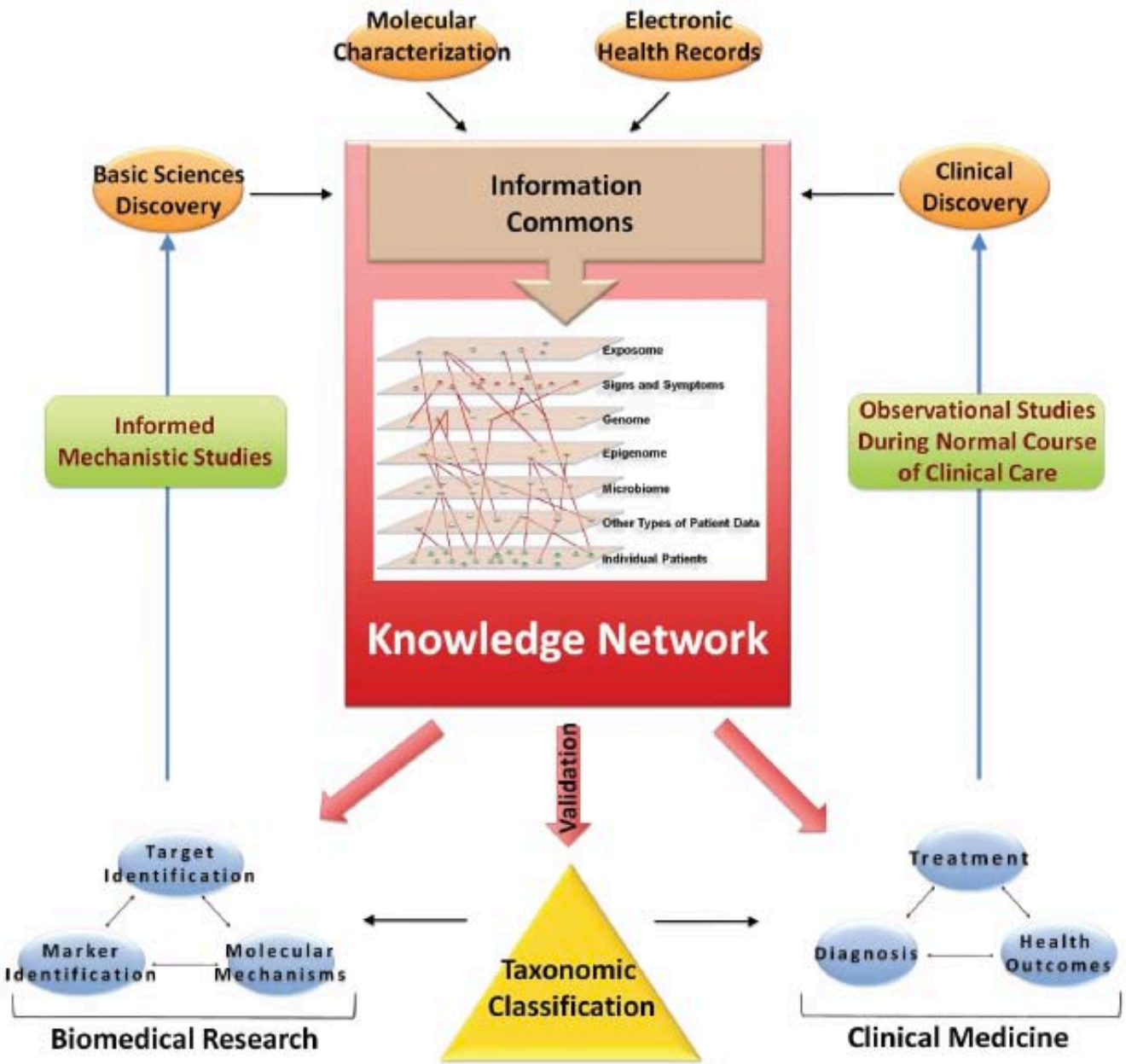
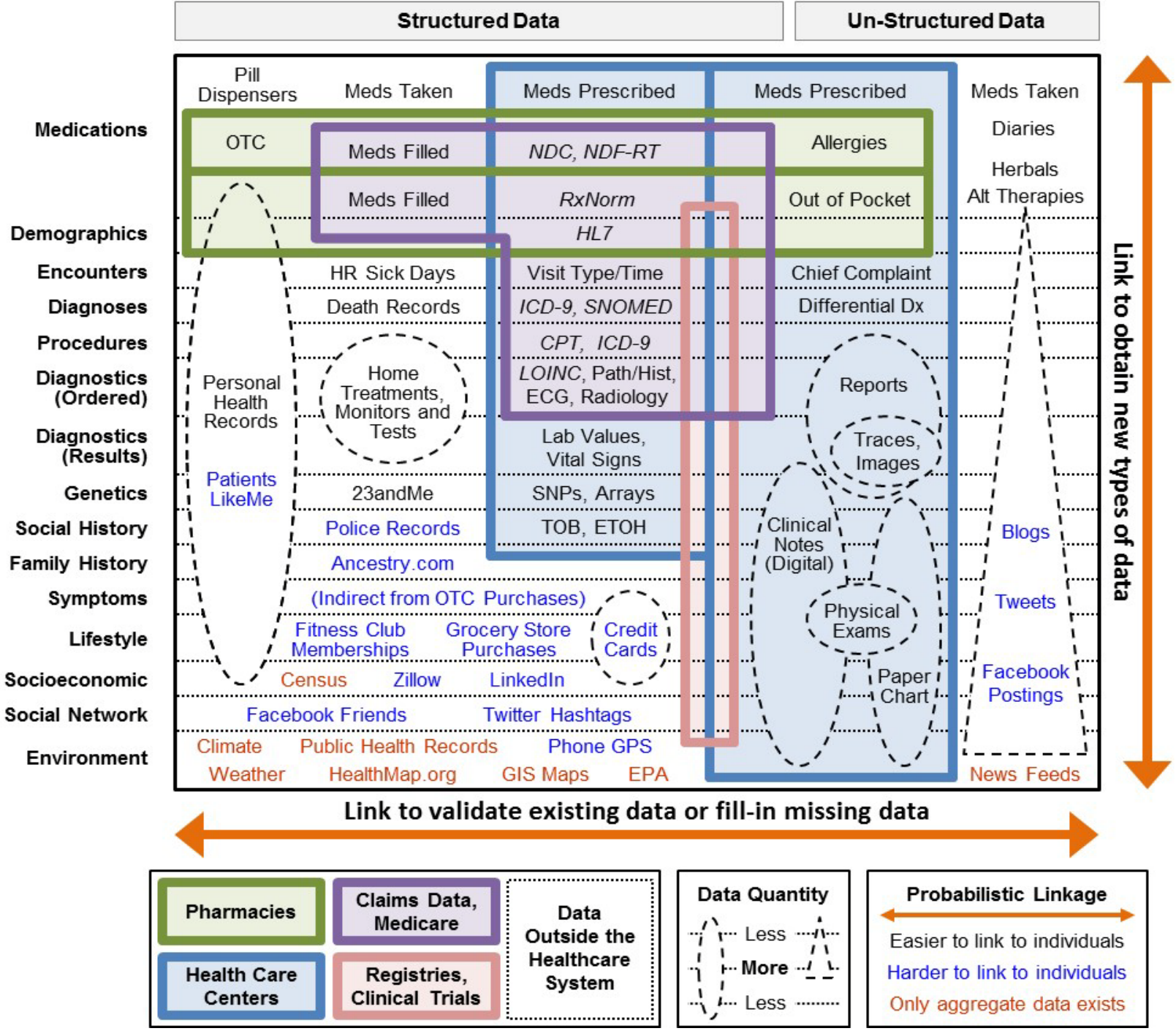
Autism and Autisms.

Isaac S. Kohane, MD, PhD

cbmi.med.harvard.edu

HARVARD MEDICAL SCHOOL


Molecular Characterization
Electronic Health Records
Basic Sciences Discovery
Information Commons
Clinical Discovery
Exposome
Signs and Symptoms
Genome
Epigenome
Microbiome
Other Types of Patient Data
Individual Patients
Knowledge Network
Informed Mechanistic Studies
Observational Studies During Normal Course of Clinical Care
Validation
Target Identification
Marker Identification
Molecular Mechanisms
Biomedical Research
Taxonomic Classification
Treatment
Diagnosis
Health Outcomes
Clinical Medicine


Structured Data
Un-Structured Data
Medications
Demographics
Encounters
Diagnoses
Procedures
Diagnostics (Ordered)
Diagnostics (Results)
Genetics
Social History
Family History
Symptoms
Lifestyle
Socioeconomic
Social Network
Environment
Pill Dispensers
Meds Taken
Meds Prescribed
Meds Prescribed
Meds Taken
Diaries
OTC
Meds Filled
NDC, NDF-RT
Allergies
Herbals
Alt Therapies
Meds Filled
RxNorm
Out of Pocket
HL7
HR Sick Days
Visit Type/Time
Chief Complaint
Death Records
ICD-9, SNOMED
Differential Dx
CPT, ICD-9
Personal Health Records
Home Treatments, Monitors and Tests
LOINC, Path/Hist, ECG, Radiology
Reports
Lab Values, Vital Signs
Traces, Images
Patients LikeMe
23andMe
SNPs, Arrays
Clinical Notes (Digital)
Blogs
Police Records
TOB, ETOH
Ancestry.com
(Indirect from OTC Purchases)
Tweets
Physical Exams
Fitness Club Memberships
Grocery Store Purchases
Credit Cards
Census
Zillow
LinkedIn
Paper Chart
Facebook Postings
Facebook Friends
Twitter Hashtags
Climate
Public Health Records
Phone GPS
Weather
HealthMap.org
GIS Maps
EPA
News Feeds
Link to obtain new types of data
Link to validate existing data or fill-in missing data
Pharmacies
Claims Data, Medicare
Health Care Centers
Registries, Clinical Trials
Data Outside the Healthcare System
Data Quantity
Less
More
Less
Probabilistic Linkage
Easier to link to individuals
Harder to link to individuals
Only aggregate data exists

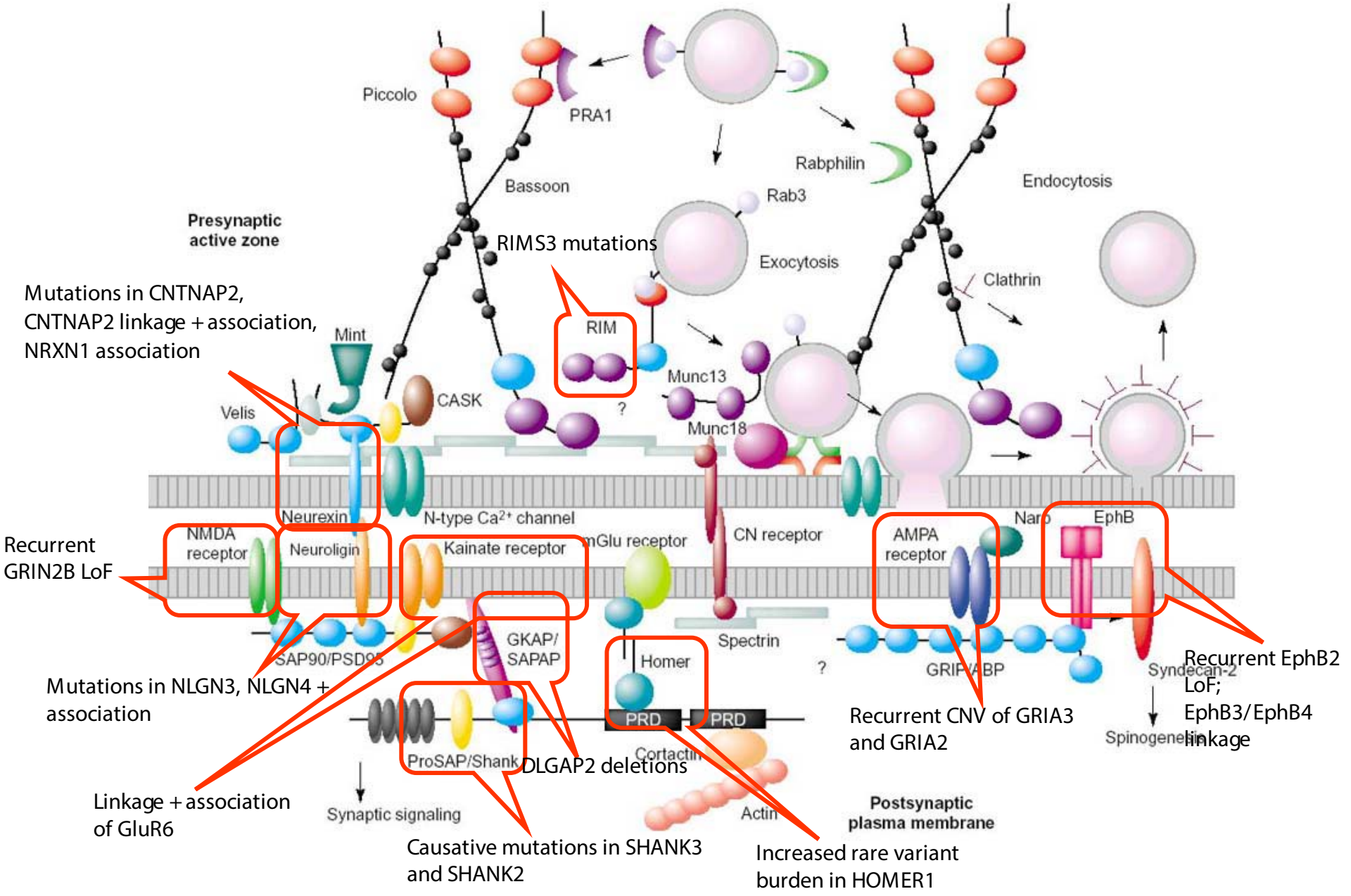
So what does precision medicine mean?

Probability of a disease *D*

Given the findings *F*

$$p(D \mid F)$$

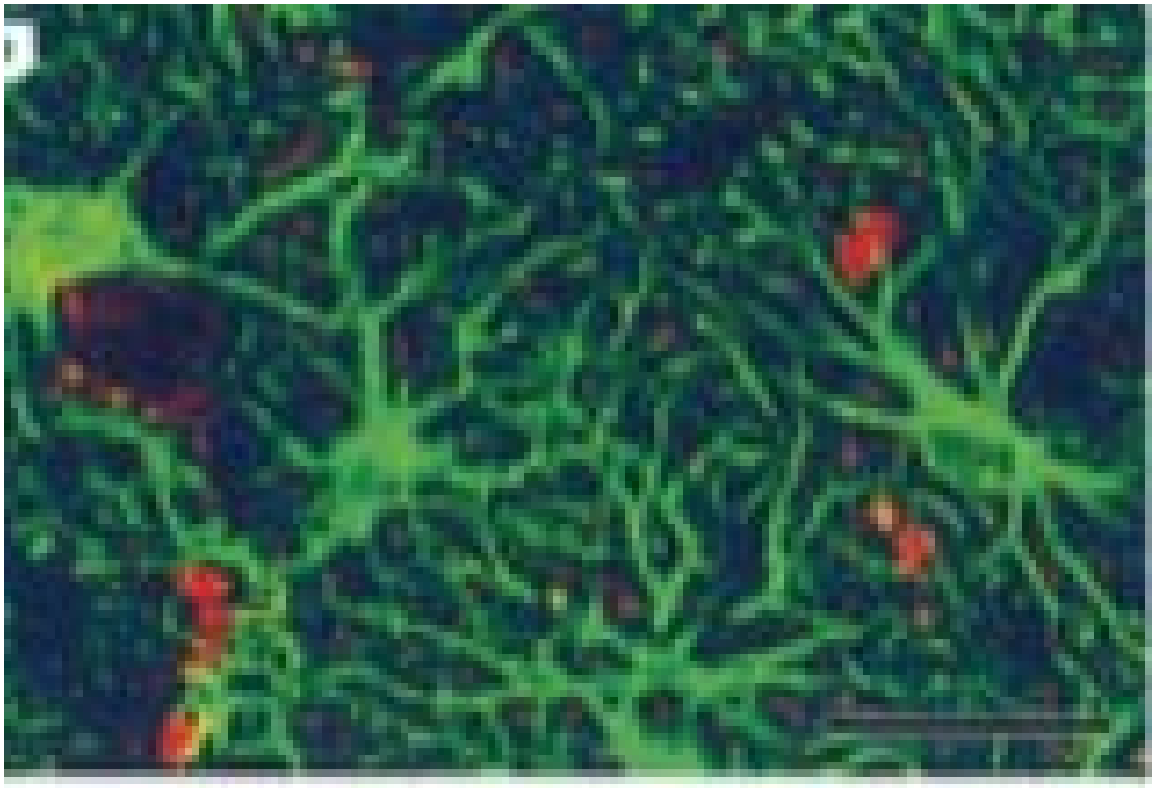
Convergence on synaptic function



Pathways from Predictor

KEGG pathways	Count	%	P-Value	FDR	Genes
Neurotrophin signaling pathway	11	4.5	0.000018	0.0018	CRK, CRKL, KIDINS220, MAP2K1, MAP3K5, MAPK8, PIK3CB, PRKCD, RPS6KA3, SH2B3, YWHAG
Fc gamma R-mediated phagocytosis	7	2.9	0.0034	0.16	CRK, CRKL, DOCK2, MAP2K1, PIK3CB, PRKCD, PTPRC
Focal adhesion	10	4.1	0.0037	0.12	ACTN1, CRK, CRKL, IQGAP2, ITGB2, MAP2K1, PDGFC, PIK3CB, PPP1R12A, ROCK1
Renal cell carcinoma	6	2.5	0.0044	0.11	CREBBP, CRK, CRKL, EGLN1, MAP2K1, PIK3CB
Regulation of actin cytoskeleton	10	4.1	0.0057	0.11	ACTN1, CRK, CRKL, IGF1R, MAP2K1, MAPK8, PDGFC, PIK3CB, PPP1R12A, ROCK1
Vascular smooth muscle contraction	7	2.9	0.0075	0.12	GNAQ, GUCY1B3, MAP2K1, PPP1R12A, PPP1R12B, PRKCD, ROCK1
Chemokine signaling pathway	9	3.7	0.008	0.11	CCR2, CRK, CRKL, DOCK2, JAK2, MAP2K1, PIK3CB, PRKCD, ROCK1
Long-term potentiation	5	2.1	0.021	0.23	CREBBP, GNAQ, MAP2K1, PPP1R12A, RPS6KA3
Chronic myeloid leukemia	5	2.1	0.029	0.28	CRK, CRKL, CTBP2, MAP2K1, PIK3CB
Fc epsilon RI signaling pathway	5	2.1	0.032	0.28	LCP2, MAP2K1, MAPK8, PIK3CB, PRKCD
Notch signaling pathway	4	1.6	0.036	0.29	CREBBP, CTBP2, MAML3, NOTCH1
Type II diabetes mellitus	4	1.6	0.036	0.29	HK2P1, MAPK8, PIK3CB, PRKCD
Progesterone-mediated oocyte maturation	5	2.1	0.044	0.31	IGF1R, MAP2K1, MAPK8, PIK3CB, RPS6KA3

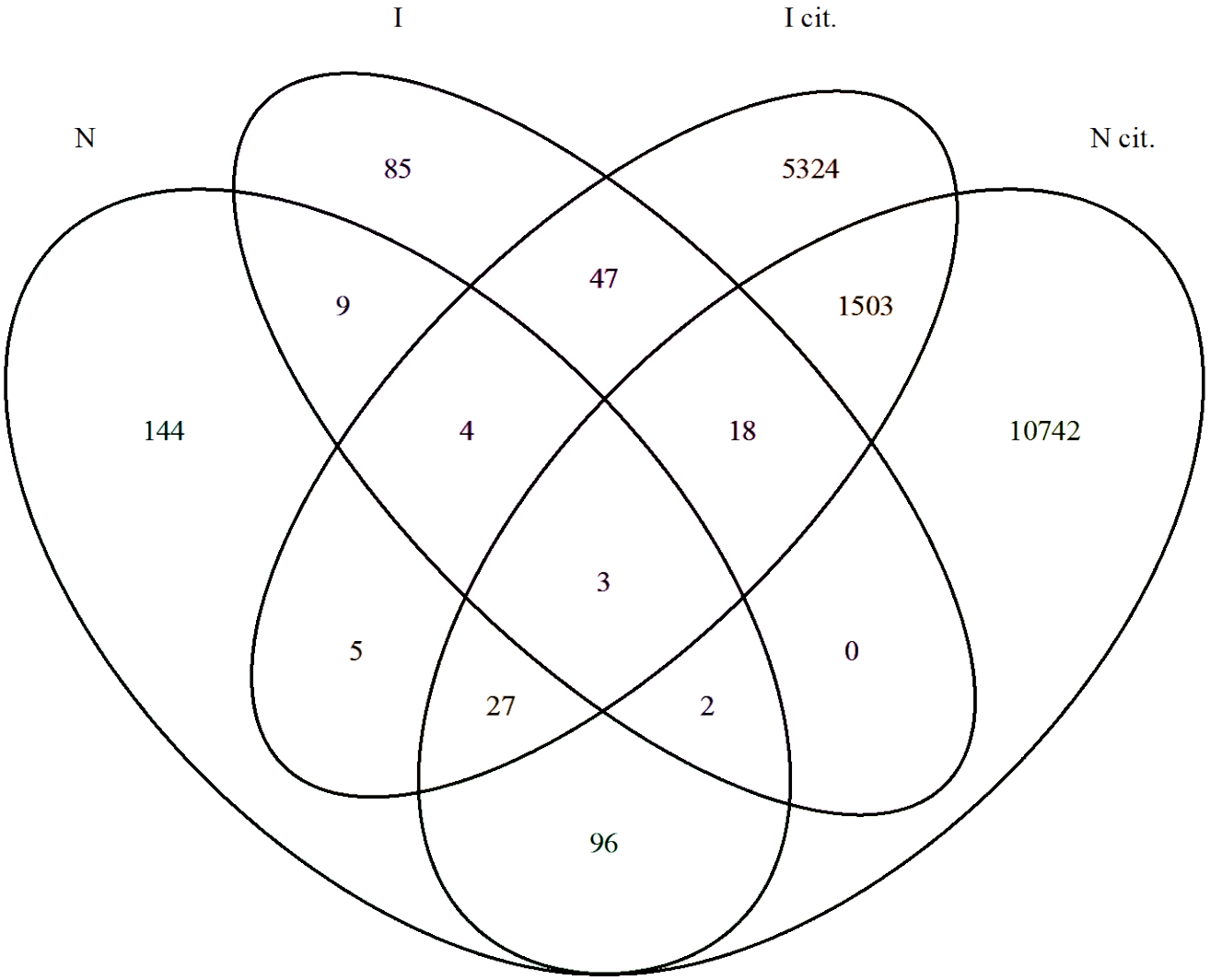
Vargas *et al. 2005*



And some interesting observations

- Mother's with RA and Father's with T1DM
- Mother's with high gestational CRP
- Intra- and peri-partum infection
- Mouse models of inflammation

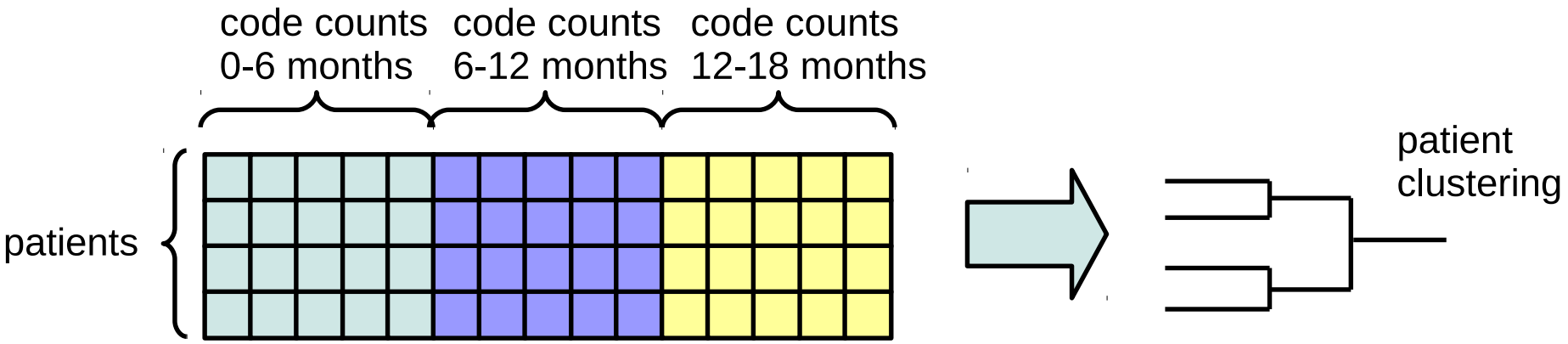
(Immune vs Synapse)+Gene



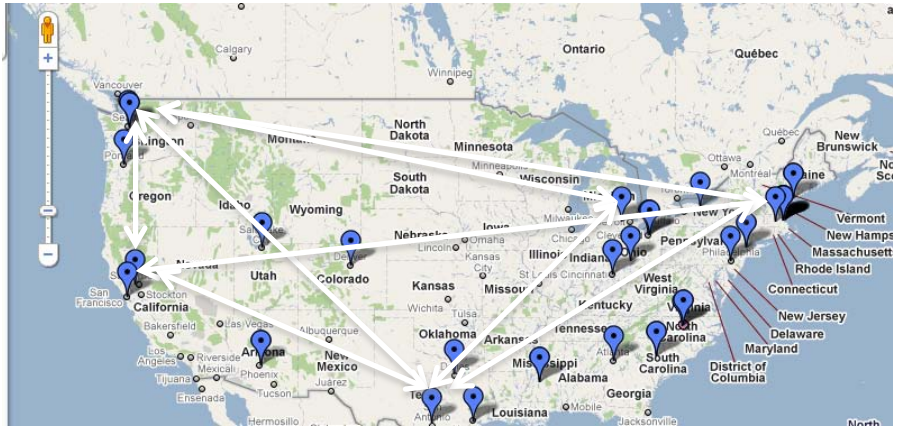
Across AHC's at a glance

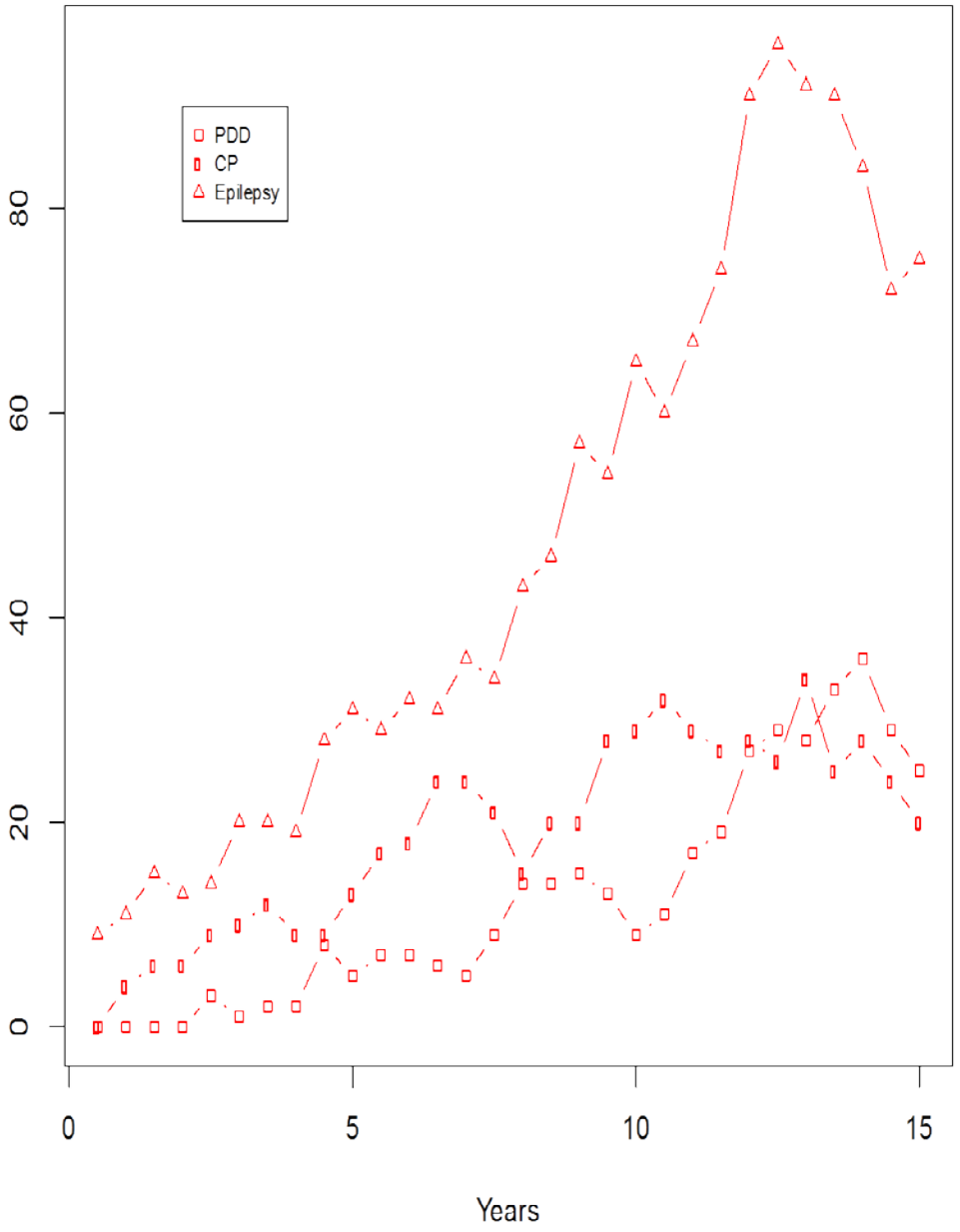
- Patients: 13750 (with basic demographics)
- ~0.5% hospital population
- M:F (5:1)
- Diagnoses: 5627
- Laboratory measurements: 3,158,234 on 3581 lab measurement types
- Medications: > 800,000 Rx's

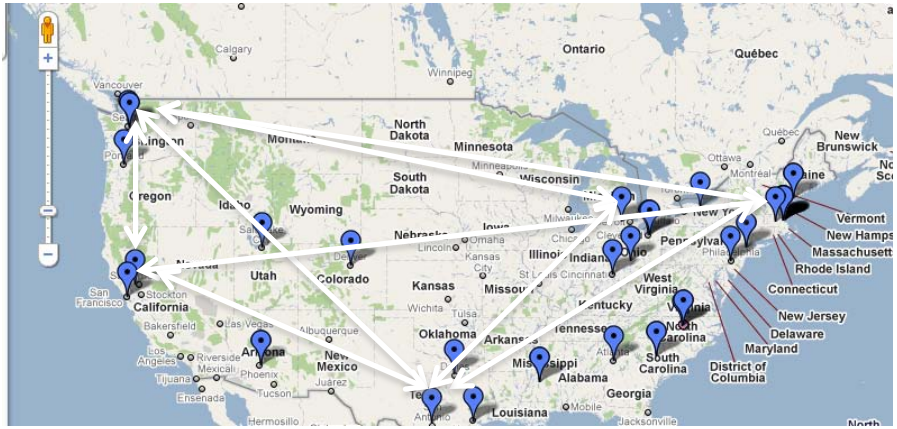
Unbiased clustering

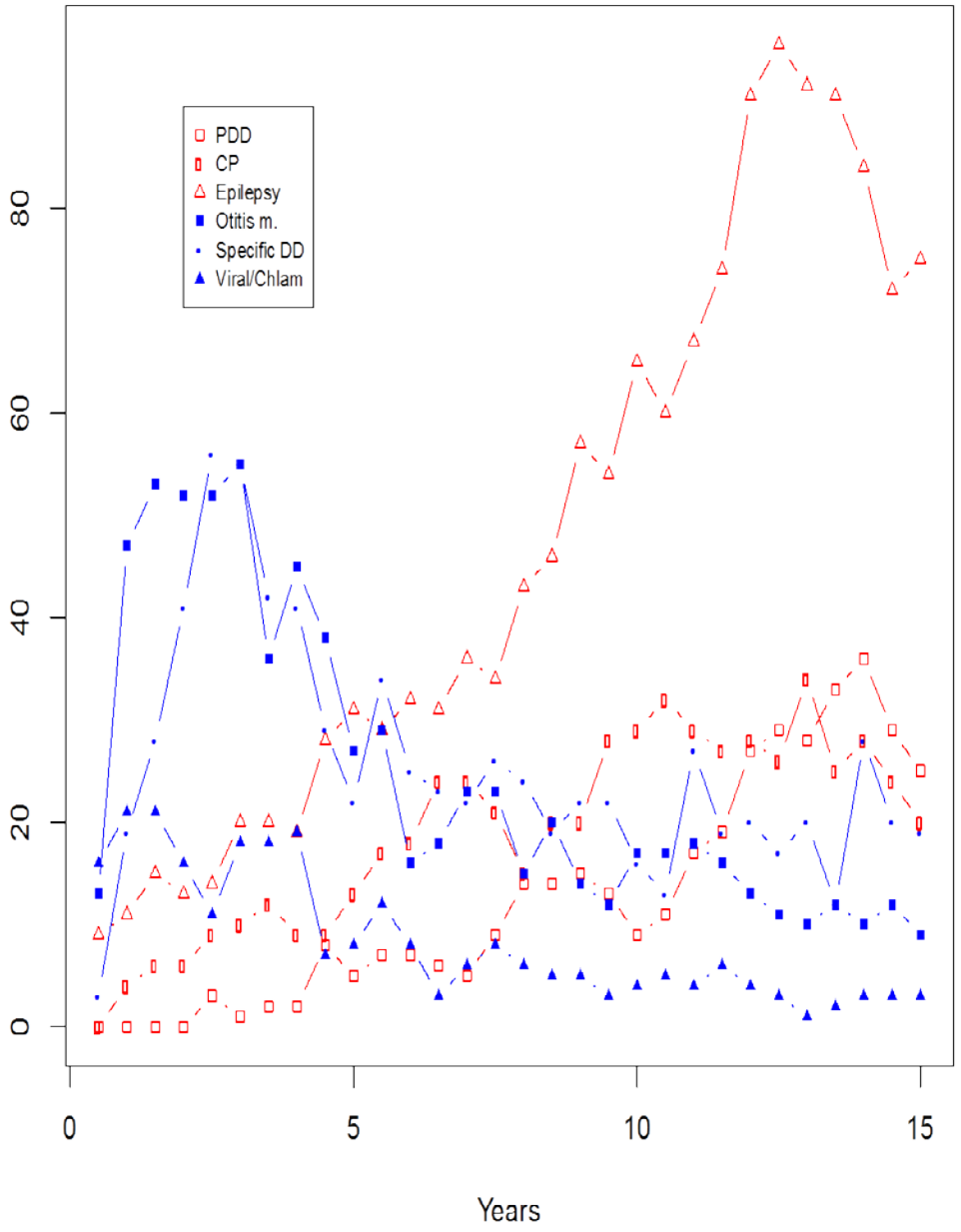


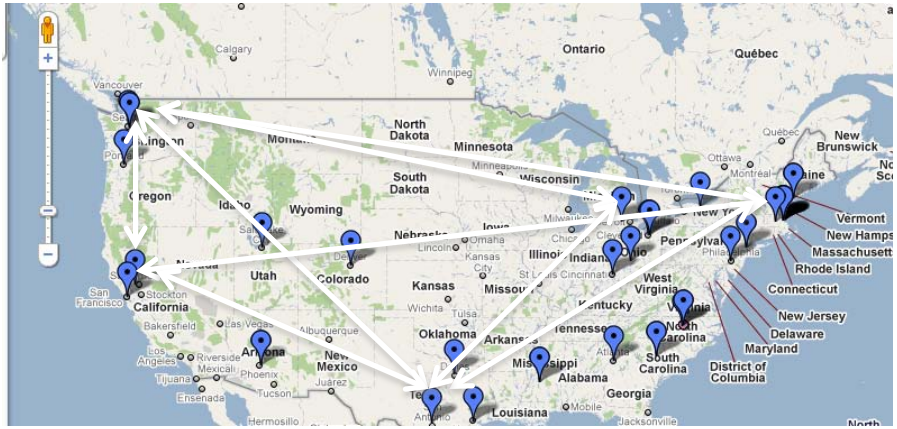
Autism or Autisms?

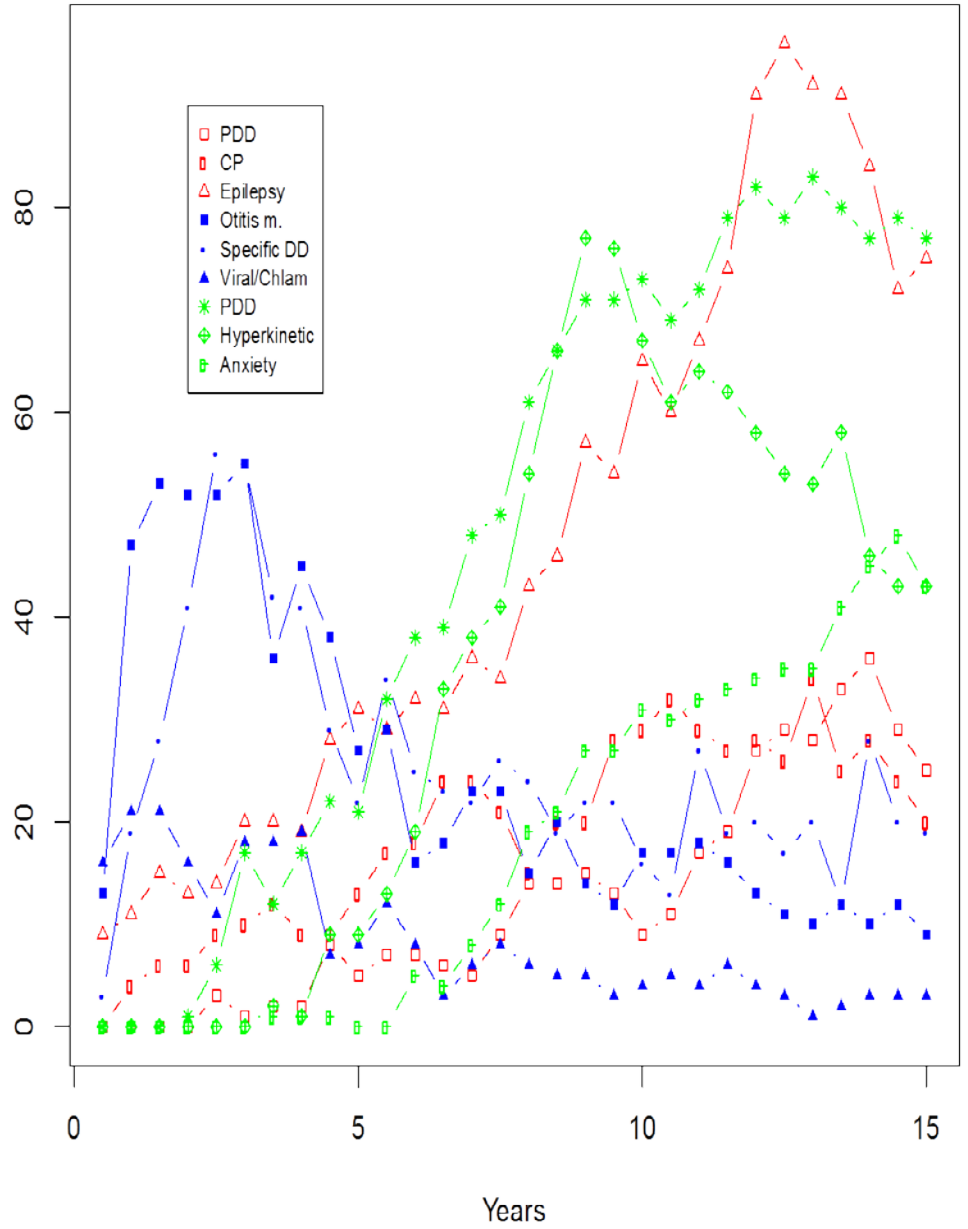




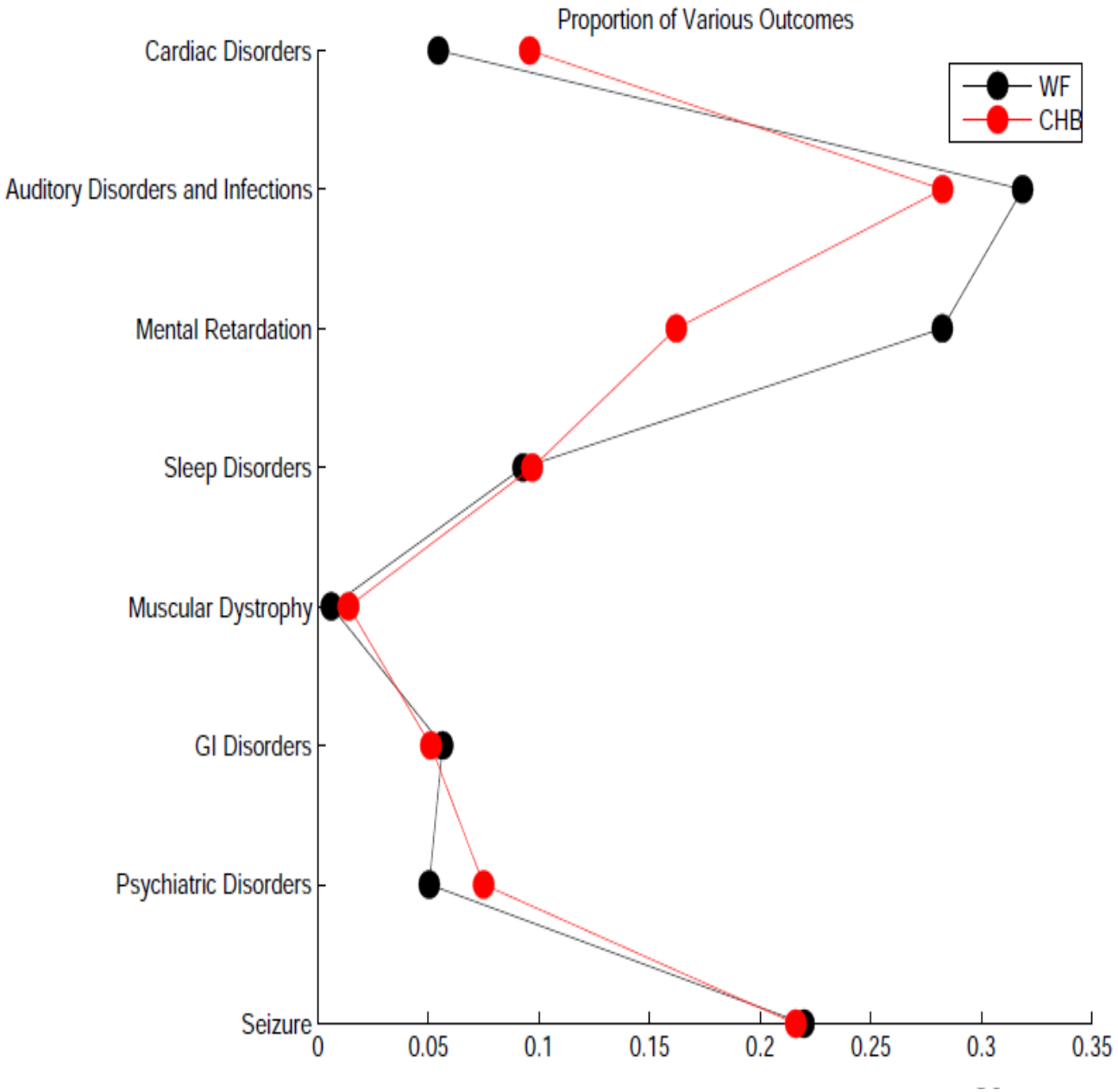

Calgary
Ontario
Québec
Winnipeg
Vancouver
North Dakota
Minnesota
South Dakota
Wisconsin
Oregon
Wyoming
Nebraska
Utah
Colorado
Kansas
Missouri
Illinois
Indiana
West Virginia
Kentucky
Oklahoma
Arkansas
Louisiana
Alabama
Georgia
South Carolina
North Carolina
Virginia
Maryland
Delaware
New Jersey
Connecticut
Rhode Island
Massachusetts
Vermont
New Brunswick
District of Columbia
California
San Francisco
Stockton
Bakersfield
Las Vegas
Albuquerque
Arizona
New Mexico
Phoenix
Tucson
Tijuana
Ensenada
Juárez
Hermosillo
Mobile
Jacksonville
Ottawa
Montréal


PDD
CP
Epilepsy
Otitis m.
Specific DD
Viral/Chlam
0
20
40
60
80
0
5
10
15
Years


Calgary
Ontario
Québec
Winnipeg
Vancouver
North Dakota
Minnesota
South Dakota
Wisconsin
Oregon
Wyoming
Nebraska
Utah
Colorado
Kansas
Missouri
Illinois
Indiana
West Virginia
Kentucky
Oklahoma
Arkansas
Mississippi
Alabama
Georgia
Louisiana
South Carolina
North Carolina
Virginia
Maryland
Delaware
New Jersey
Connecticut
Rhode Island
Massachusetts
Vermont
New Brunswick
California
Arizona
New Mexico
District of Columbia


PDD
CP
Epilepsy
Otitis m.
Specific DD
Viral/Chlam
PDD
Hyperkinetic
Anxiety
0
20
40
60
80
0
5
10
15
Years

Reproducibility

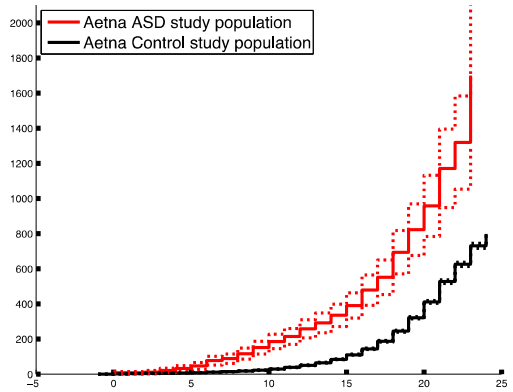


Thought experiment

- Autism, a common disease with prevalence ~1%
- Let's say causal variants ~1% frequency, RR = 2
- Then with 80% power (5% alpha) ~23,000
- What if ASD is really 10+ diseases?
  - Many reasons why it will not be clinically noticed.
- Then causal variants will have 10% frequency and with same power/alpha ~2300 subjects
- But until now phenotype-first was unaffordable.

Validating IBD Findings

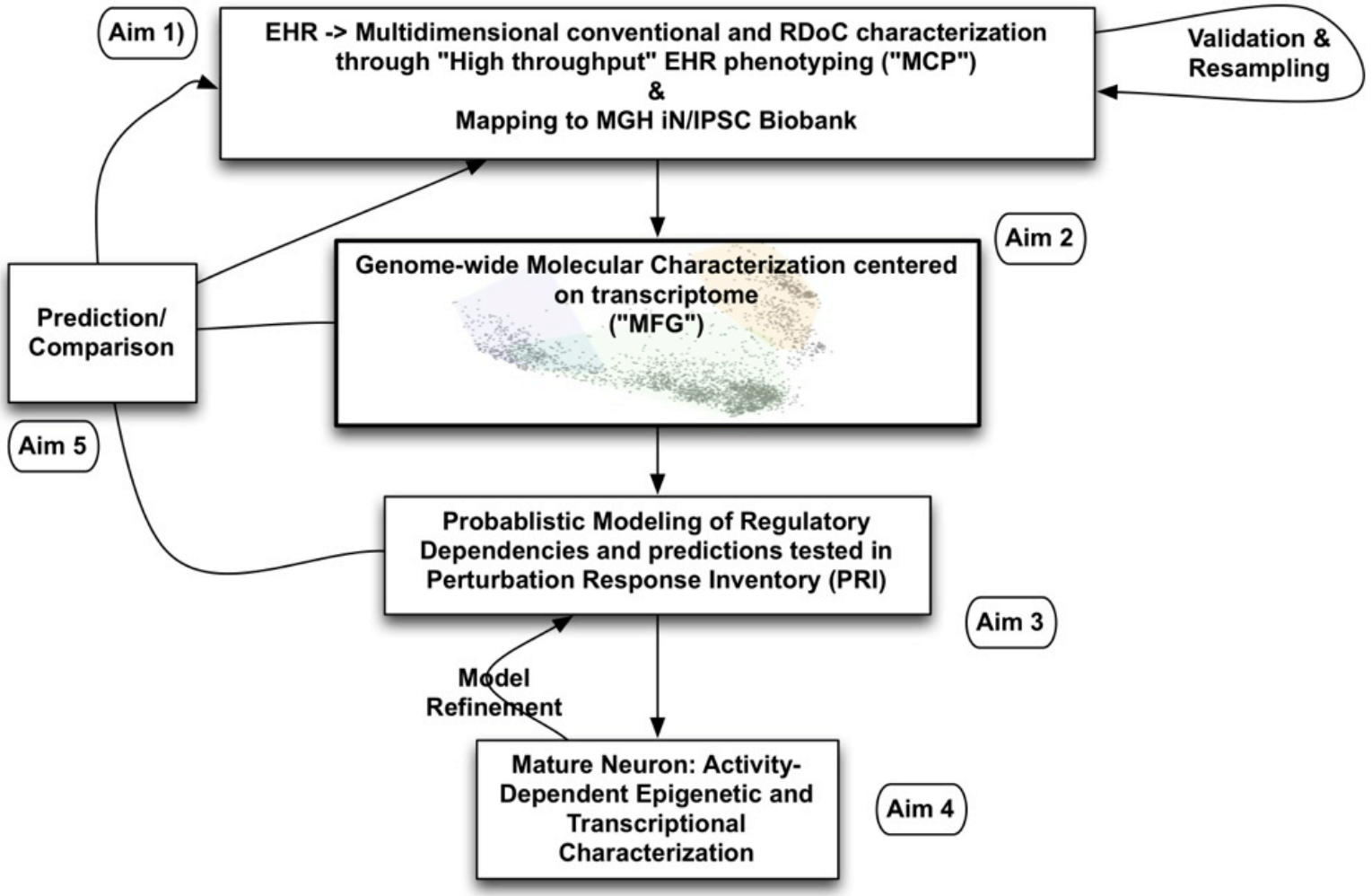
Study Population	Control	ASD
Aetna	186.4 (183.3-189.5)	446.0 (376.8-529.8)
BCH	433.4 (417.5-449.9)	630.6 (462.1-859.9)
WFBMC	282.9 (260.7-306.7)	361.9 (140.0-1824.8)



Summary

- The conventional wisdom regarding the causes of autism is incomplete, divided and obscured.
- Phenotype-first strategies may massively accelerate discovery of genetic architecture.
- There is a lot of shared pathobiology across autism
- There is a lot of undiscovered heterogeneity and distinctive pathobiology within conventionally labeled diseases.
- Aggressively ecumenical approach to integrative data analysis will accelerate discovery.

Neuropsychiatric Genome-Scale and RDoC Individualized Domains (N-GRID)



Thank you

Psychiatric Genetics
Roy Perlis
Jordan Smoller

Biomedical Informatics
Nathan Palmer
Sek Won Kong
Alal Eran
Chirag Patel
Arjun Manrai
Andrew Beam
Susanne Churchill
Stan Shaw
Peter Szolovits
Tianxi Cai
Finale Doshi
Daria Prilutsky
David Margulies

Developmental Medicine
Lenny Rappaport
Chuck Nelson
Stephanie Brewster

Genomics
Lou Kunkel