Rare Variant Discovery With Extended Pedigrees: Human Subcortical Volumes

Dr. John Blangero

South Texas Diabetes and Obesity Institute

Director, Genomics Computing Center STDOI

UTRGV/UTPA/UTHSCSA

Yale UNIVERSITY

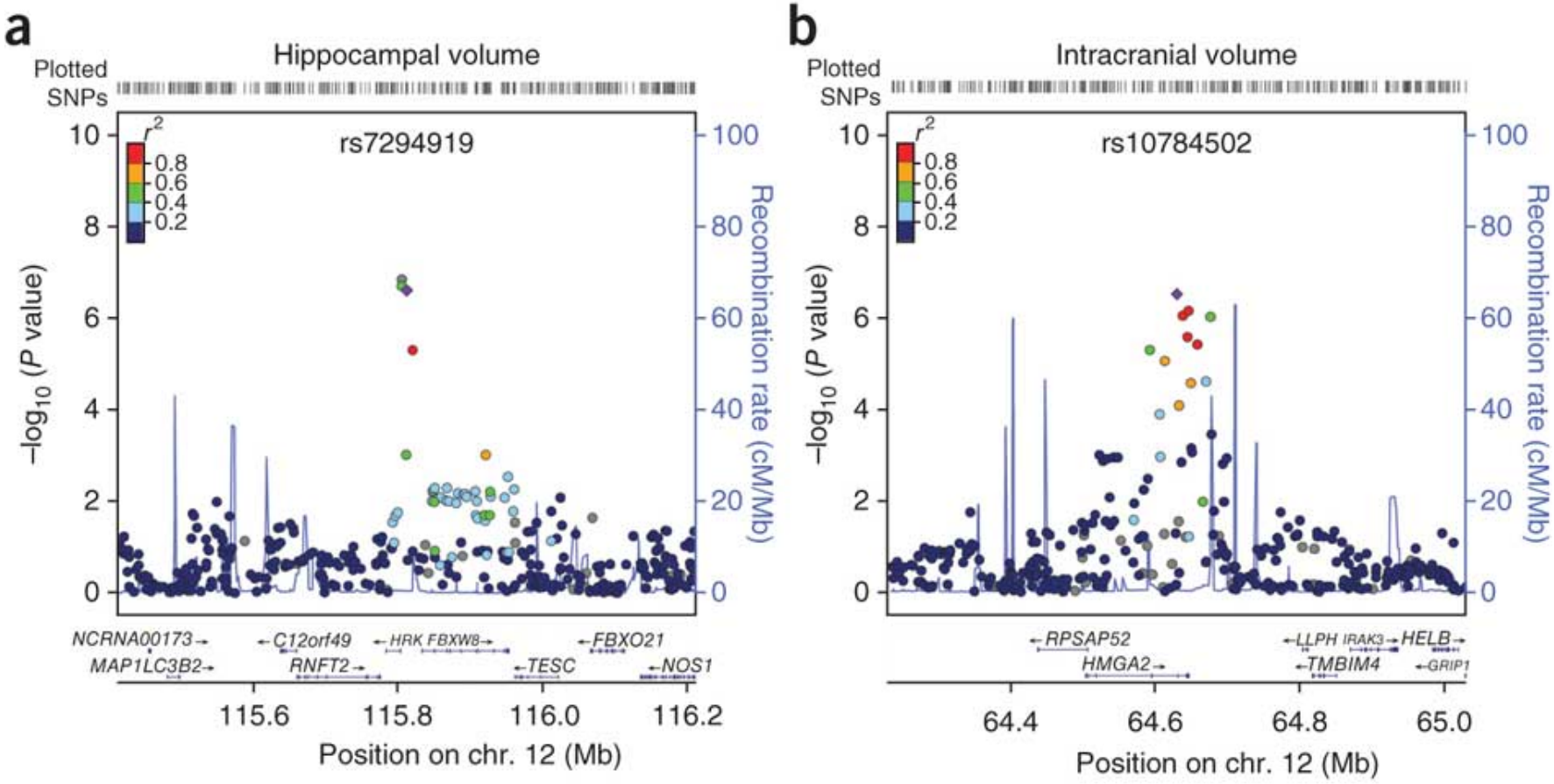
THE INSTITUTE OF LIVING

OHBM, June 14, 2015

No Conflict of Interest

- I will not discuss off label use and/or investigational drugs in my presentation
- I have no financial relationships to disclose (SureScore; 4TELX)
- Financial support was provided by (currently funded grants):
  - U01 MH105630; R01 MH106324; R01 MH078143; R56 MH097940; R01 HL113323; NARSAD; DoD/CDMRP; U54 E020403; R01 MH104284; R21 DA038381; R01 MH094524; R01 MH096957

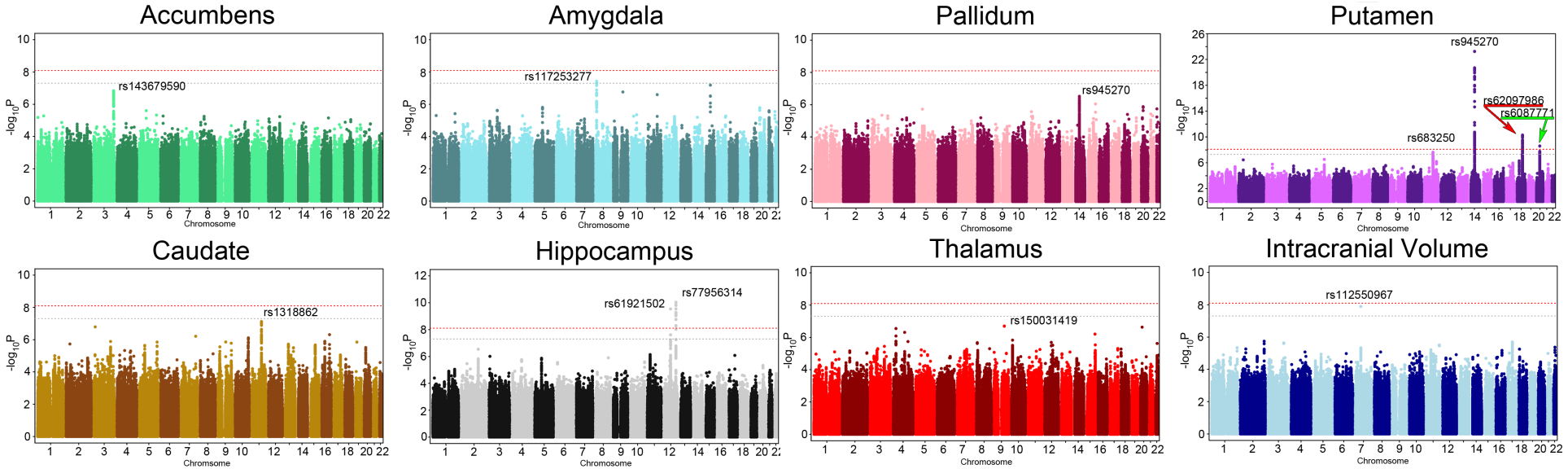
Genome-Wide Significant QTLs For Hippocampus and Intracranial Volume



21,151 Unrelated Individuals

Stein et al., *Nat Genet*, 2012

ENIGMA2: Genome-Wide Significant QTLs for Subcortical Nucli



29,556 Unrelated Individuals

Hibar et al., *Nature*,2015

Genetic Analysis of Complex Traits

QTL Localization

Where in the genome is the QTL located?

QTL Identification

What is (are) the gene(s) involved?

QTL Allelic Architecture

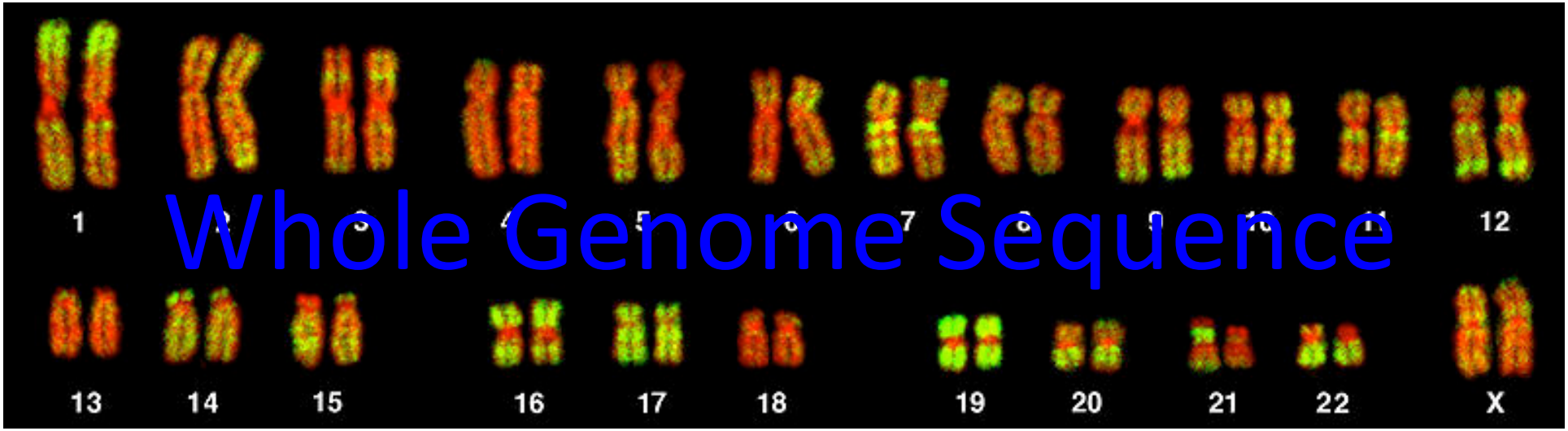
What are the specific QTNs? How many QTNs?

What are their frequencies and effect sizes?

Status of Human QTL Localization: Current Status

- Accumulated large numbers of QTL localizations but VERY FEW gene identifications.
- QTL region size from linkage: ~10-15Mb
- QTL region size from association: ~ 500kb
- Identification of the underlying genes will require deep comprehensive sequencing of these regions to find functional variants.

Genetic Search Space is Huge but Finite



23 Chromosomes

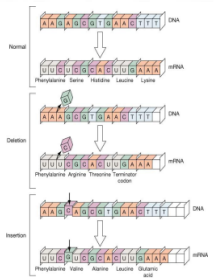
~20-25,000 genes

~3 billion base pairs

How Many Human DNA Variants?

6.77 billion people on Earth

Each has ~80 new mutations

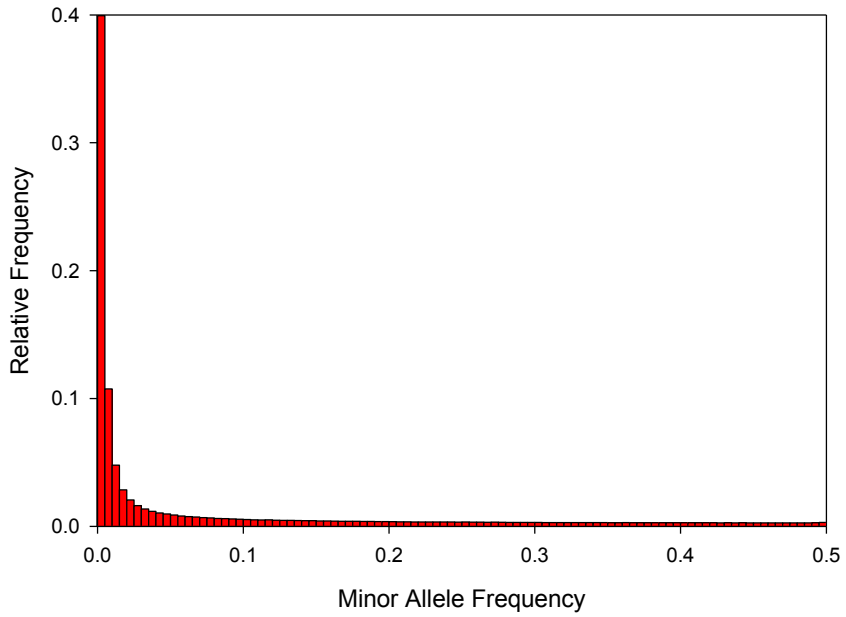


~536 billion new human mutations on Earth

~162 variants per each site in the human genome!
(except for those incompatible with life)

Growing Realization from Whole Genome Sequencing Studies

~67% have MAF < 0.01



- Vast amount of rare (even private) sequence variation in human populations
- Common variation accounts for only a small proportion of the observed heritability of a given trait
- Increasing evidence that rare functional variants contribute to complex traits

Rare Variant Hypothesis

- A substantial component of quantitative trait variation is due to “rare” sequence variants in multiple genes
- Larger effects of rare variants will make gene discovery easier. It is easier to perform required functional experiments on variants with larger mean displacements

QTL Effect Sizes

QTL Type	Heritability due to Variant	α (sdu) Displacement between Genotypic Means
Rare monogenic	1.0	>3.5
Rare partially penetrant	>0.10 in a given pedigree	0.50 < α < 3.5
GWA-derived common variant	<0.01	<0.15

How Can We Study Rare Variants: Return of the Family Study

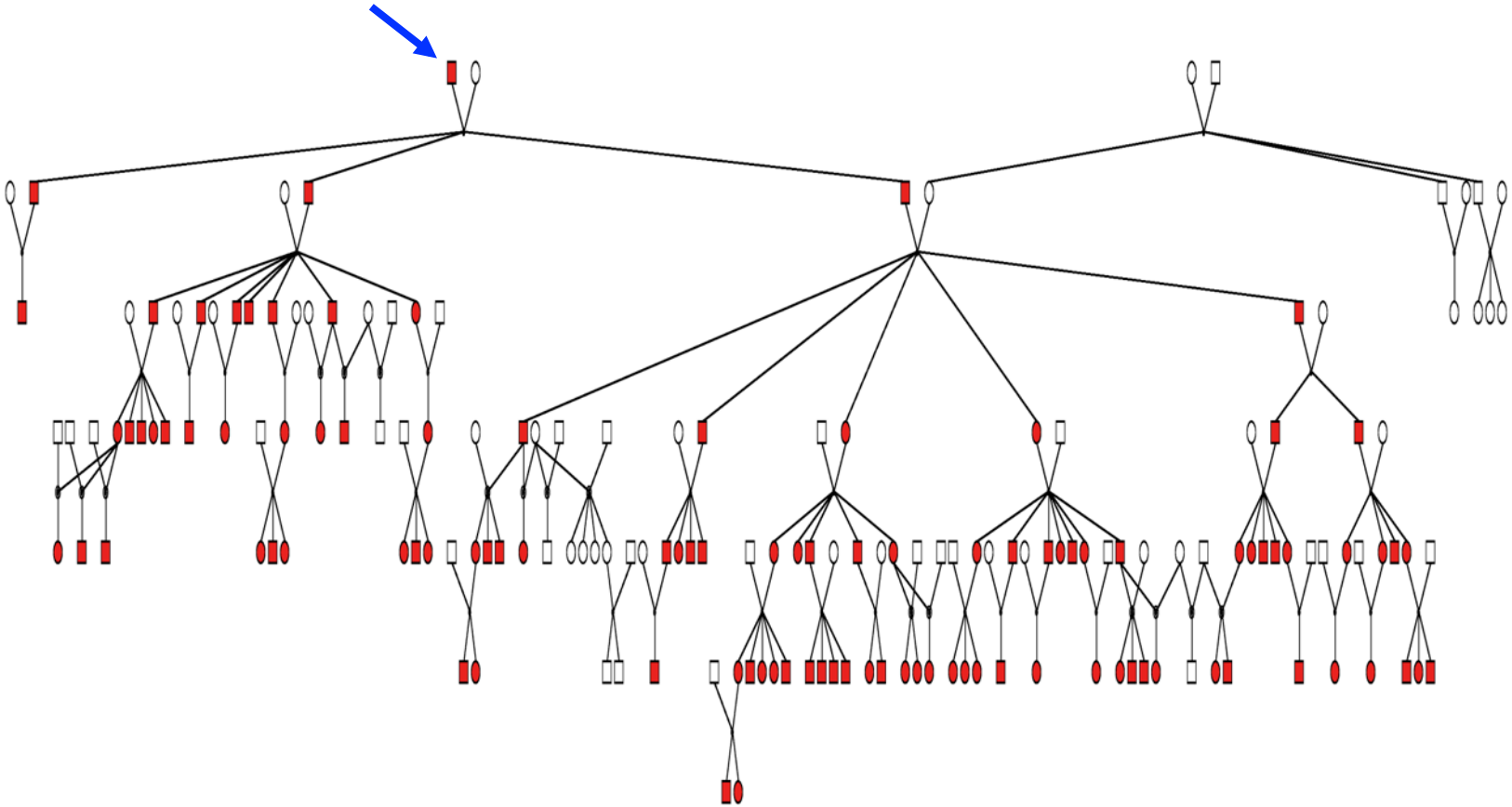
- Rare functional variants are best detected (and their genes identified) using a large pedigree-based design
- Pedigrees allow observation of multiple copies of a private variant
- Large pedigrees with large lineages are optimal for the study of private variants because of the potential to generate sufficient numbers of private allele copies

Large Numbers of Private Exonic Variants

- Seen at least twice in one pedigree: N=45,977
- Novel (not in ESP): N = 24,037
- Avg. of 60 novel variants per founder

Special thanks to Alisa Manning of T2D-GENES Project 2 for identifying these.

San Antonio Family Study Pedigree 1: Founder 1 Lineage



105 possible copies of founder private variant

Strategies for Gene Identification with Whole Genome Sequence Data

Prior hypotheses needed as naïve search of WGS destroys power to detect an effect

Strategies to reduce number of variants tested:

- Protein altering variants
  - Predicted to be damaging
- Putative regulatory variants
- Conserved variants
- Variants showing evidence of natural selection
- QTL regions from linkage or association

Defining Deleterious Variants

- **PolyPhen2** (**Poly**morphism **Phen**otyping v**2**) predicts the impact of an amino acid substitution on the structure and function of a human protein
- **SIFT** (**S**orting **I**ntolerant **f**rom **T**olerant) predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids
- The correlation between PolyPhen2 and SIFT for 800 non-synonymous variants associated with brain-related traits was r^2=0.45, suggesting poor comparability.
- Thus all bioinformatic prediction needs to be biologically confirmed

Genetics of Brain Structure and Function Study

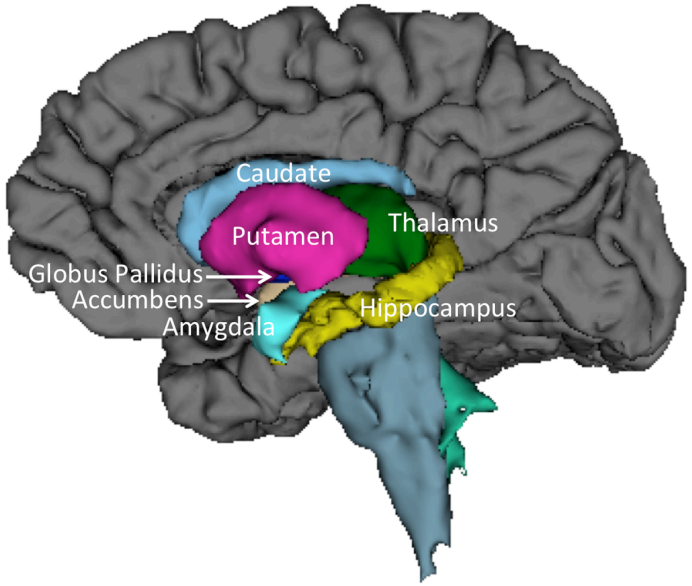
- Extension of San Antonio Family Study (n=~3000)
- 2,000 Mexican-American individuals from ~60 randomly ascertained extended families
  - 1,824 individuals examined to date
- Genotyping:
  - 1M SNPs
  - ~500 whole exome sequence
  - ~2000 whole genome sequence
- Transcriptional profiling of lymphocytes
- Phenotypes: structural & functional brain imaging, neurocognitive assessments, psychiatric diagnoses

Glahn et al, *Biol Psychiatry*, 2012

Sample & MRI Assessment

- 890 individuals with anatomic scans and sequence data for this study
- T1-weighted, 3D structural sequence:
  - 800 micron isotropic voxels (FOV=200 cm)
  - TurboFLASH, adiabatic inversion recovery pulse
  - TR=2000ms, TI=795, flip angle=8 degrees
  - 6 separate images, retrospective motion correction
  - Gray/white contrast of ~25% and SNR~25

FreeSurfer Pipeline



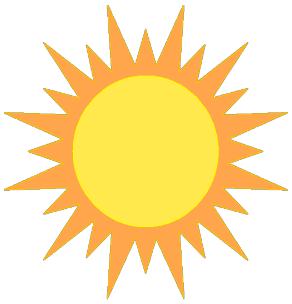
Jung et al., *Plos One*, 2014

http://surfer.nmr.mgh.harvard.edu

- Removal of non-brain tissue/ intensity normalization
- Linear spatial normalization
- Segmentation of subcortical volumetric structures
- Tessellation of white-matter
- Automated topology correction
- Surface deformation / inflation
- Registration to a spherical atlas
- Parcellation into gyral structure

Fischl & Dale, *Proc Nat Acd Sci*, 2000; Winkler et al., *NeuroImage*, 2010

Genetic Analyses: SOLAR



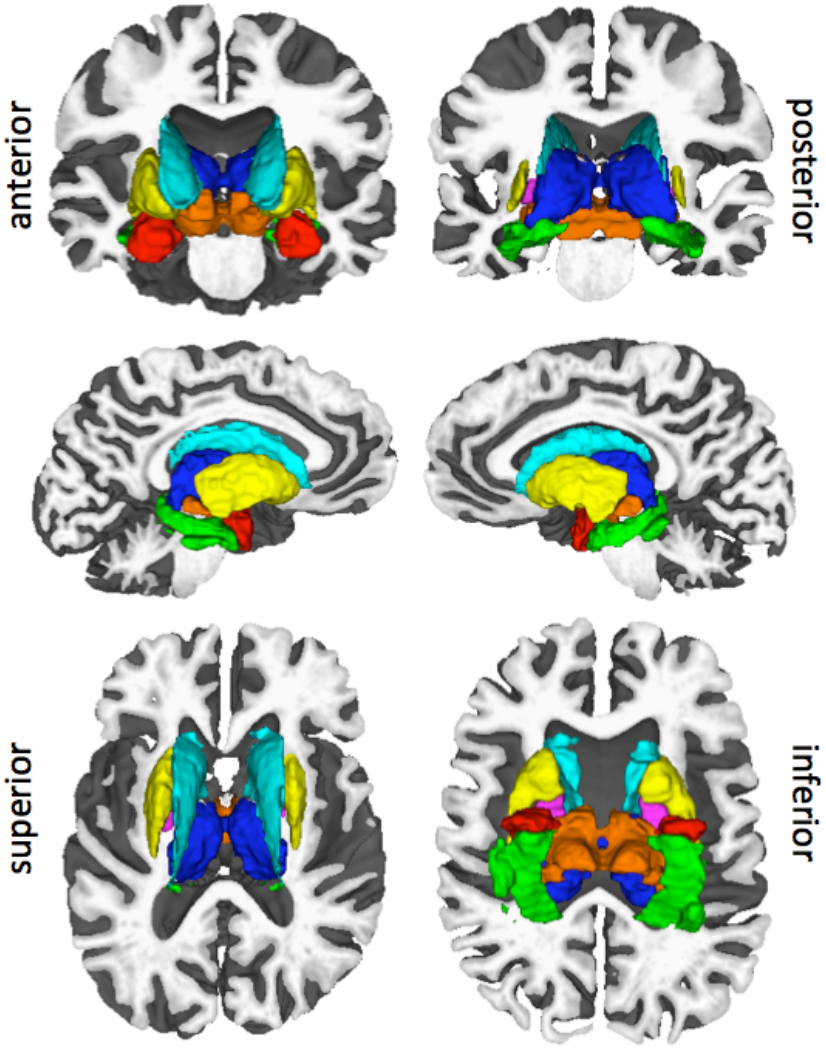
Sequential Oligogenic Linkage Analysis Routines

http://solar.txbiomedgenetics.org

- Statistical genetic analyses were performed with SOLAR and included:
  - Estimation of Heritability
  - Pleiotropy Analysis
  - Genome Association Analysis

Almasy & Blangero, *Am J Hum Genet*, 1998

Subcortial Nuclei Are Heritable



Subcortical nuclei	h^2	p-Value
Amygdala	0.73	1.4×10^{-31}
Caudate	0.74	5.8×10^{-31}
Hippocampus	0.70	7.6×10^{-33}
Pallidum	0.59	4.7×10^{-20}
Putamen	0.74	2.8×10^{-28}
Thalamus	0.73	7.4×10^{-32}
Ventral diencephalon	0.66	7.9×10^{-28}

1010 Family Members

Adjusted for covariates:
Age, Age^2, Sex, Sex*Age, Sex*Age^2, Intracranial Volume

McKay et al., *Brain Imaging Behav, 2014*

Whole Genome Sequencing

1018 individuals had direct whole genome sequence data available:

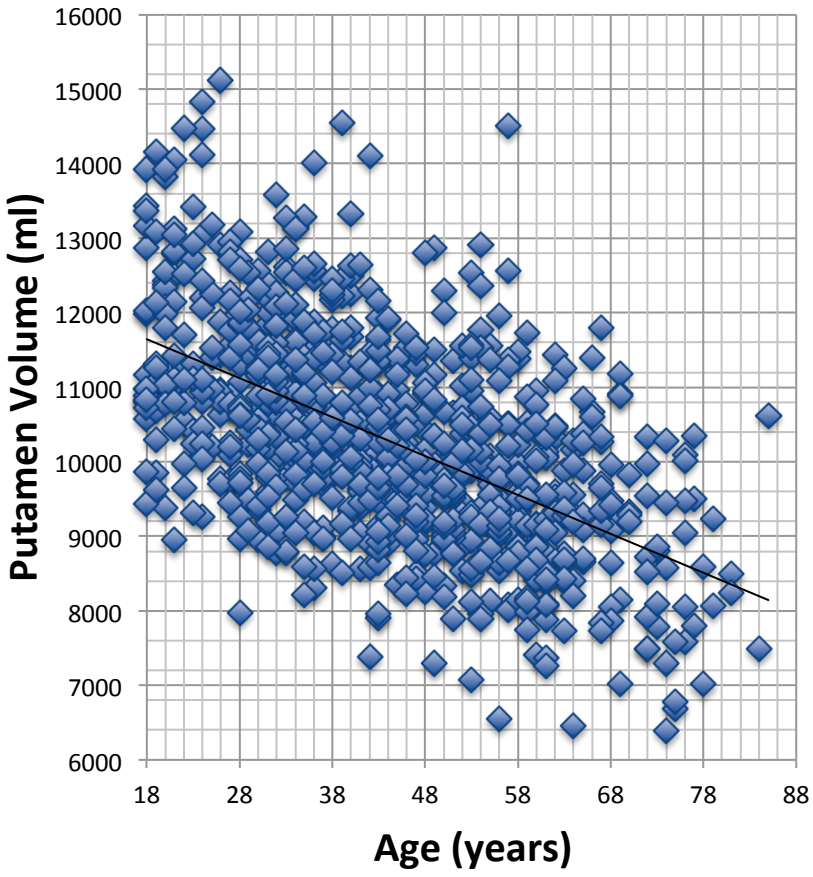
- Complete Genomics (n=607)
- Illumina (n=411)

	Complete Genomics
Total Variants	29,652,265
Non-Coding Variants	29,389,088 (99.1%)
Coding Variants	263,177 (0.9%)
Synonymous	96,436 (36.6%)
Non-Synonymous	128,653 (48.9%)
Highly Deleterious, Non-Synonymous	87,993

6,145 HD NS variants with 5 or more minor alleles

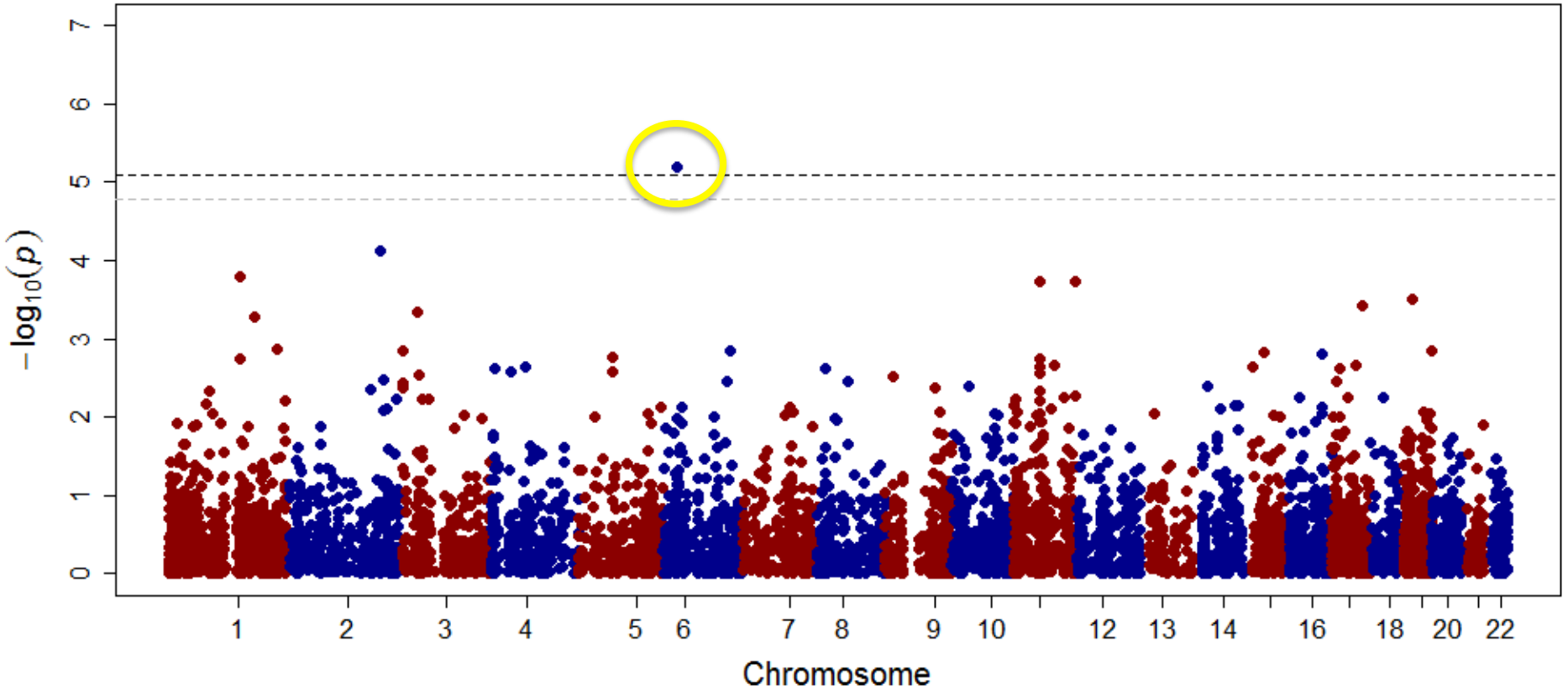
Putamen Phenotype

- Average 10,301.74 µl
- min 6,391 µl
- max 15,121 µl
- Standard Deviation 1402.71 µl
- Correlation with age, r^2=-0.55



McKay et al., *Brain Imaging Behav, 2014*

Genome-Wide Search of Highly Deleterious Non-Synonymous Variants



Bonferroni correction = 0.05/6,145 = 8.14×10^{-6}

Chromosome 6, Position 29627222

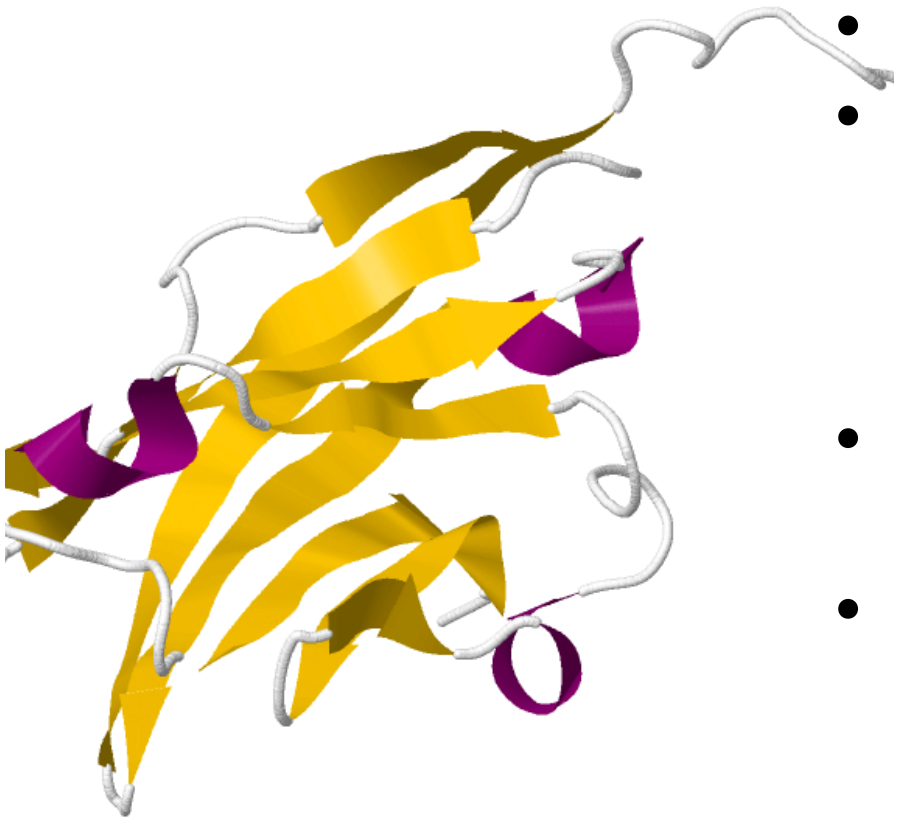
- 6p22.1 rs138169338
- $\chi^2 = 20.40$, $p = 6.27 \times 10^{-6}$
- Polyphen Score 0.994
- Missense mutation, amino acid change: uc003nmy.2:c.C215A:p.P72H
- 0.01 MAF (1000 Genome)
- 0.24 MAF in our sample
- A allele accounts for 0.46 standard deviation decrease in putamen volume (645.24ml change or ~12.5 year decrease)

6p22.1

- SNP is on the edge of the major histone compatibility region
- Five genome-wide significant variants associated with schizophrenia within 85kB:
  - rs2746149, $p=2.62 \times 10^{-8}$, distance 84kB
  - rs1233493, $p=3.33 \times 10^{-8}$, distance 61kB
  - rs1235162, $p=3.41 \times 10^{-8}$, distance 18kB
  - rs404240, $p=6.7810^{-8}$, distance 5kB
  - rs2746150, $p=8.06 \times 10^{-8}$, distance 77kB

Schizophrenia Psychiatric GWAS Consortium, *Nature Genetics*, 2011

Myelin Oligodendrocyte Glycoprotein (MOG) Gene



- SNP located within exon 1 of *MOG*
- A membrane protein expressed on the oligodendrocyte cell surface and the outermost surface of myelin sheaths
- Primary target antigen involved in immune-mediated demyelination
- This protein may be involved in completion and maintenance of the myelin sheath and in cell-cell communication

Conclusions

- QTL localization is still very relevant for detecting effects of rare variants
- Pedigrees are good for identifying specific rare variants. Need large families but every family may be informative
- Whole genome sequencing is going to be very useful for identifying genes underlying human QTLs.
- *MOG,* a novel candidate gene for putamen volume
- *NDST4*, a novel candidate gene for amygdala volume

Ackno rators

Olin Neurops l San Antonio

& Yale Univer

- Emma Knowles
- Sam Mathias
- Karen Hodgson
- Nailin Yao
- Jen Barrett

hers

y

n

kler

mann

U Texas Rio G

- Marcio de Alme
- **John Blange**
- Joanne Curran
- Thomas Dyer
- Ravi Duggirala
- Juan Peralta

v

ck

Financial Supp