“Monogenic” contributions to complex traits: scaling pleiotropy

Judy H. Cho, M.D.
Ward-Coleman Professor of Translational Genetics
Director, Charles F. Bronfman Institute for Personalized Medicine
Icahn School of Medicine at Mount Sinai

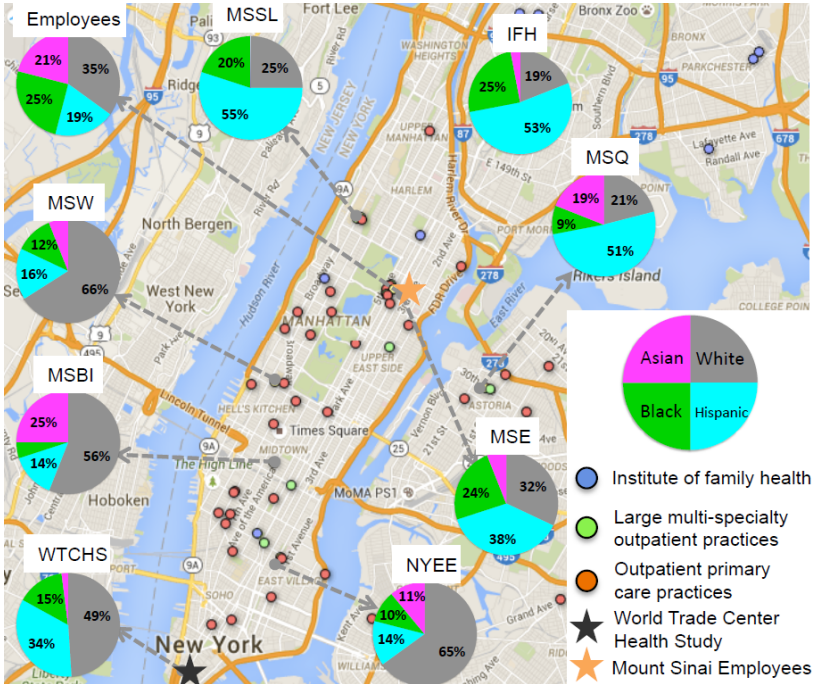
Mount Sinai

March 1, 2018

The Short-term Challenge for BioMe

The Mount Sinai BioMe – EMR-linked Biobank

1. Founded in September 2007 (by Dr. Erwin Bottinger),
2. an ongoing, consented EMR-linked bio- and data repository, additional data from questionnaire
3. A longitudinal design, including past, current and future records,
4. enrolls participants non-selectively from MSMC's patient population of 80,000 inpatient and 700,000 outpatient visits annually
5. Currently 41,000 participants,
6. participants are 40% Hispanic, 35% European, 25% African-American
7. recall based on phenotype or/and genotype for in-depth follow-up studies,



Ruth Loos, Ph.D.

Eimear Kenny, Ph.D.

Advantages to Health-System based Biobanks

1. "Case" vs. "control" status is overly-simplistic
2. Health Systems have enormous "phenotype" data: labs, radiology, pathology data
3. Development of disease is *time*-dependent
4. Flipping the genetic paradigm

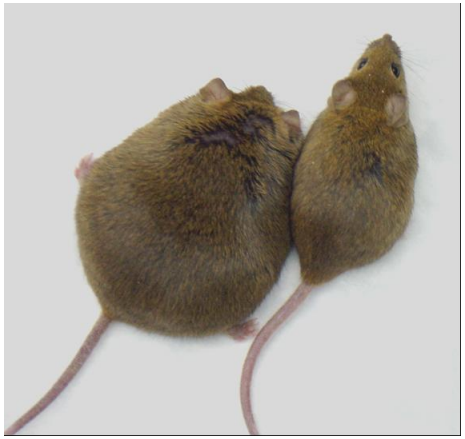
Redefining human health & disease through a much deeper study of human traits

Regeneron exome sequencing the entire BioMe Biobank: 32K sequences back

- Large non-European ancestry cohorts linked to a health system biobank
- Identifying new loss-of-function mutations, never before identified in humans (Human Knockout Project)

Studying humans with naturally occurring gene knockouts identified by sequencing

- *Safety* information for targeting/blocking a pathway
- Identifying drug targets for protective, loss-of-function variants—GPCRs, kinases, channels

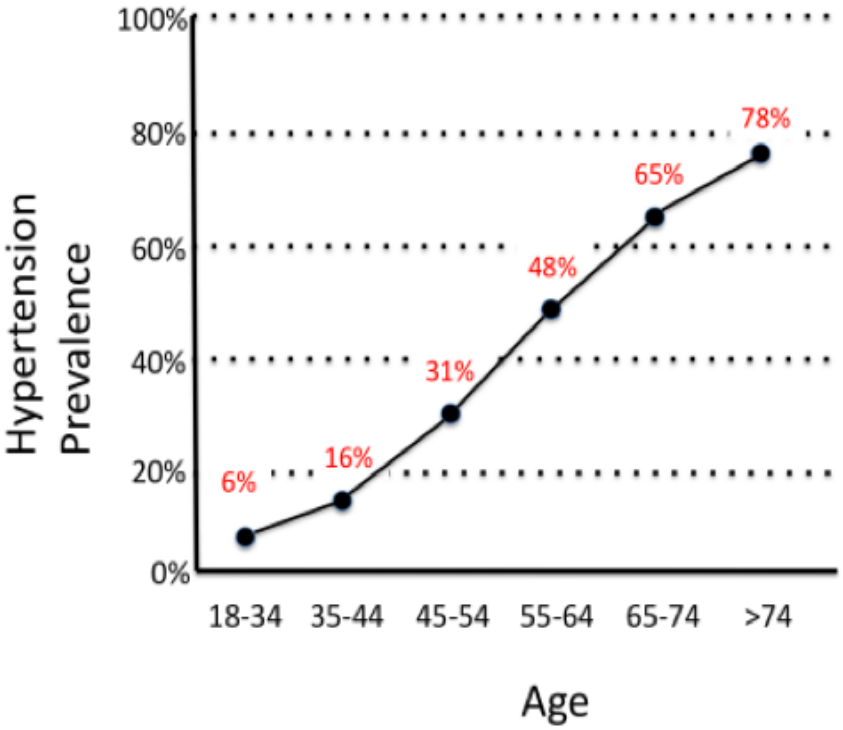


Leptin knockout

- Phenotypes for actionable variants: penetrance estimates, *phenotypic variability*

Of horses and zebras

I. Age-dependent prevalence of hypertension



Typical definition: >140/90

New definition: >130/80

"Monogenic" forms of hypertension: estimated 3-5% of all cases. Genetics first approach to identify <u>early cases</u>

II. Primary Immunodeficiencies: under-diagnosed diseases?

Infections are very common—viral, bacterial

Likely continuum of genetic differences in the capacity to fight infections—major evolutionary selection

Crucial time element: decreased immune capacities at age extremes (newborns, elderly)

For consideration: protocols to alert health care providers re risk factors for immunodeficiences—e.g.

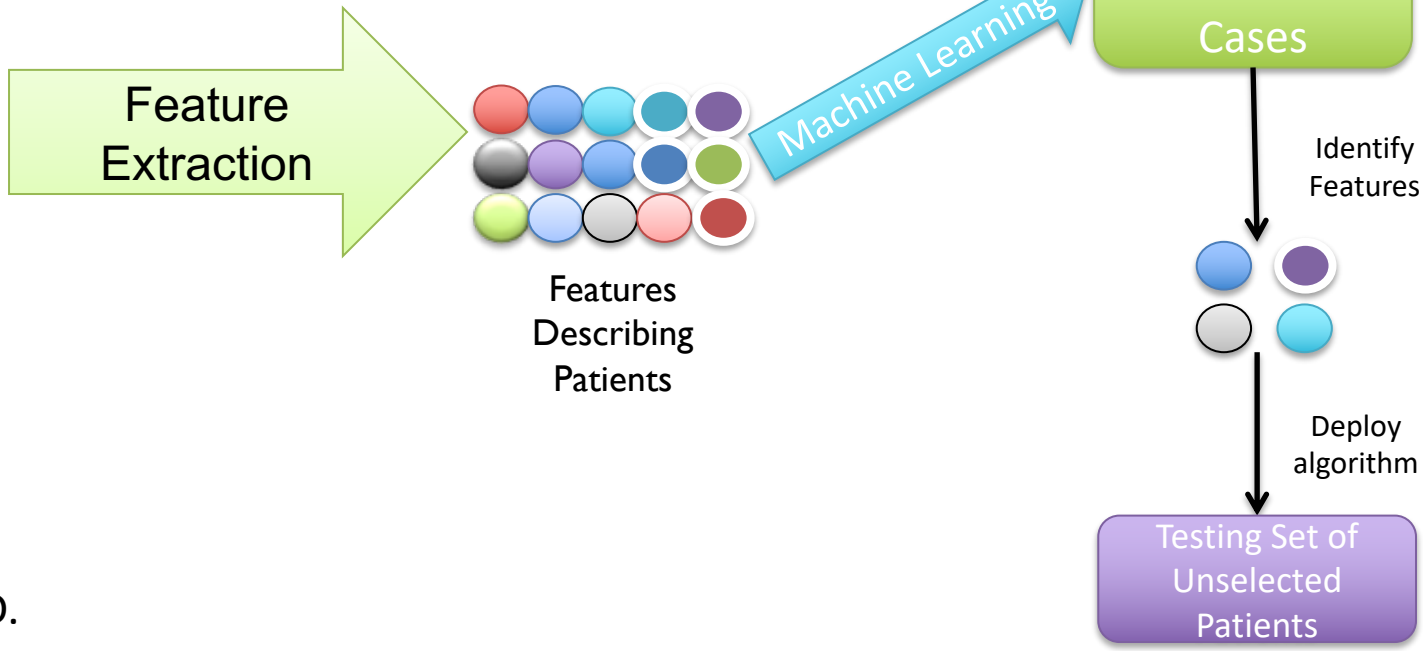
- *Two major infections*
- *In a non-alcoholic*
- *Below the age of 50*

Chronic granulomatous diseases: NADPH oxidases, infections and increased risk for IBD

III. Scaling rare disease discovery within health systems through machine learning

- ~ 6,000 rare diseases that are known
- Most of these diseases are not diagnosed precluding appropriate therapy

Girish Nakardni, M.D.



This will enable us to set up automated disease surveillance programs so that rare diseases are promptly diagnosed

“Expected” & “Unexpected” Pleiotropy

IL23R is an IBD (Crohn's disease & ulcerative colitis) gene

- Crohn's disease (CD) *and* ulcerative colitis (UC)
- Multiple independent alleles, notably a *protective,* LOF minor allele Arg381Gln (2-3 fold protection)
- IL-23 immune pathway: key role in Th17 cells, Tc17, ILC

Majority of IBD loci show similar trends between CD vs. UC

Science 2006; 314: 1461-3

Across autoimmunity: a) MHC class I vs. II; b) PTPN22 Arg620Trp, and c) IL23R

Table 1 Major genetic association signals across autoimmune diseases

	MHC class	IL23R	PTPN22	CTLA4[a]
Type 1 diabetes	Class II		Arg620**Trp**	Non-coding
Juvenile idiopathic arthritis	Class II		Arg620**Trp**	
Autoimmune thyroid disease	Class II		Arg620**Trp**	Non-coding
Rheumatoid arthritis	Class II		Arg620**Trp**	Non-coding
Multiple sclerosis	Class II			
Celiac disease	Class II			Non-coding
Systemic lupus erythematosis	Class II		Arg620**Trp**	
Psoriatic arthritis	Class I	Distinct alleles		
Psoriasis	Class I	Arg381**Gln**		
Ankylosing spondylitis	Class I	Arg381**Gln**		
Inflammatory bowel disease	Class II	Arg381**Gln**	**Arg**620Trp	

Underlined and bolded codons represent the allele associated with disease. **[AU: what do colors mean?]**

[a]Non-coding variants associated with the *CTLA4* region may be distinct between diseases.

IL23R-associated diseases respond to blocking IL-23 pathway

Cho and Feldmann, *Nature Medicine* 2015; 730

Cystic fibrosis: poly-organ recessive Mendelian disease—chloride channel defect

Organ manifestations

1. Lung: frequent infections
2. Male infertility
3. Frequent pancreatic *insufficiency* requiring enzyme replacement

Heterozygous CFTR carriers present with recurrent pancreatitis phenotypically distinct from CF patient manifestations

Large, collaborative biobank studies: Systematic evaluation of heterozygous carriers of (AR) Mendelian diseases

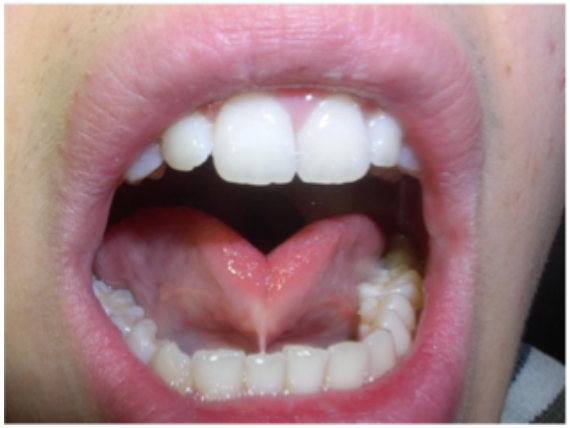
Mendelian diseases co-segregating with IBD

Table 2:

Relative Risk and Bonferroni Corrected *P*-values for Significant ($P < 0.05$) Mendelian Diseases Associated with CD and UC.[a]

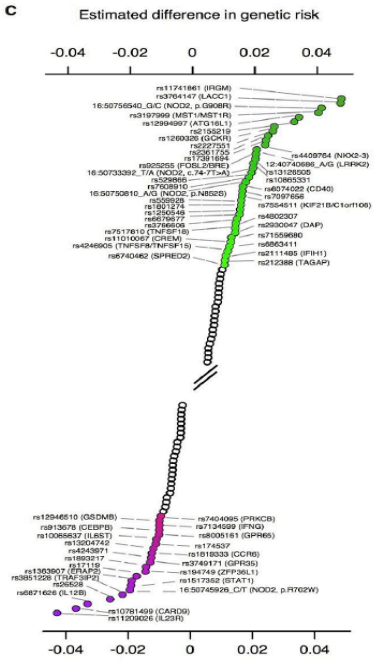
Mendelian Disease	Cases No.	CD RR	CD *P*-Value	UC RR	UC *P*-Value
Disorders of Phosphorous Metabolism	176,087	6.36	0.00E+00	6.42	0.00E+00
Long QT Syndrome	37,916	3.62	4.08E-112	3.68	1.97E-112
Haemophilia	53,855	3.47	6.48E-145	3.03	7.27E-100
Disorders of Urea Cycle Metabolism	18,042	4.04	3.36E-68	4.26	9.18E-74
Genetic Anomalies of Leukocytes	3,917	6.66	1.47E-39	6.83	6.20E-40
Tongue Tie	59,368	0.14	1.98E-49	0.2	4.04E-39
Lipoprotein Deficiencies	83,695	1.9	1.44E-38	2.03	2.93E-47
Disorders of Copper Metabolism	5911	7.25	6.53E-70	5.21	3.17E-36
Thalassemia	46,956	2.19	1.25E-35	2.29	1.24E-39
Chronic Granulomatous Disease	8058	4.13	9.31E-32	3.94	1.47E-27

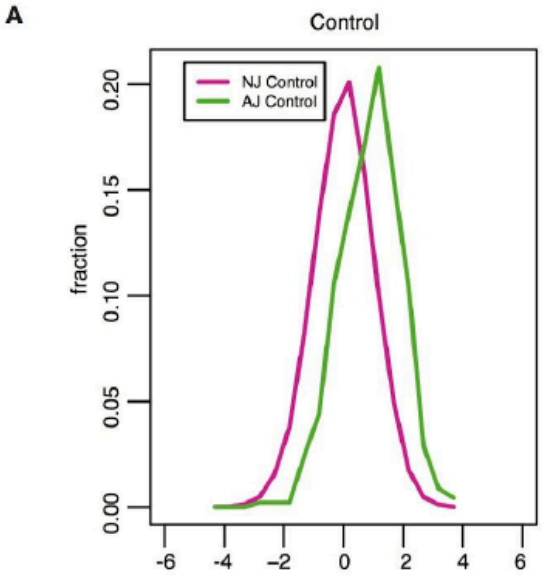
67 million U.S. patients EHR

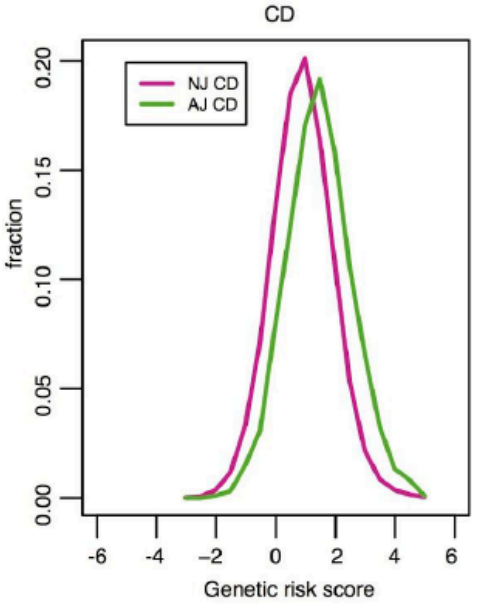


Han et al., *IBD Journal* 2018, 471

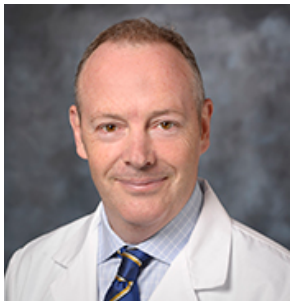
Ashkenazi Jews with a three-fold higher *prevalence* of IBD vs. non-Jewish Europeans







Manny Rivas



Dermot McGovern

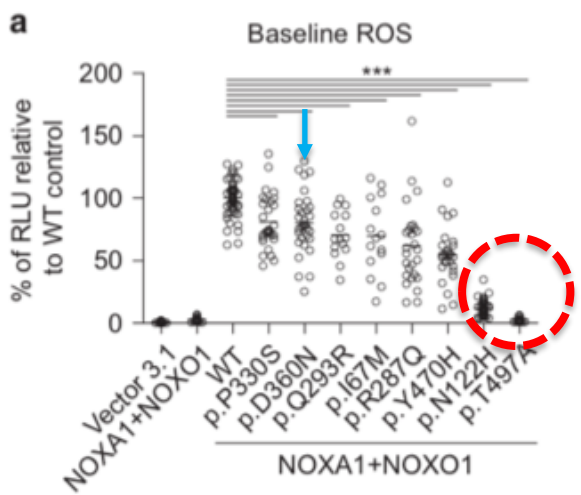
Mark Daly

In press, Plos Genetics

NOX1 (NADPH oxidase 1) D360N: present in both AJs and non-AJ EA, higher effect sizes in AJs

	Case allele freq	Control allele freq	Odds ratio	P-value
All AJ UC	2.2%	0.9%	2.44	0.005

- NOX1 is X-linked: Male > female effects



Holm Uhlig

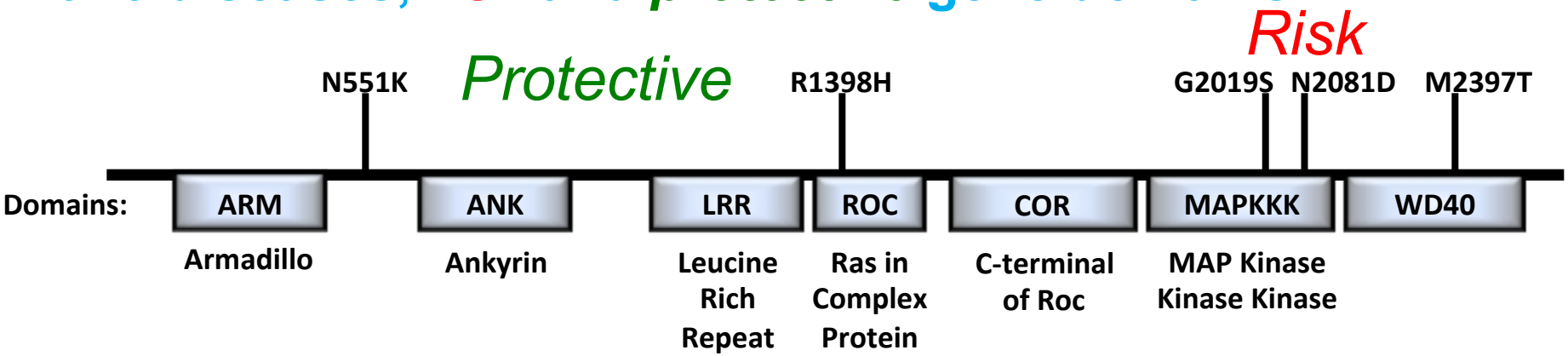
Hayes et al., CMGH 2015; 489

Very early onset mutation: identifies NOX1 variants with much greater functional effects than *D360N*

Schwerd et al., Mucosal Immunology 2018; 562

“Unexpected” Pleiotropy

Jewish-predominant mutations: One gene (LRRK2), two diseases, *risk* and *protective* gene domains



Crohn's disease

1. Intestine
2. Disease onset: 15-30
3. Episodic inflammation

Parkinson's disease

1. Brain
2. Late onset: 50's & above
3. Inexorable degeneration

LRRK2 mutations: alterations in tissue-based phagocytes

Hui et al., Sci Trans Med, 2018

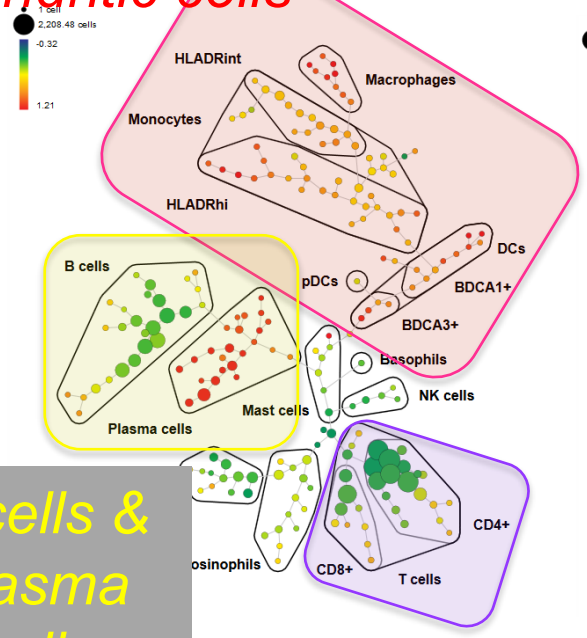
Inga Peter

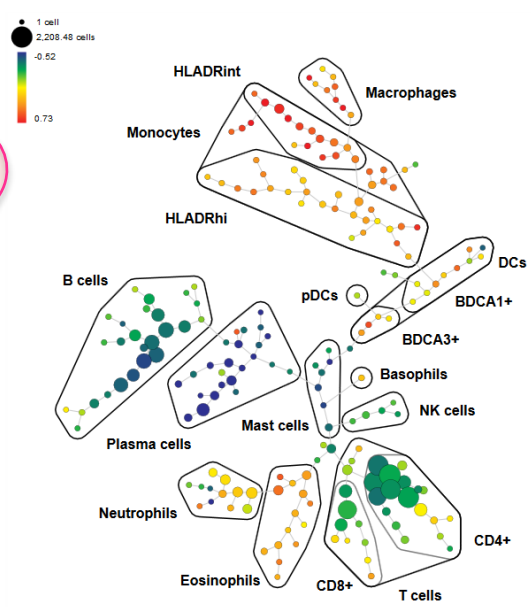
Uncommon risk alleles of innate immunity: Cytof terminal ileal biopsies from resected tissues

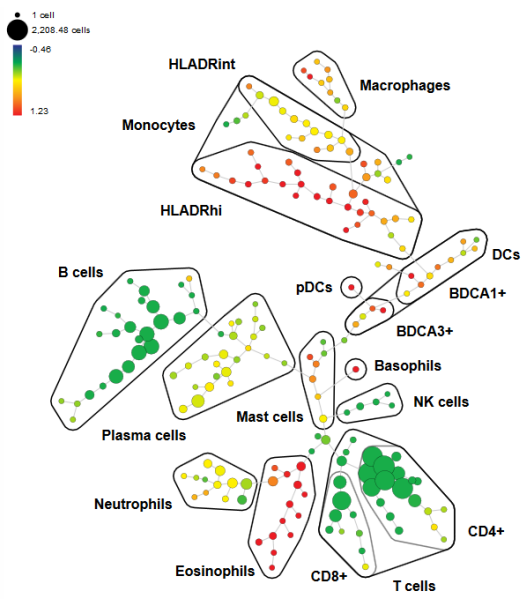
Monocytes, macrophages & dendritic cells

B cells & plasma cells

T cells







At-risk populations with higher disease prevalences

Higher polygenic/all common variant risk scores: different effect sizes for uncommon alleles?

Disease prevalences estimates across populations

- challenges of Hispanic definitions.
- Sample size limitations & under-study of non-European populations

U.S. populations & CDC-based estimates of disease prevalence

"Monogenic" genes in complex traits: Conclusions & next steps---*continua* of allele, gene effects

1. Gene-centric view of disease pathogenesis

2. Systematic analysis of biobank-based genetic data will provide specific biologic/medical context to phenotypic variability of monogenic/high-effect genetic variants

3. Population differences: rare/uncommon variant differences & disease prevalence. *Domain-based* sequence annotation

4. Emerging recognition of the ubiquity of pleiotropy: defining causality & age/time-based factors

5. "Phenome coverage more limiting than genome coverage"–Visscher & Yang 2016, Nature Genetics, 707