediates, and iii) subset-specific perturbations

3. Ontogenet program developed from the ImmGen dataset will predict a set of transcriptional regulators which are unknown so far

4. The power of immGen dataset is not only in the analysis of the data within it, but in the integration with external systemic datasets, which can amplify informational outputs.

5. ImmGen software and App can be easy to analyze such a big database and will apply to our good little science.

Thank you very much!