UNIVERSITY OF UTAH
HEALTH SCIENCES

Family Studies and Complex Disease

Lynn Jorde

Mark and Kathie Miller Presidential Professor and Chair

Department of Human Genetics

May 1, 2018

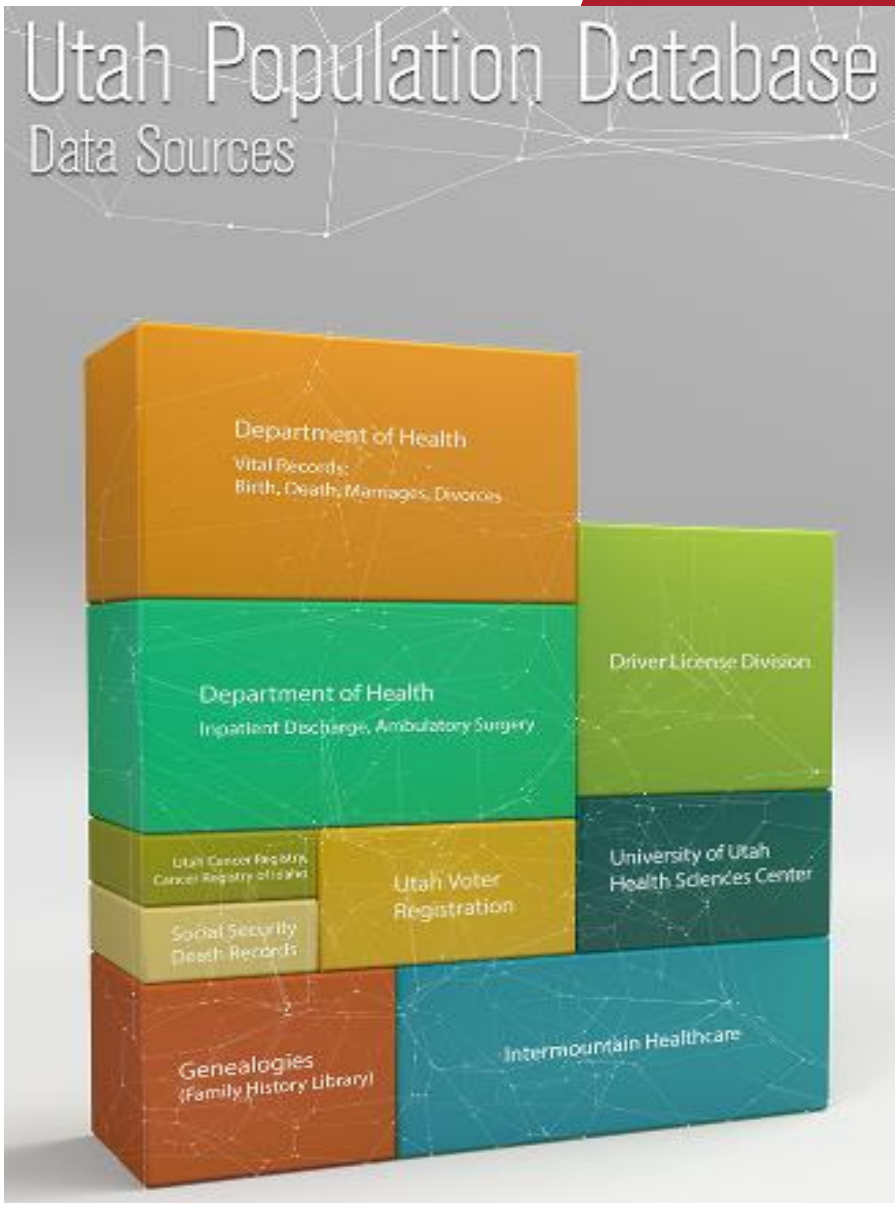
Family studies and complex disease: Applications

- Detection of causal *de novo* mutations
  - CNVs: e.g., autism, schizophrenia
  - SNVs: e.g., intellectual disability (Gilssen *et al.*, 2014, *Nature*), congenital heart disease (Jin *et al*., 2017, *Nat. Genet.*)
- Parent-of-origin effects
  - e.g., paternal CNV bias for autism (Brandler *et al*., 2018, *Science*)
- Germline and somatic mosaicism
- Detection of Mendelian subsets with complex disease phenotypes (*BRCA1, BRCA2, APC, etc.*)
- Multigenerational pedigrees allow detection of shared genomic segments containing rare disease-causing variants
- Families provide opportunity for long-term, longitudinal studies and return of clinically significant results

A Utah nuclear family

Utah Population Database (UPDB)

- University of Utah resource to support medical research, initiated in the 1970s
- 10 million people in large, multigenerational pedigrees are linked to more than 25 million medically relevant records
- Geocode information available
- Now expanded to 100 million people



Utah Genome Project (UGP): >7,000 cases sequenced; >50 diseases

WGS/WES + large pedigrees from 10-million-member Utah Population Database

- Amyotrophic lateral sclerosis
- Genetics of extreme thinness to develop obesity interventions
- Spontaneous preterm birth
- Familial childhood cancers (Ewing sarcoma, Wilms tumor, germ cell tumors)
- DNA repair genes and cancer
- Hematologic cancers: CLL, CML, multiple myeloma
- Common cancers: breast, prostate, colorectal
- Psoriasis and psoriatic arthritis
- Juvenile idiopathic arthritis
- Crohn disease
- Osteoarthritis
- Idiopathic pulmonary fibrosis
- Familial cardiac arrhythmia
- Autism
- Suicide
- Primary ovarian insufficiency
- WGS of Utah CEPH pedigrees (n= ~600)

Evaluating *de novo* mutations (DNMs) using Utah CEPH pedigrees

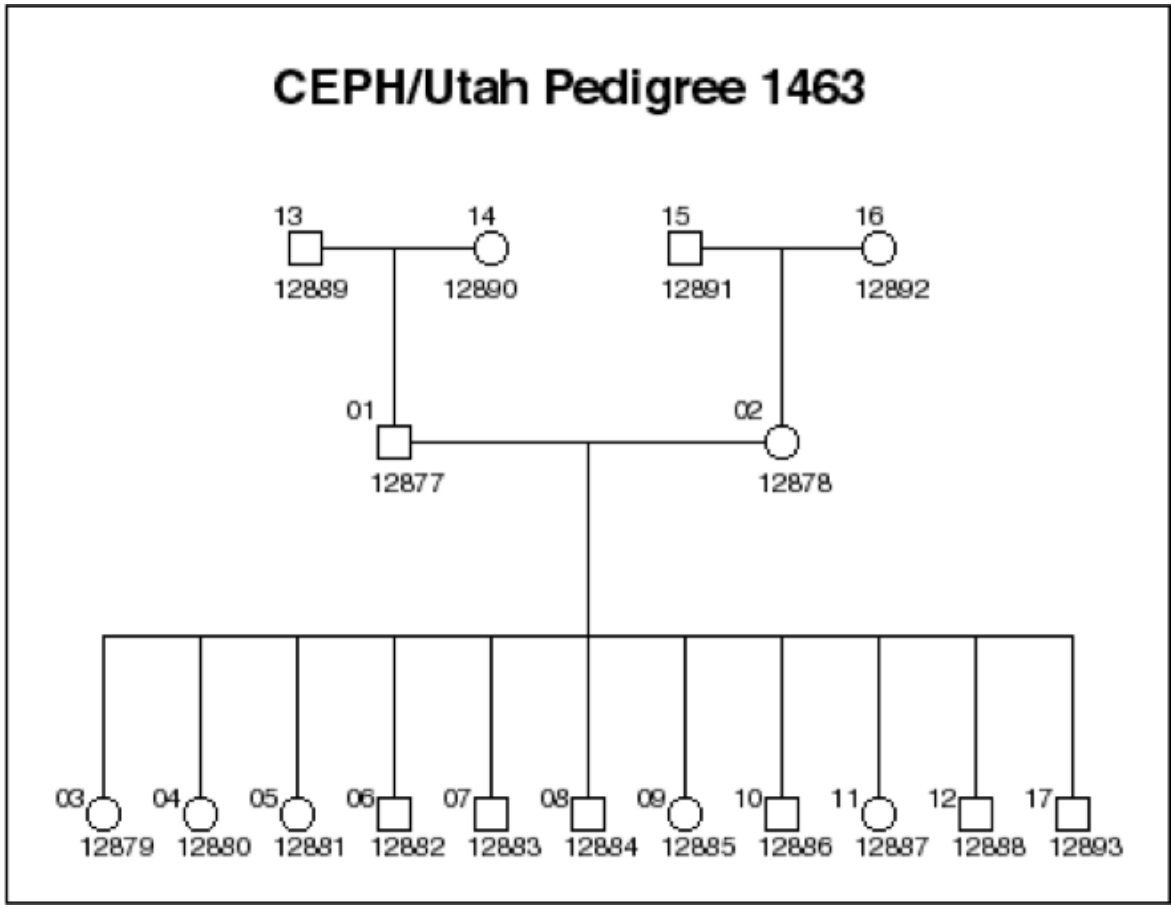
Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative Genetic Mapping of the Human Genome

JEAN DAUSSET,* HOWARD CANN,* DANIEL COHEN,* MARK LATHROP,* JEAN-MARC LALOUEL,† AND RAY WHITE†,[1]

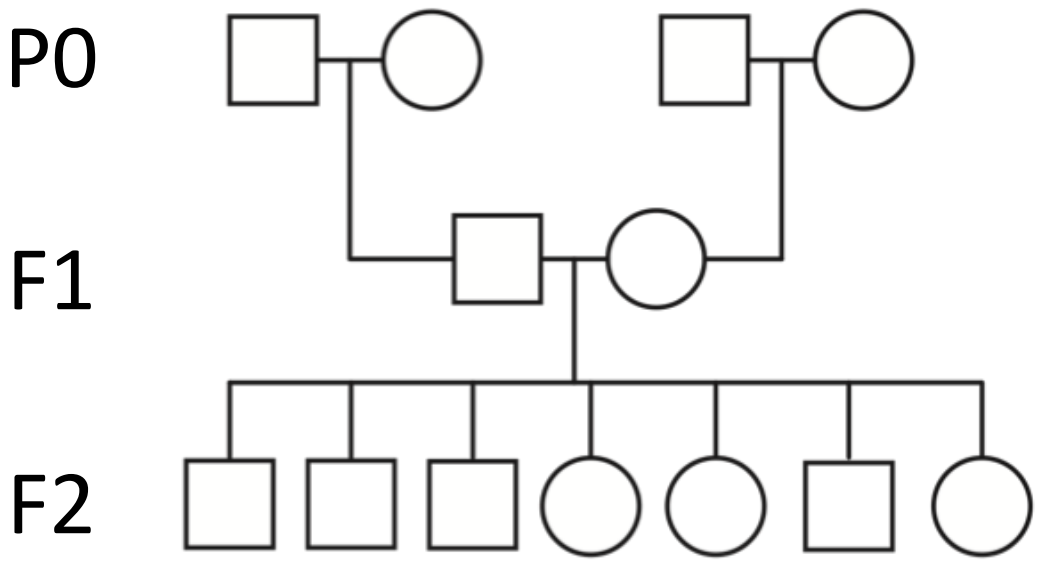
*Centre d'Etude du Polymorphisme Humain (CEPH), 27 rue Juliette Dodu, 75010 Paris, France; and †Howard Hughes Medical Institute and Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, Utah 84132

Genomics, 1990

Tom Sasani,
PhD student



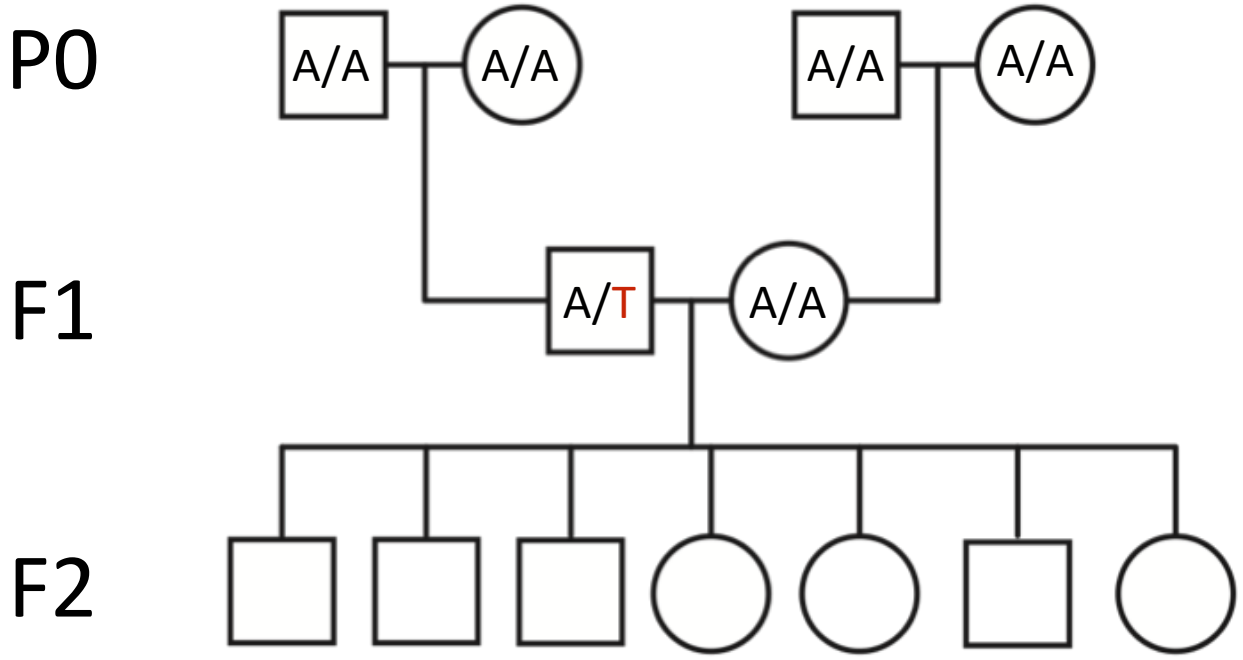
Overview of CEPH/Utah dataset



603 individuals
34 families
4-16 F2 per family
DNA from original blood draws

Illumina sequencing @ ~32X coverage for all individuals

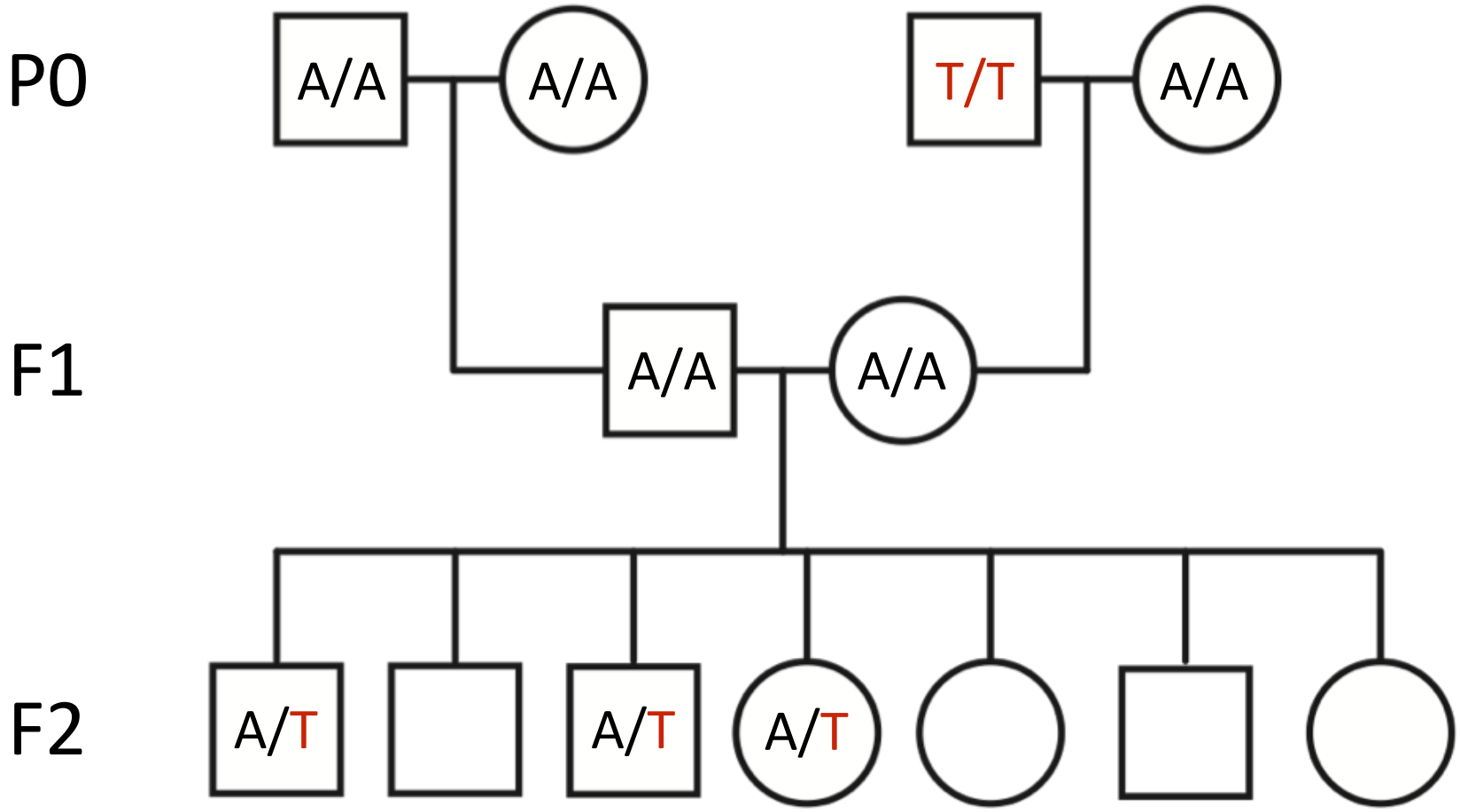
CEPH/Utah pedigrees offer excellent power for detecting and validating DNMs by transmission



P (not inherited) 0.5 x 0.5 x 0.5 x 0.5 x 0.5 x 0.5 x 0.5

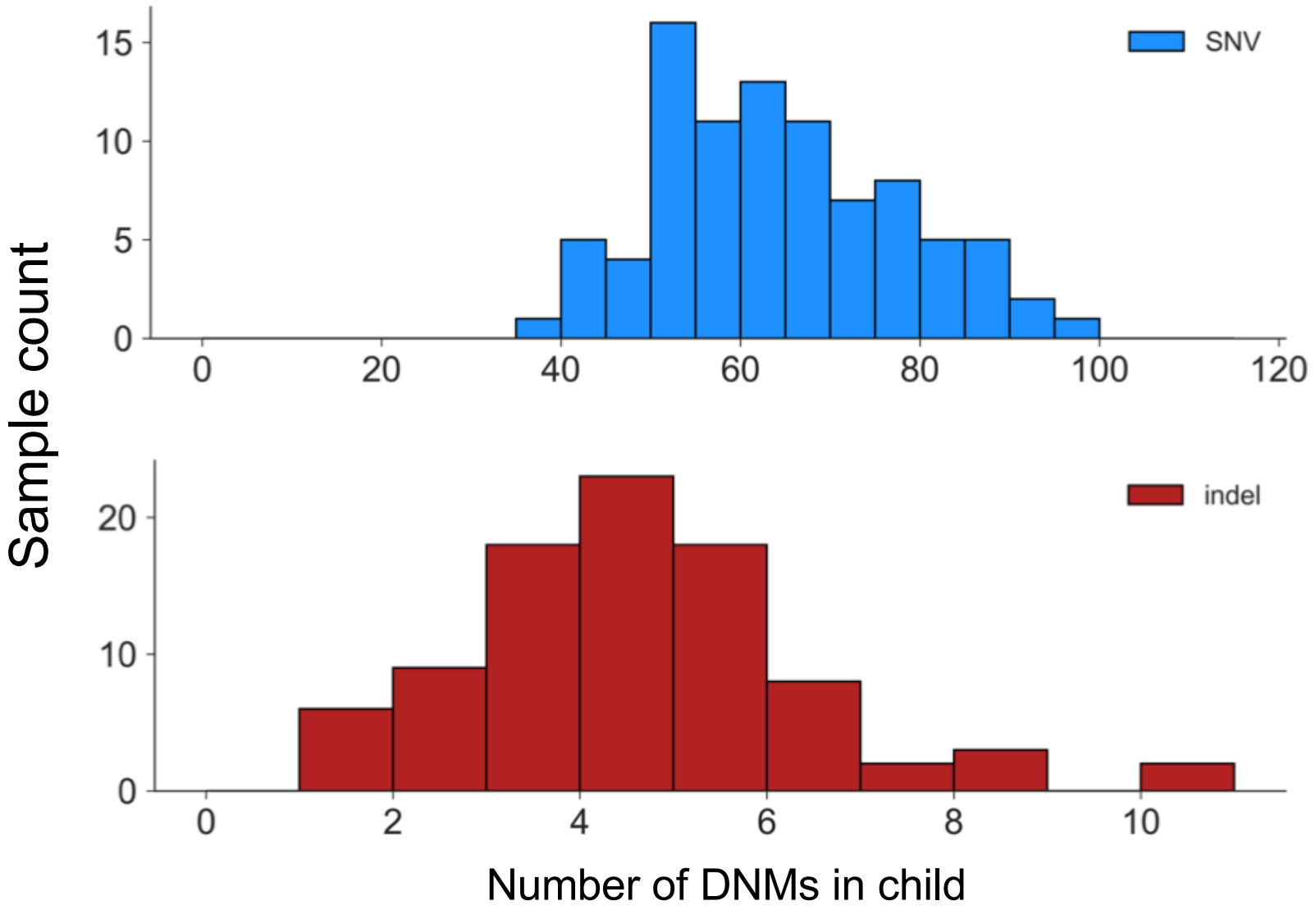
$= 0.5^7 = 0.0078$ – a likely false positive

Estimating a false negative rate of DNM detection

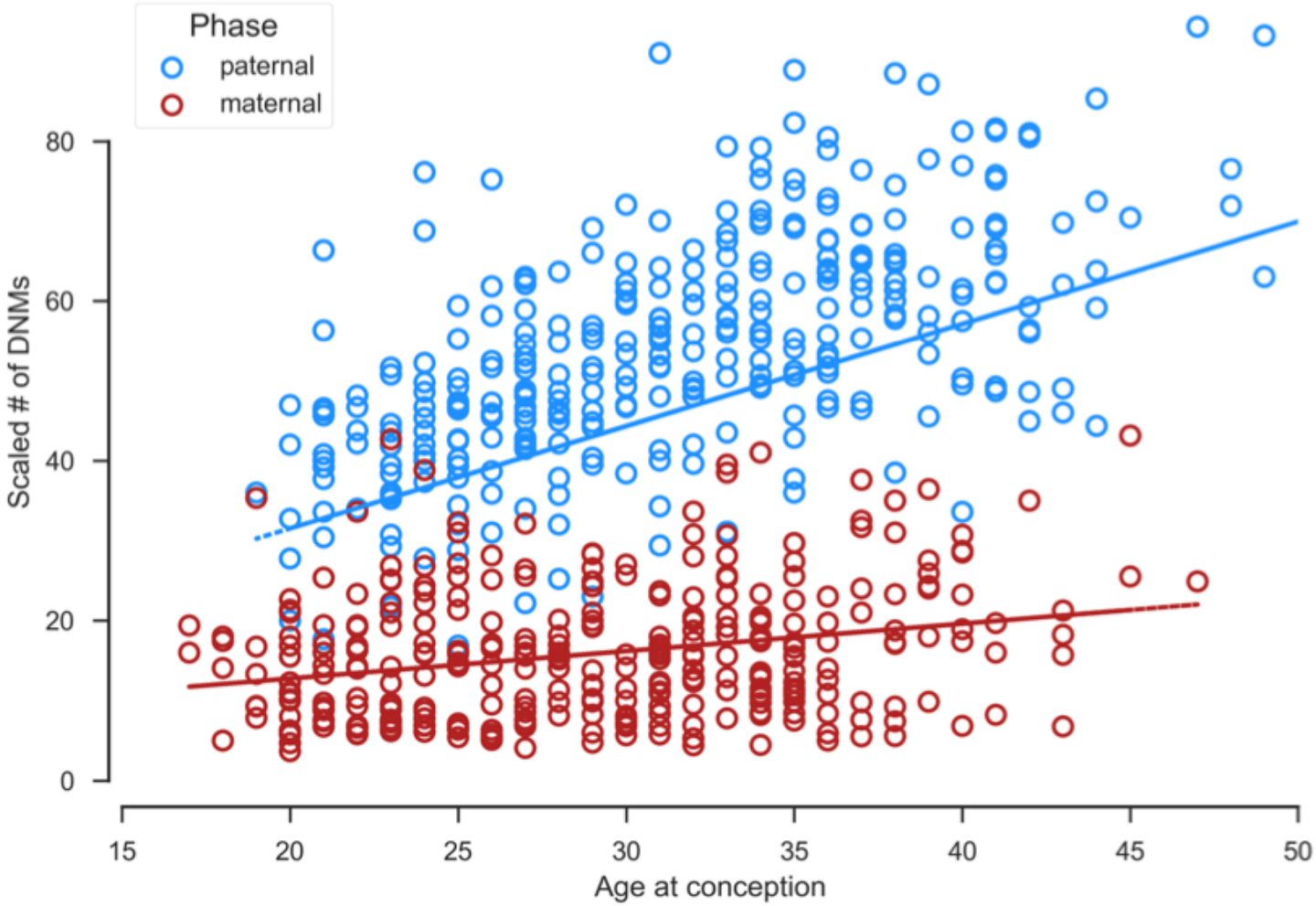


False negative rate is < 4%

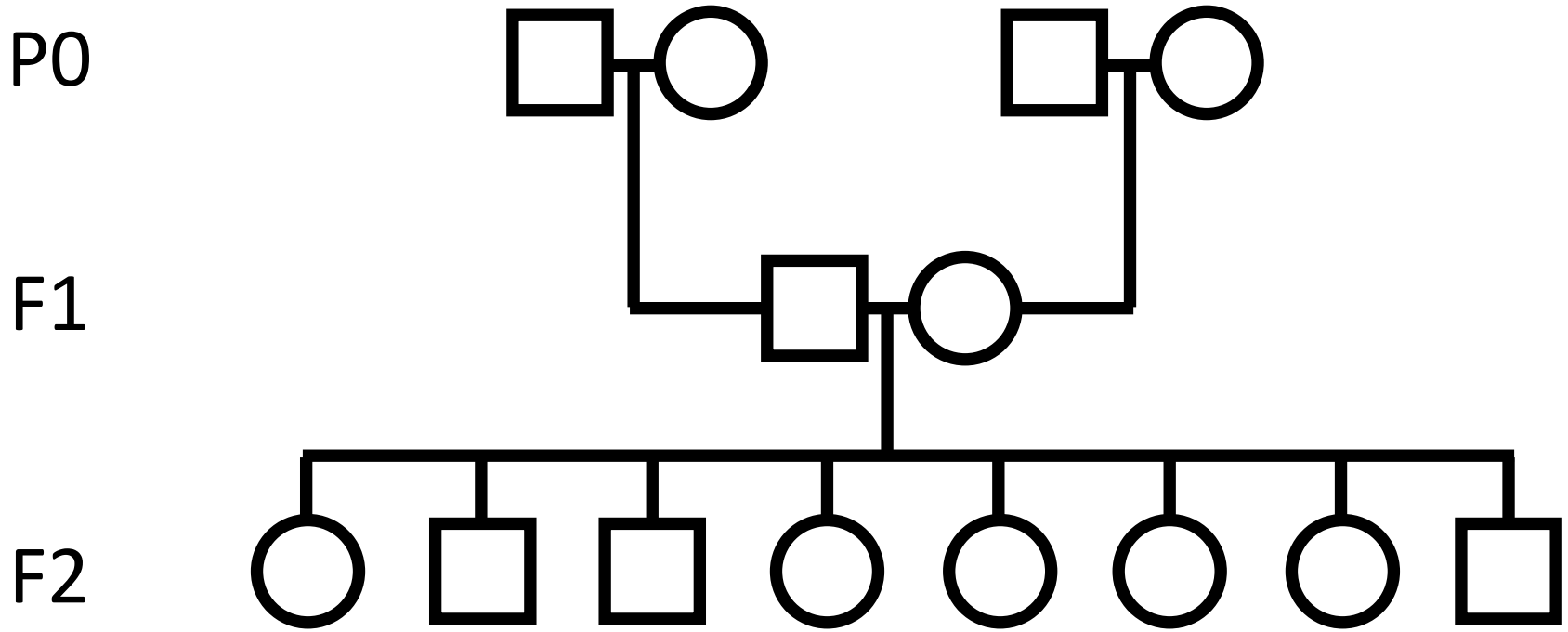
By validating DNM in the F1 generation by transmission, we find an average of 71 DNMs per individual ($\mu = 1.3 \times 10^{-8}$)



We observe ~1.3 additional DNM per year of paternal age; 0.3 DNM per year of maternal age

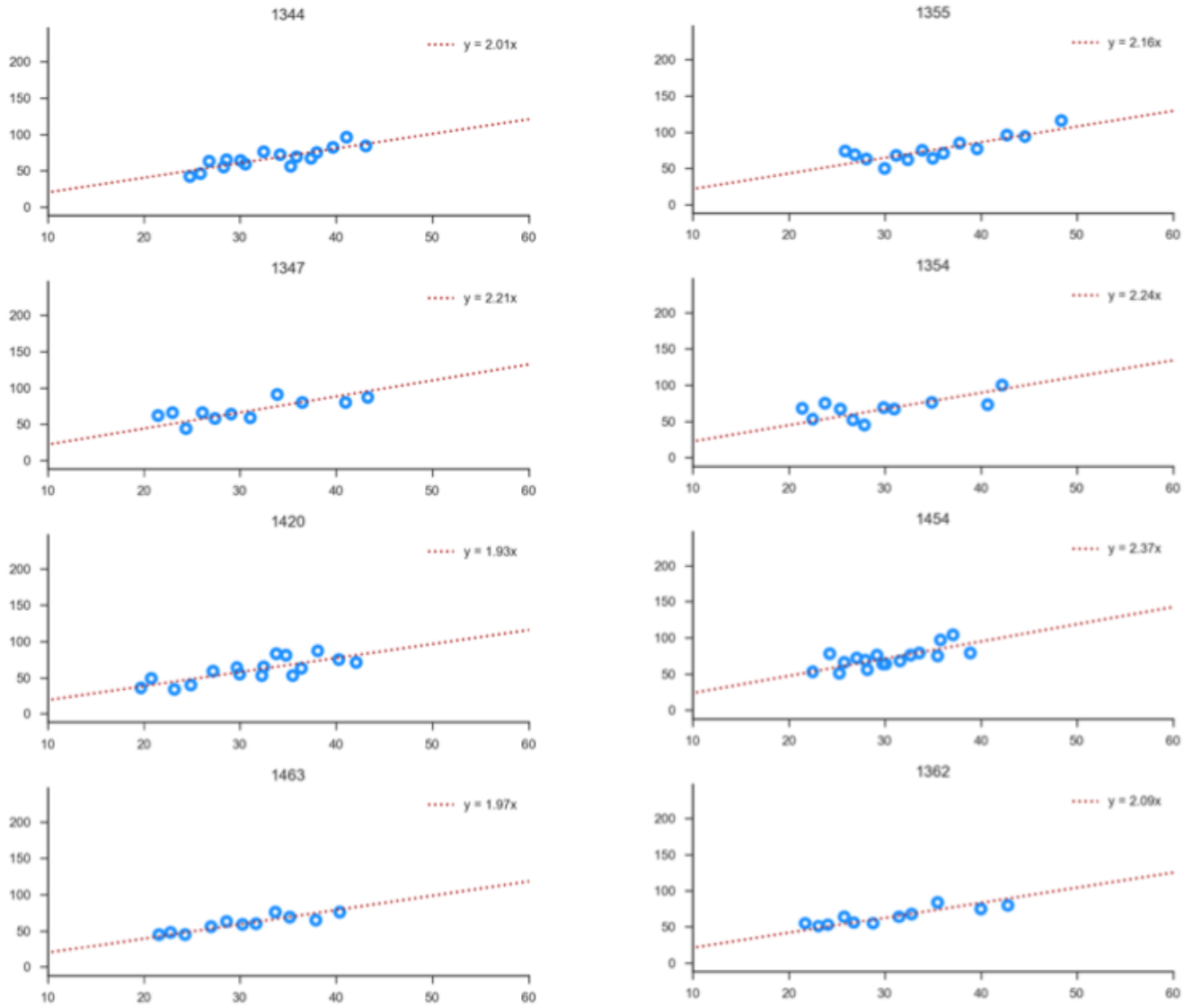


Large F2 allows longitudinal evaluation of DNM rate

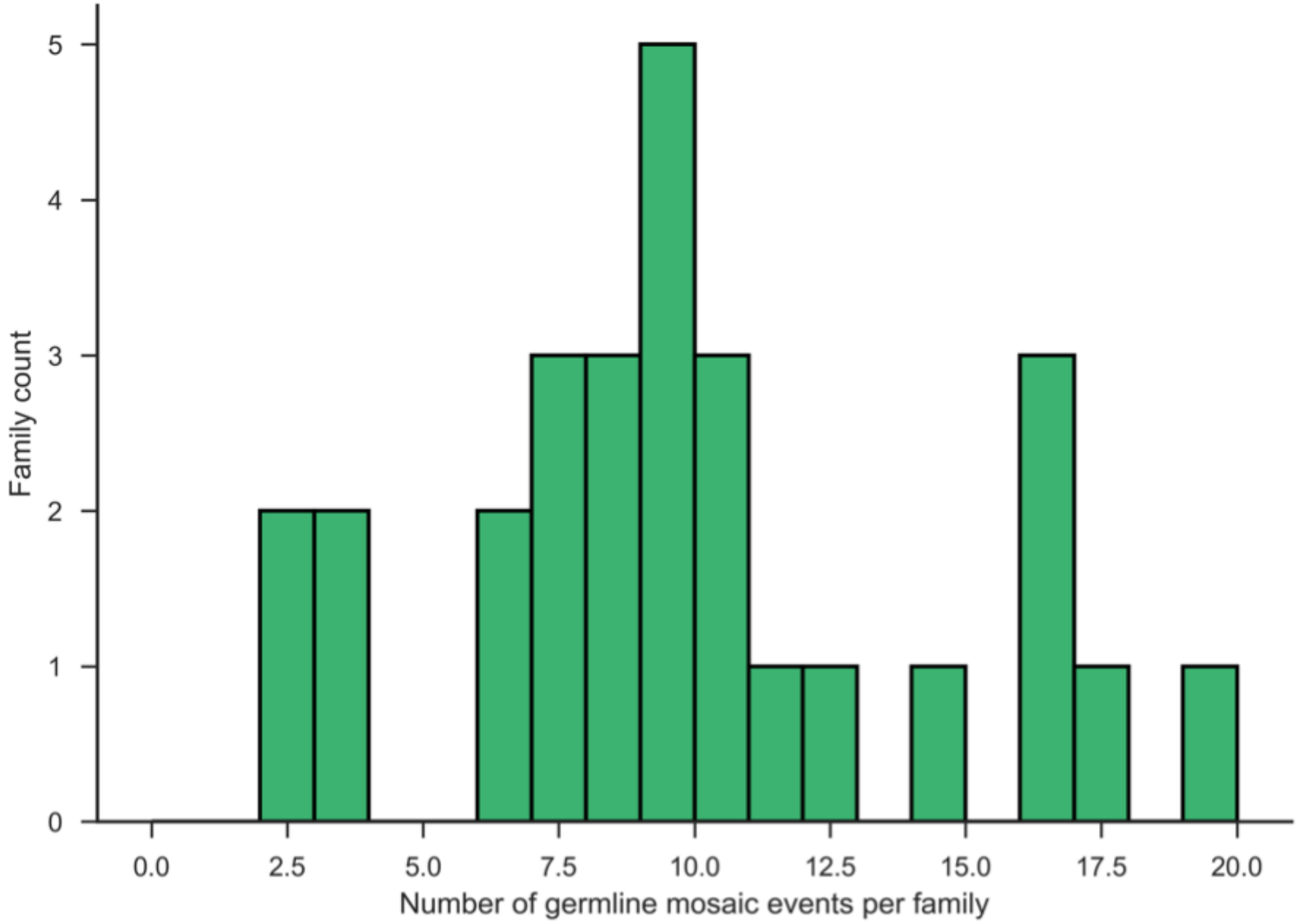


Father's age at conception	20	21	22	23	26	27	35	39
# DNMs	27	47	50	44	64	59	70	78

Paternal age effect within individual CEPH pedigrees



339 germline mosaic events (de novo variants in multiple F2s) identified across CEPH families



"Return to CEPH"

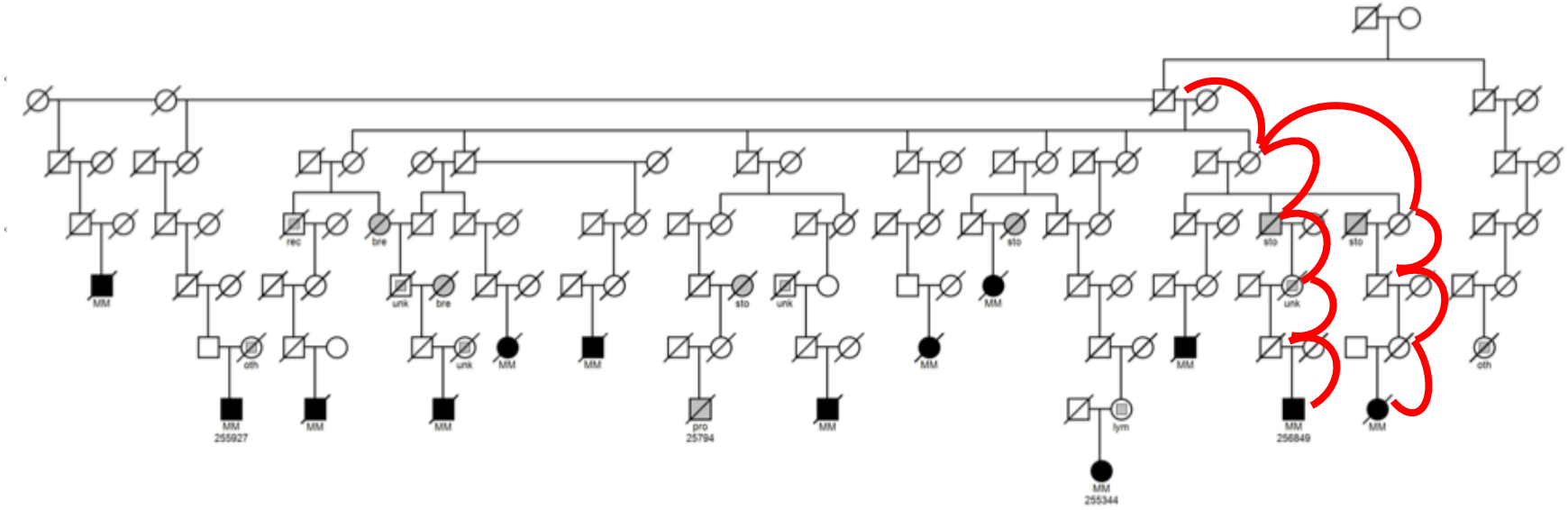
- Families were recontacted in 1990s
- 180 phenotype variables collected
  - Lipid profiles
  - Blood chemistry panel
  - Blood pressure, ECG, echocardiogram
  - Anthropometrics
  - Behavioral assessments
  - Plasma, urine samples
- Fourth-generation data collection to be initiated with recontact of families and follow-up

Prescott *et al*., 2008, *Annu. Rev. Genomics Hum. Genet.*

Finding new genes in large pedigrees: Shared Genomic Segments (SGS)

Nicola Camp, PhD

- Ideal for large pedigrees -- inherited genomic sharing unlikely in very distant relatives
  - Inherited sharing across 15 meioses is unlikely across a genome
  - Shared environment unlikely for distant relatives
  - Small sample sizes have good power for gene detection



Waller RG, …, Camp NJ, 2018, *PLoS Genetics*

Finding new genes in large pedigrees: Shared Genomic Segments (SGS)

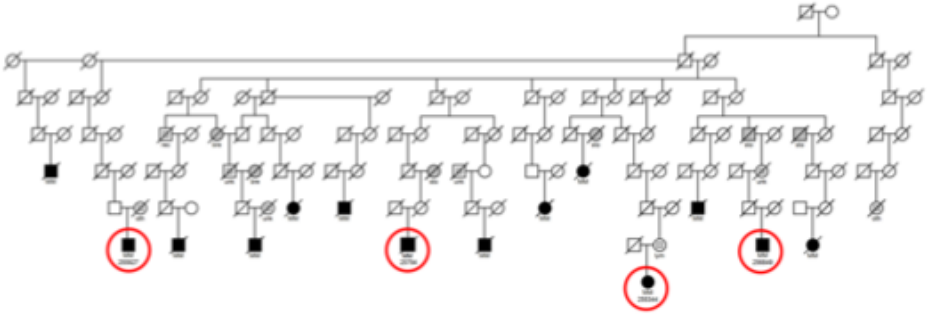
- Ideal for pedigrees -- inherited genomic sharing unlikely in very distant relatives
  - Inherited sharing across 15 meioses is unlikely across a genome

- High-density SNPs provide good resolution of genomic regions
  - SNP microarrays
  - SNPs extracted from WGS

Finding new genes in large pedigrees: Shared Genomic Segments (SGS)

- Ideal for pedigrees -- inherited genomic sharing unlikely in very distant relatives
  - Inherited sharing across 15 meioses is unlikely across a genome
- High-density SNPs provide good resolution of genomic regions
  - SNP microarrays
  - SNPs extracted from WGS
- Simple concept -- identity-by-state (IBS) used to infer identity by descent (IBD)
  - Length shared IBS is significantly longer than expected by chance, then likely IBD
  - Empirical estimate of genome-wide significance
    - Pedigree structure
    - Marker and LD map
  - Intra-familial heterogeneity assessed with all-subsets approach

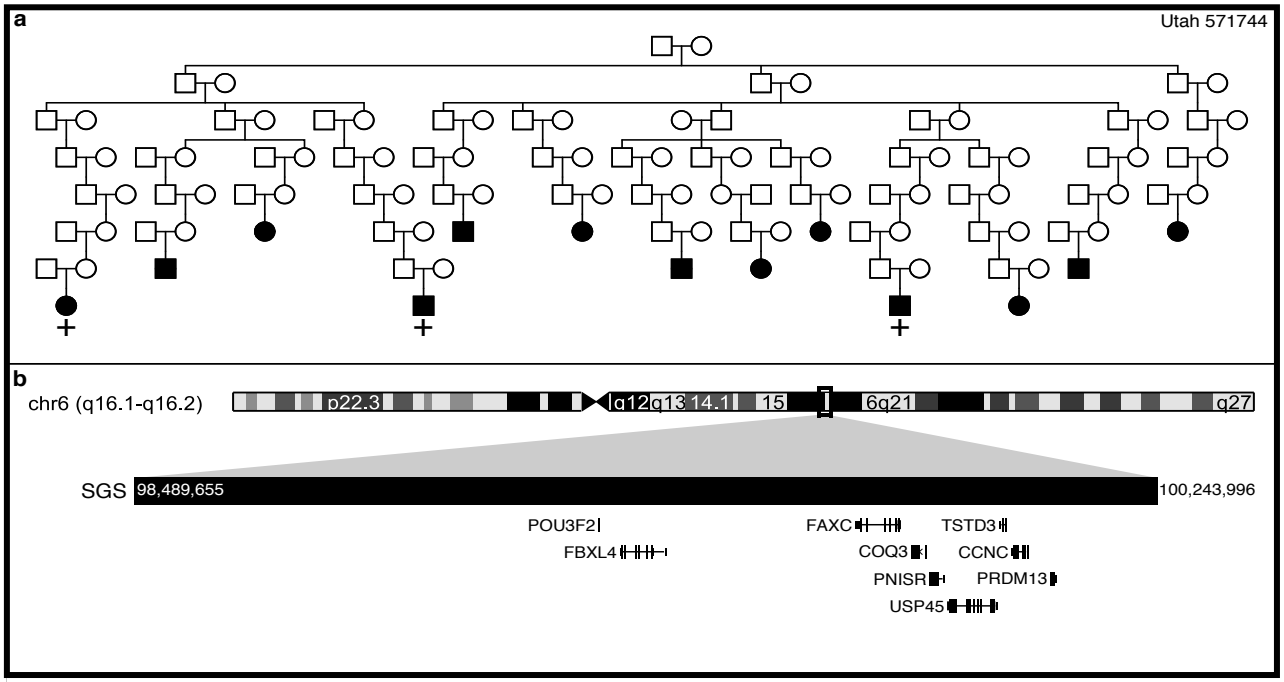
SGS applied to multiple myeloma



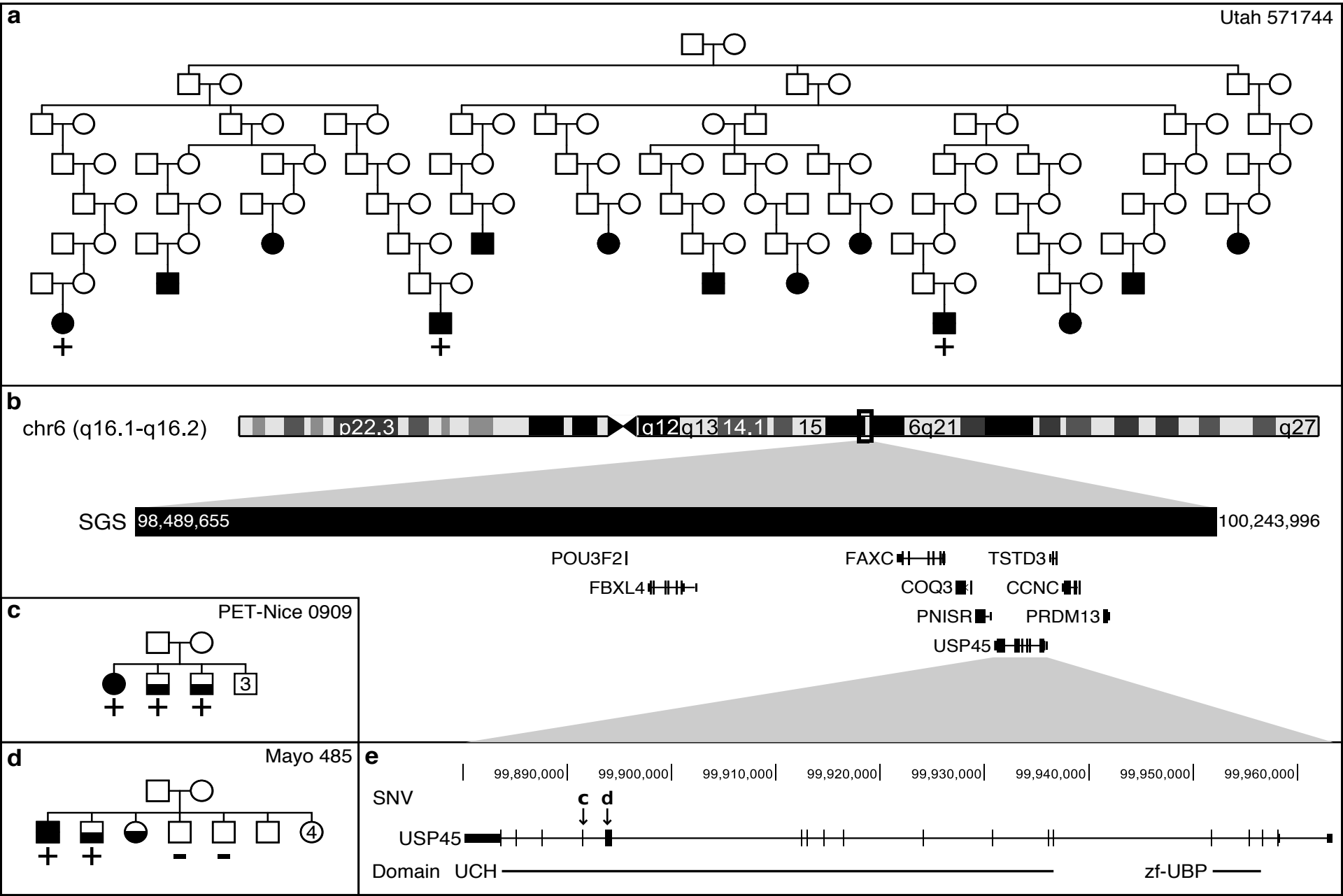
- Late-age onset hematologic cancer
- <50% five-year survival
- GWAS accounts for ~20% of heritability (Mitchell *et al.,* 2016, *Nat. Comm.)*
- 11 high-risk MM pedigrees identified from UPDB
  - Genotype SNP array
  - 3-4 sampled cases per pedigree
  - 8-23 meioses between cases
- Shared Genomic Segment analysis
- Exome sequencing
  - Utah pedigrees
  - Mayo pedigrees
  - INSERM, France pedigrees
  - Cornell pedigrees
  - 200 early onset MM
  - Publicly available data
    - coMMpass MM sporadic cases
    - controls

Waller RG, …, Camp NJ, 2018, *PLoS Genetics*

One genome-wide significant SGS region (1.8 Mb) containing nine genes



Exome sequencing in the region



One gene, *USP45*, contains pathogenic variants

Table 1. Significant

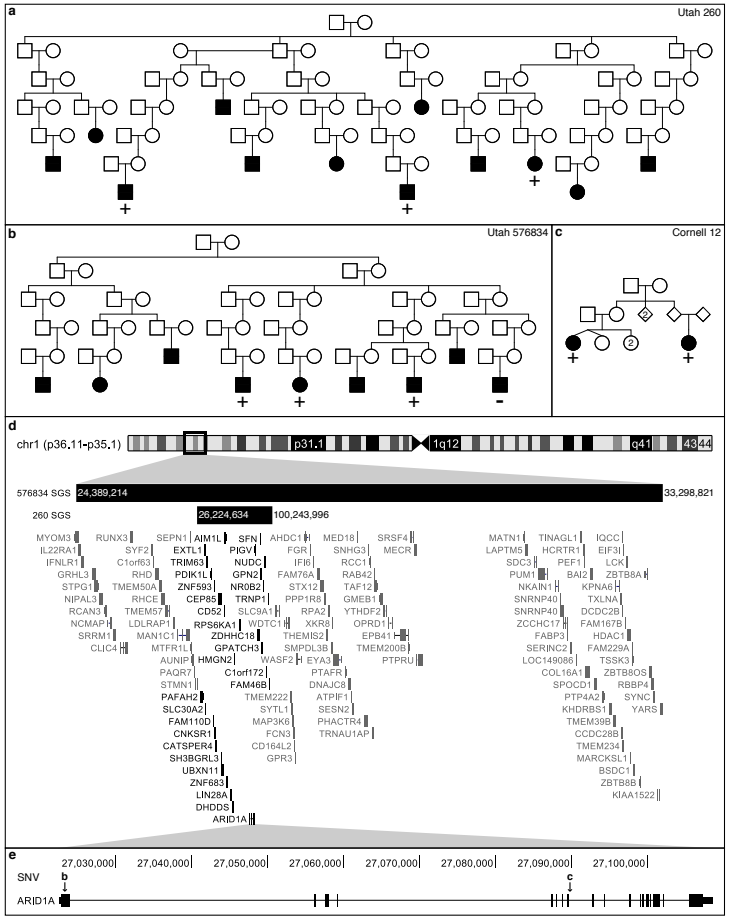
Family	MM	MGUS	Meioses	Type	Position~	Length[l]	P-Value	Gene	Consequence	Impact	AAF^
Utah 571744	3		20	SGS	6:98,489,655—100,243,996	1.8	$3.3x10^{-6}$*				
PET-Nice 0909	1	2	3	SNV	6:99,891,443			*USP45*	p.Gln691*	Stop Gain	None
Mayo 458	1	1	2	SNV	6:99,893,787			*USP45*	p.Gln621Glu	Missense	None

~Build HG19. [l]In mega-bases. *Genome-wide significant. ^Alternate allele frequency based on the non-TCGA, non-Finnish, European ExAC individuals.

USP45 is a DNA repair gene

- Deubiquitylates ERCC1, a catalytic subunit of the ERCC1-ERCC4 DNA repair complex
- Mouse USP45 knockout models
  - higher levels of ubiquitylated ERCC1
  - cells are hypersensitive to UV radiation and DNA inter-strand cross-links
  - DNA damage repair is impaired
- Well-established role in tumor resistance to chemotherapy

Regions identified by multiple pedigrees



ARID1A involved in chromatin remodeling

Intolerant to LOF and missense variation

Prominent role in cancer somatic mutations
--tumor suppressor

ARID1A

Table 1. overlapping SGSs and segregating SNVs.

Family	MM	MGUS	Meioses	Type	Position~	Length[]	P-Value	Gene	Consequence	Impact	AAF^
Utah 576834	3		12	SGS	1:24,389,214—33,298,821	8.9	$3.0x10^{-4}$				
Utah 260	3		16	SGS	1:26,224,634—27,384,988	1.2	$2.1x10^{-4}$				
Utah 576834	3		12	SNV	1:27,023,162			*ARID1A*	p.Ser090Gly	Missense	0.0000
Cornell MM12	2	0	4	SNV	1:27,089,712			*ARID1A*	p.Met890Val	Missense	0.0001

~Build HG19. []In mega-bases. *Genome-wide significant. ^Alternate allele frequency based on the non-TCGA, non-Finnish, European ExAC individuals.

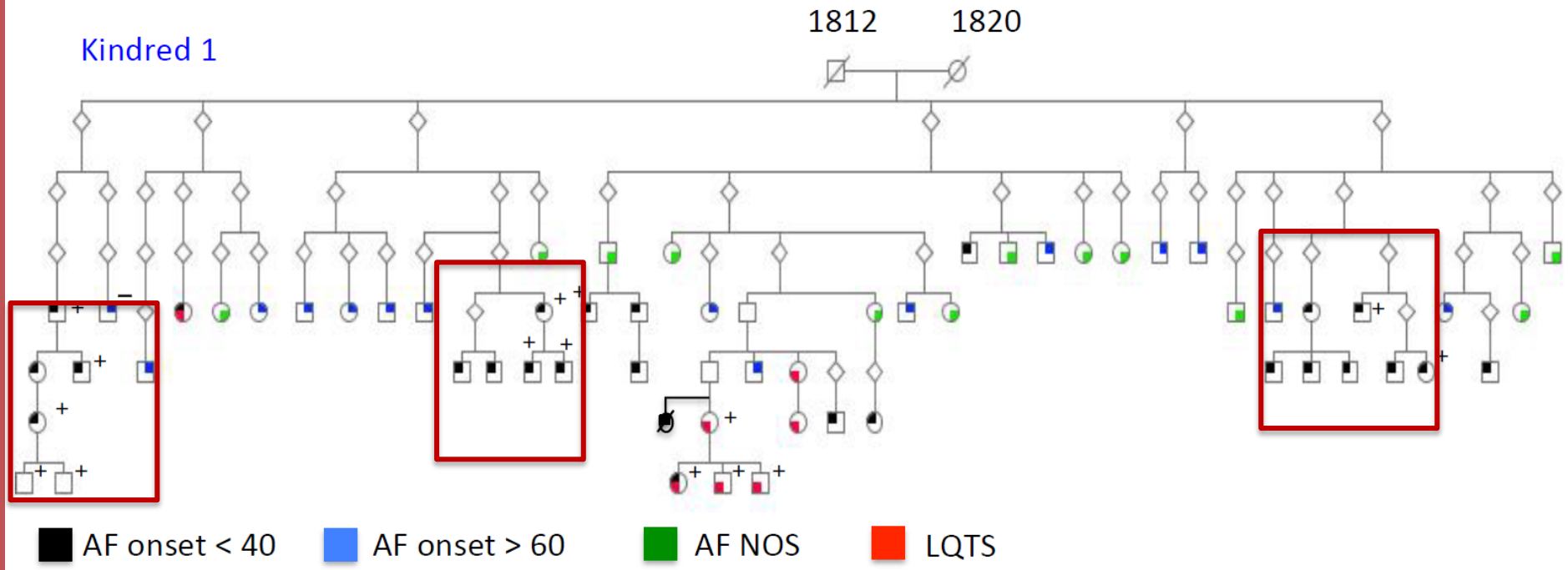
Ding **(2017)** ***Cancer Research***. Pediatric ALL mutation profiles. ARID1A potential tumor suppressor; Rao **(2017)** ***Carcinogenesis***. 5 most frequently somatically deleted genes in liver cancer; **(2017)** ***Cancer Discovery***. NEWS HIGHLIGHT. ARID1A deficiency promotes colon cancer via enhancer dysregulation; Mathur **(Dec 2016)** ***Nature Genetics***. ARID1A loss impairs enhancer mediated regulation and drives colon cancer in mice. Coatham **(Dec 2016)** ***Modern Pathology***. Inactivating mutations and loss of ARID1A expression in aggressive ovarian and endometrial cancer; Agrihoti **(Nov 2016)** ***Nature Genetics*****.** Recurrent mutations in ARID1A in nerve sheath tumors; Berns **(Nov 2016)** ***Clinical Cancer Research*****.** Reduced ARID1A expression controls resistance in breast cancer (HER2/P13K/mTOR targeting agents);….

Other examples of SGS analyses in large Utah pedigrees

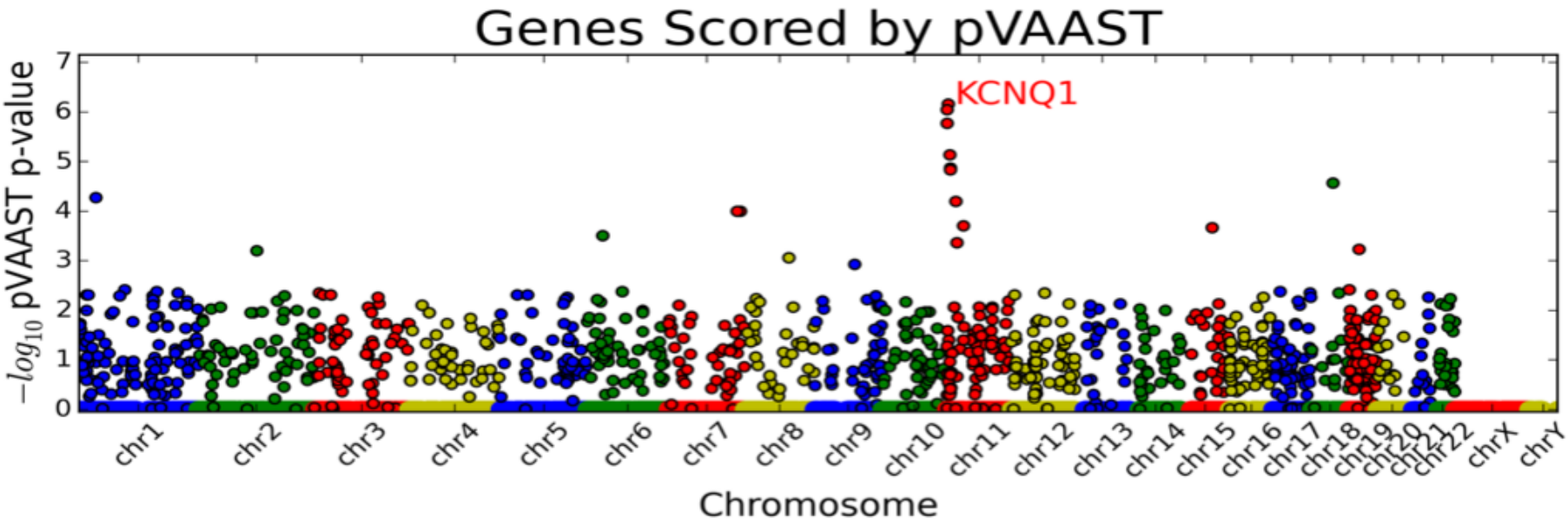
- Preterm birth
- Amyotrophic lateral sclerosis
- Small intestinal neuroendocrine tumors
- Autism
- Suicide
- Crohn disease

Family data can identify at-risk pedigree members: Young-onset atrial fibrillation

Martin Tristani-Firouzi, MD

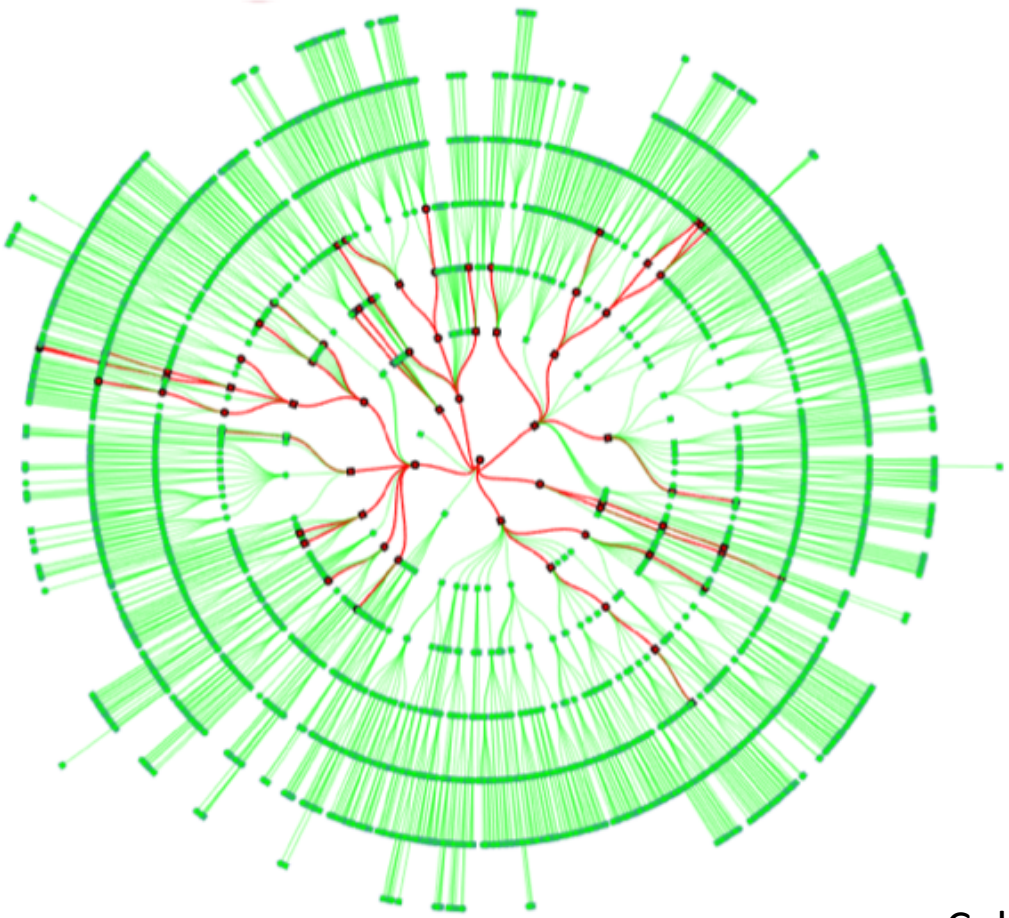


KCNQ1 mutation causes atrial fibrillation in all family branches



KCNQ1 (*LQT1*) originally identified as a long QT syndrome gene in a large Utah family (Keating *et al.,* 1991, *Science*)

Projecting medical records onto family histories



Collaboration underway
with AncestryDNA

SUMMARY: THE VALUE OF FAMILIES

- Evaluation of *de novo* events
- Parent-of-origin effects
- Mosaicism detection
- Discovery of rare variants difficult to identify in samples of unrelated subjects
- Follow-up with communication of testing results

ACKNOWLEDGMENTS

Jorde Lab

Julie Feusier

Karin Chen, MD

Kristi Russell

Jainy Thomas, PhD

Scott Watkins, MS

Jin Xing, PhD

David Witherspoon, PhD

Programmers:

Andrew Farrell, PhD

Matt Velinder

CEPH Project

George S and Dores Eccles

Benning Society

Mark Leppert

Ray White

DNA samples: Lisa Baird

UGP: Barry Moore

Shawn Rynearson

Deb Neklason

Breanna Caruso

DNA Core Facilities

Derek Warner

HCI sequencing core

ARUP sequencing core

WashU Genome Center

Quinlan Lab

Aaron Quinlan

Thomas Sasani

Ryan Layer, PhD

HUNTSMAN CANCER INSTITUTE UNIVERSITY OF UTAH and the The University of Utah Health Sciences Center