A non-human primate GTEx

Nelson Freimer

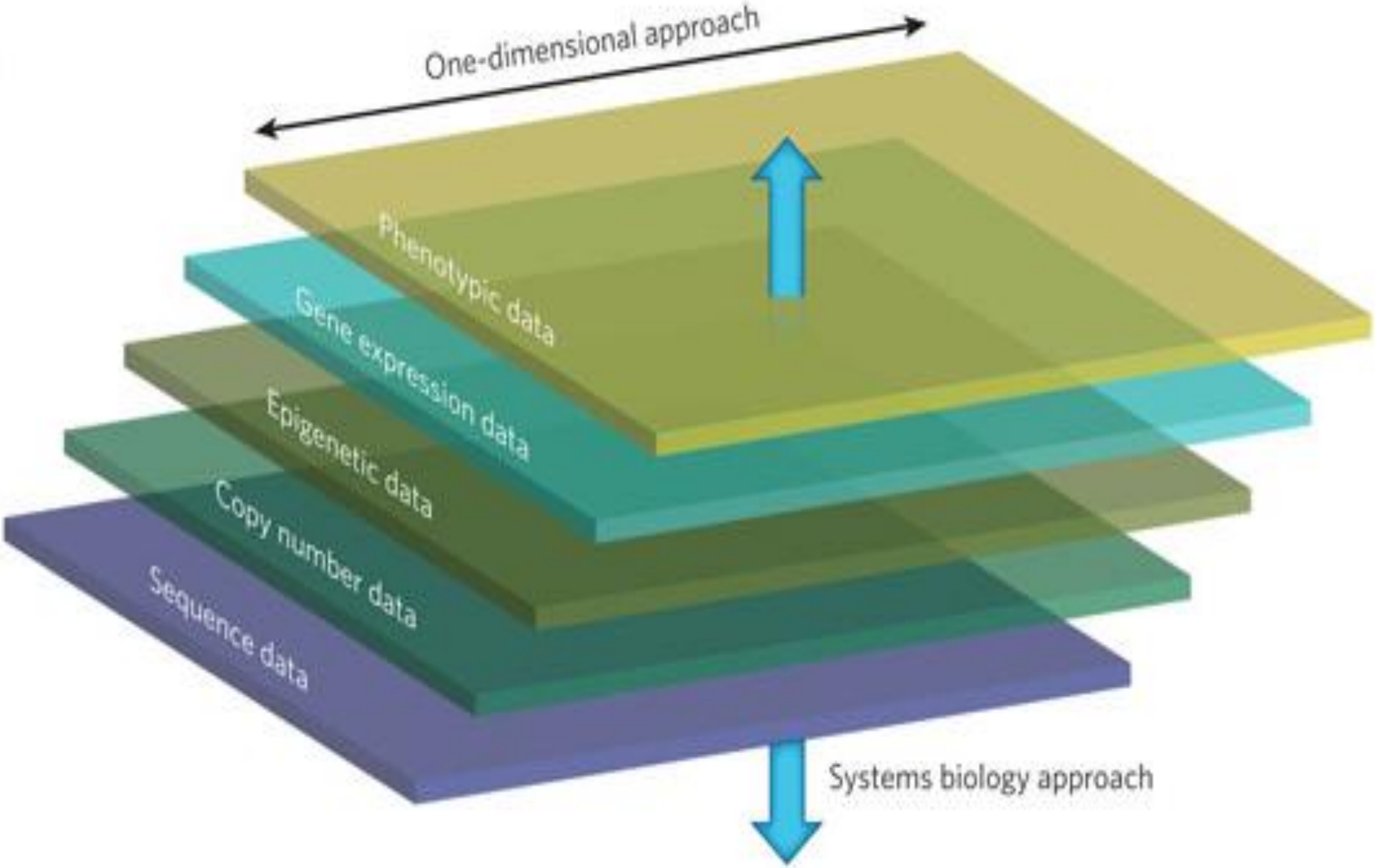
UCLA

PGC, December 9, 2016

Outline

- The need for non-human primate (NHP) systems biology models
- Introduction to the vervet monkey system
- Vervet sample and genomic resources
- Vervet multi-tissue transcriptome & eQTLs
  - Caudate, hippocampus, pituitary, adrenal, fibroblast, blood
- Future NHP GWAS

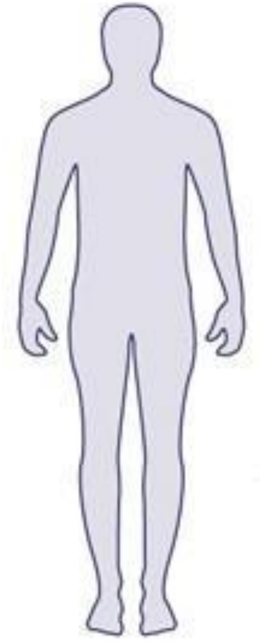
Systems biology offers a framework for understanding the genetic contribution to complex brain phenotypes



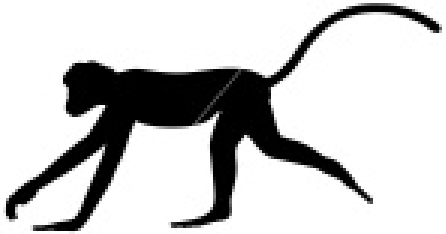
Geschwind & Konopka Nature, 2009

NHPs provide uniquely valuable systems biology models

- Environmental variability
- Limited feasibility of:
    - Longitudunal studies
    - Invasive studies



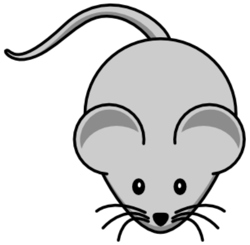
≈



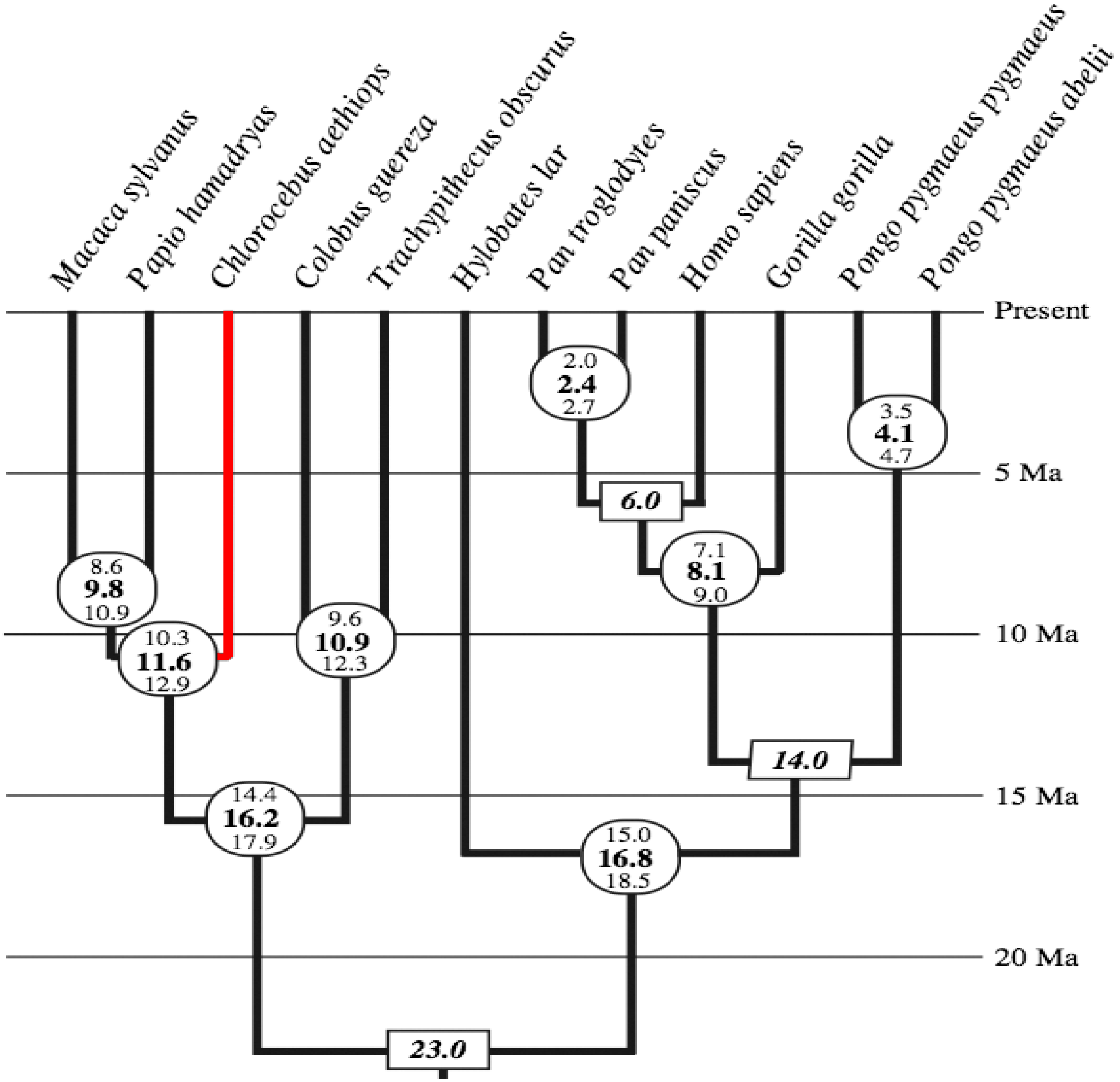
- Low sequence and synteny divergence
- Functional conservation
- Conservation of brain and behavior
- Ecological conservation
- Controlled environment
- Longitudinal studies
- Invasive studies

≠

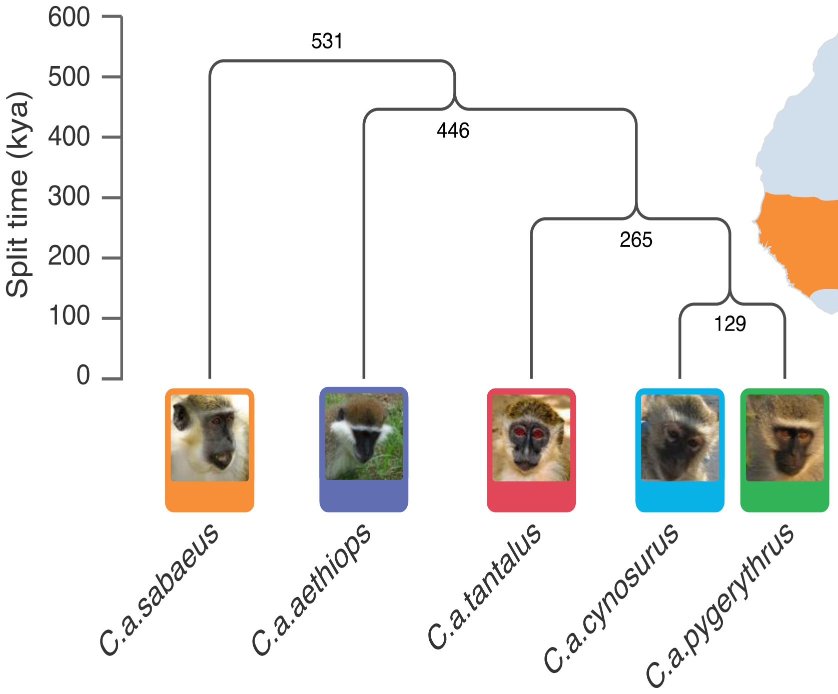
- Sequence and synteny divergence
- Functional divergence
- Substantial brain and behavior differences
- Different ecologic niche

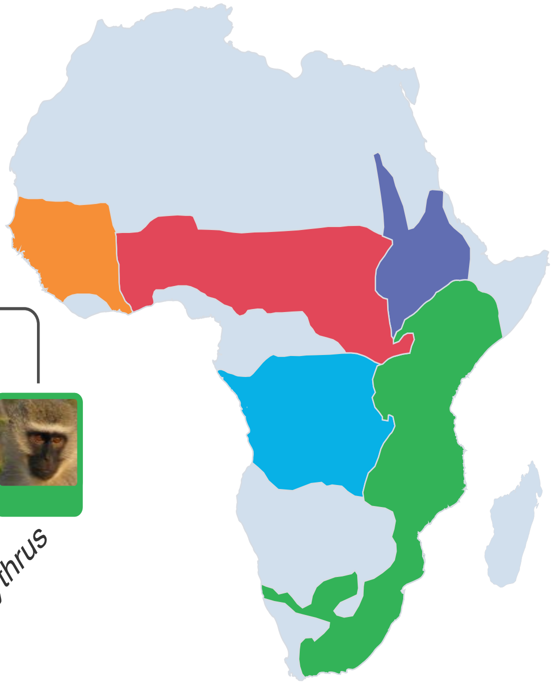


Vervet monkey model (African green monkey)

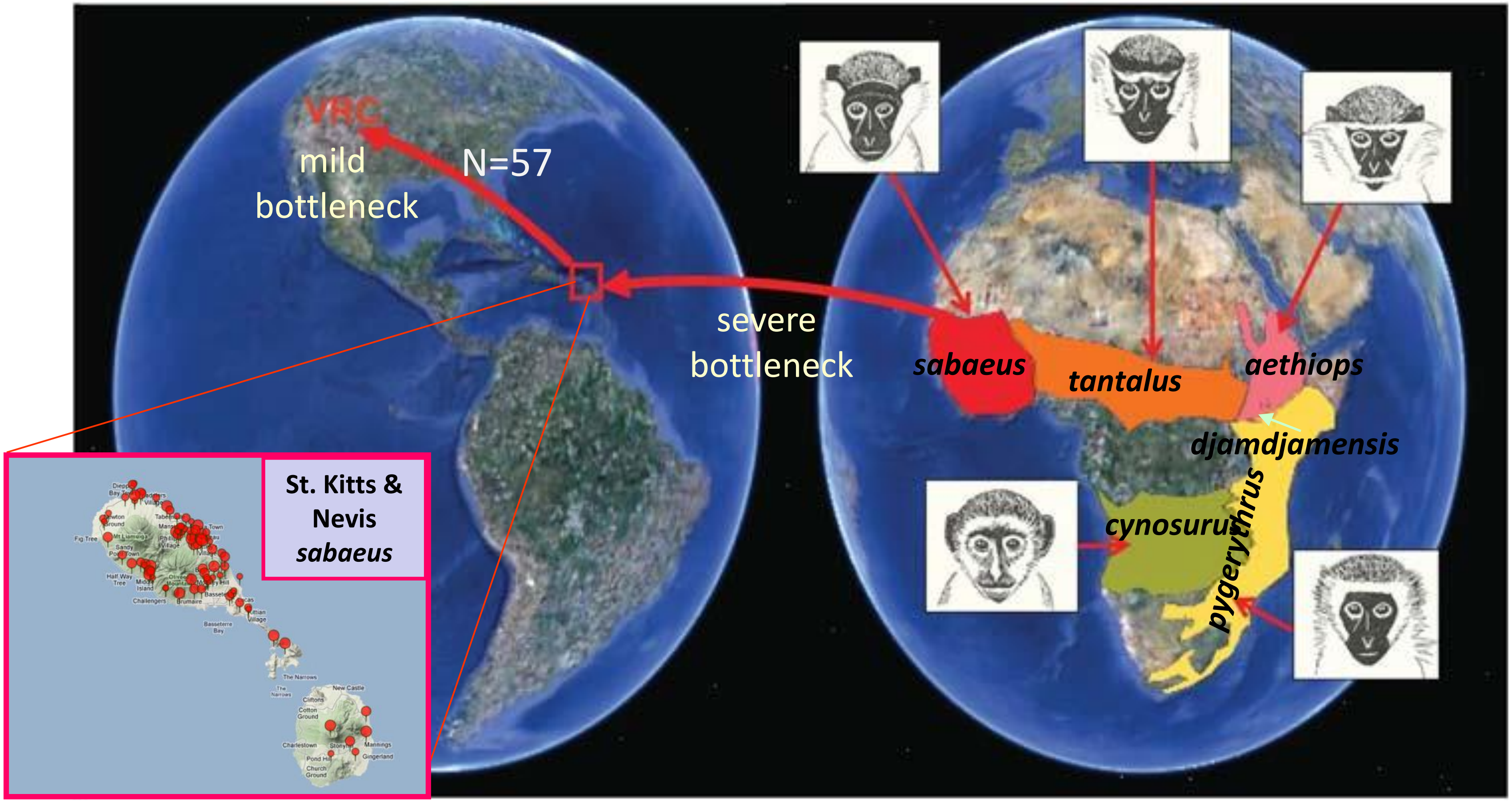


Distribution of vervet subspecies across Sub-Saharan Africa





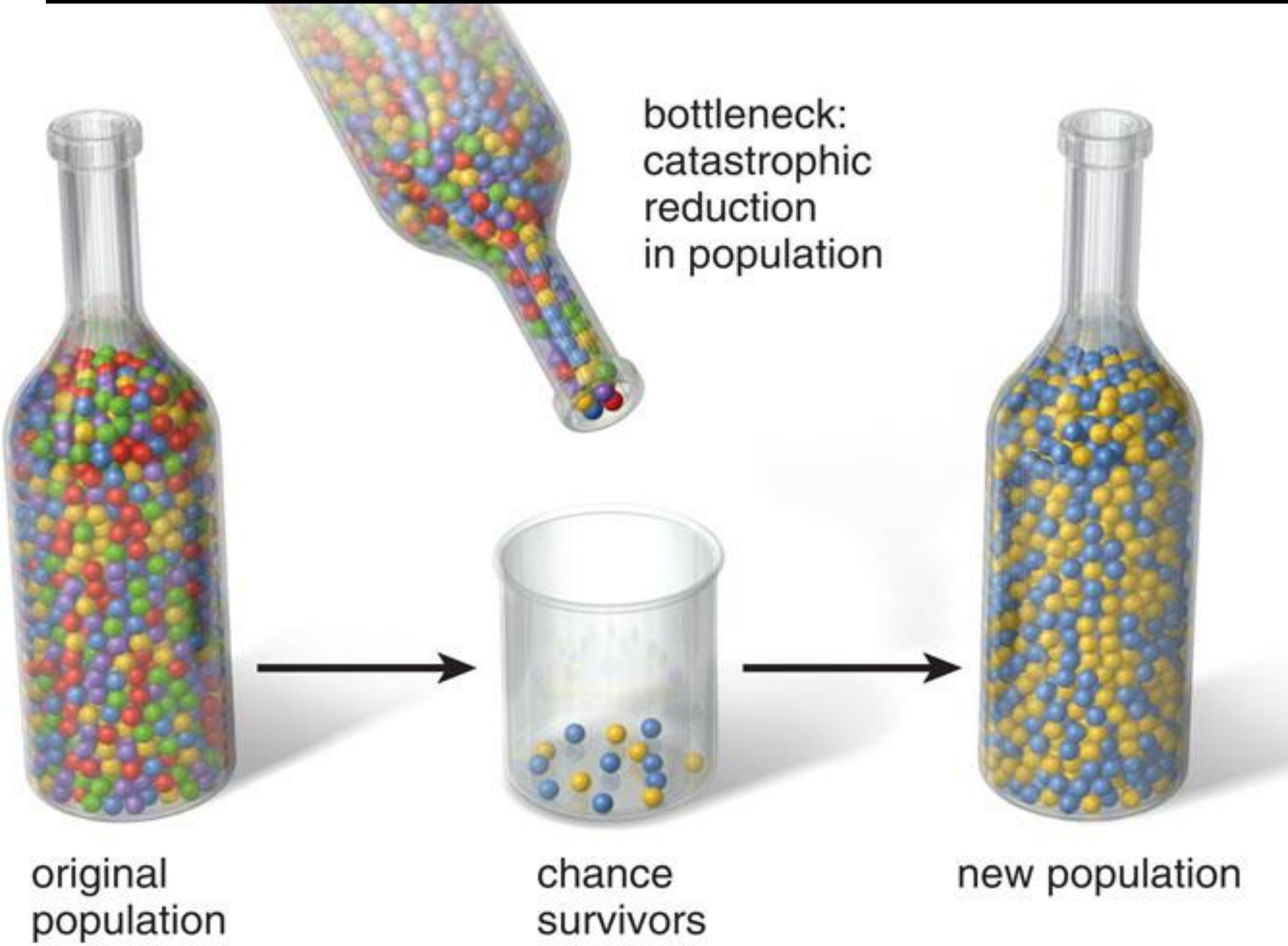
The Caribbean vervet as a genetic model: a bottlenecked founder population



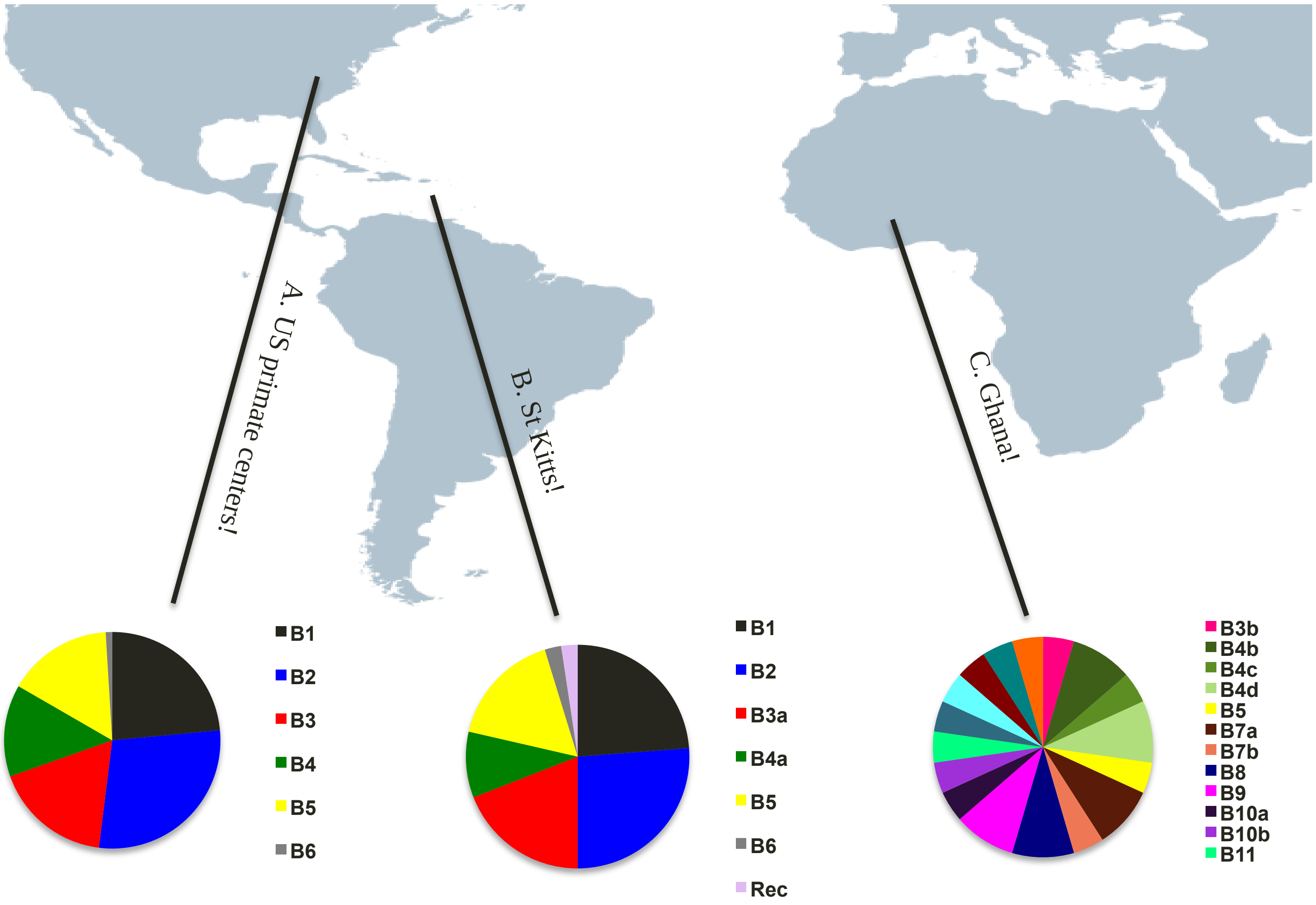
VRC = Vervet Research Colony

Jasinska *et al.* **Hum Mol Genet** 2012

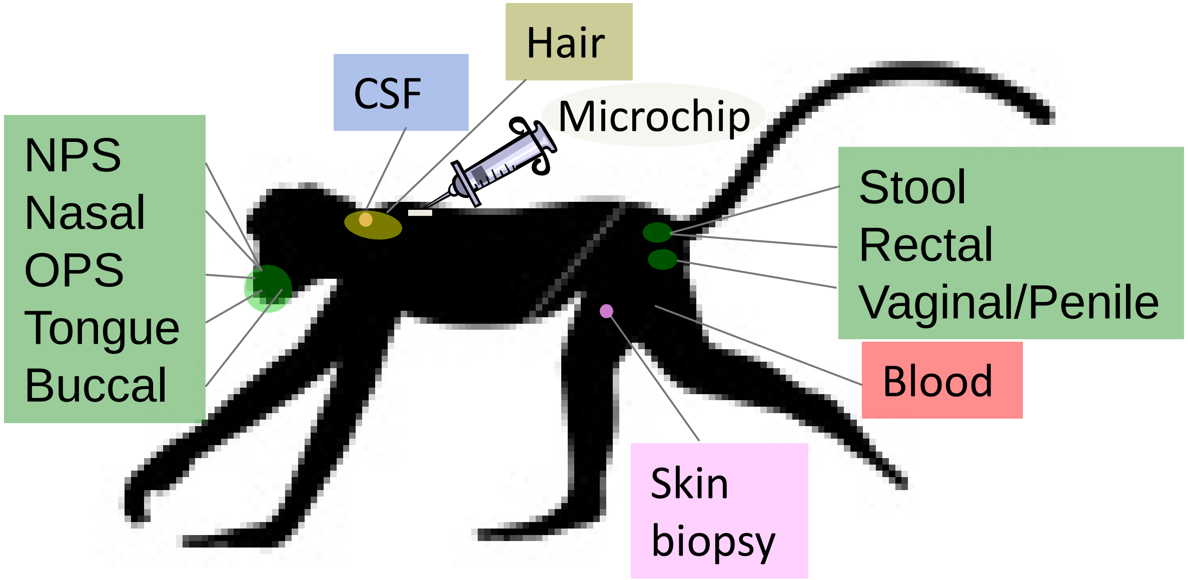
Founder populations:
The impact of growth from a bottleneck



Reduced MHC diversity: severe vervet Caribbean bottleneck



Biomaterial and data repository from world vervet populations

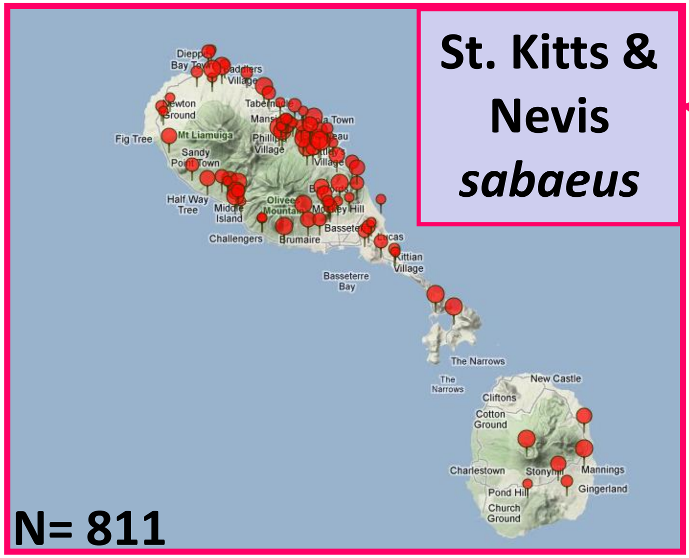


- General health state
- Reproductive status
- Body measurements
- Morphometrics
- Dentition based age
- Multi-tissue RNA, DNA
- Fibroblast cell lines

Jasinska *et al.* **Hum Mol Genet** 2012

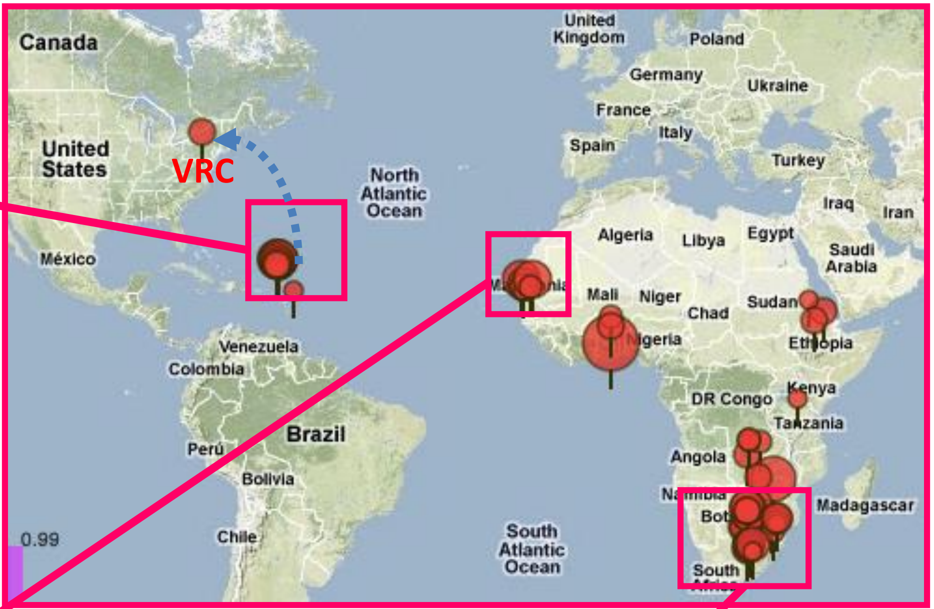
> 1,500 Wild vervets now sampled

(~130 with whole genome sequencing, WGS)



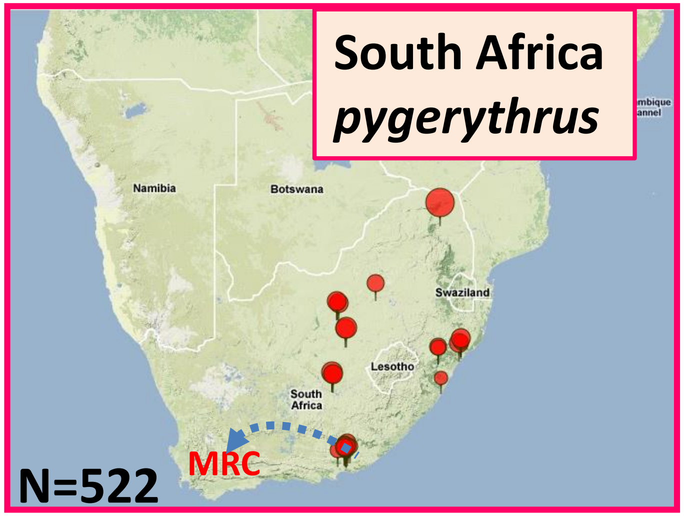
St. Kitts & Nevis
sabaeus

N= 811



Gambia
sabaeus

N=156

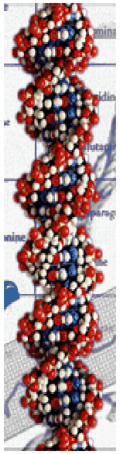


South Africa
pygerythrus

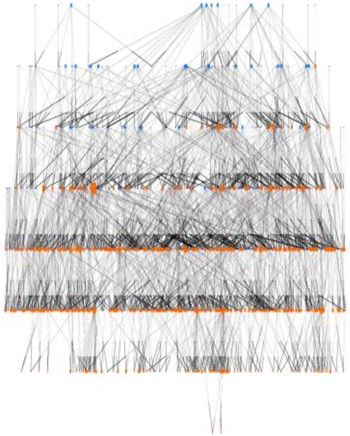
N=522

Key resources of the Caribbean vervet genetic model

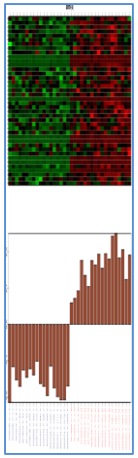
- Reference genome assembly (Warren et al 2015)
- Large pedigree samples, Vervet Research Colony, VRC
- Larger population samples, St. Kitts/Nevis, SKN (N~ 50,000)
- Genetic variation, Whole Genome Sequencing, WGS (VRC, N = 721)
- Functional genomics: Multi-tissue/across-development transcriptomes (N=90)



Vervet 1.1 reference

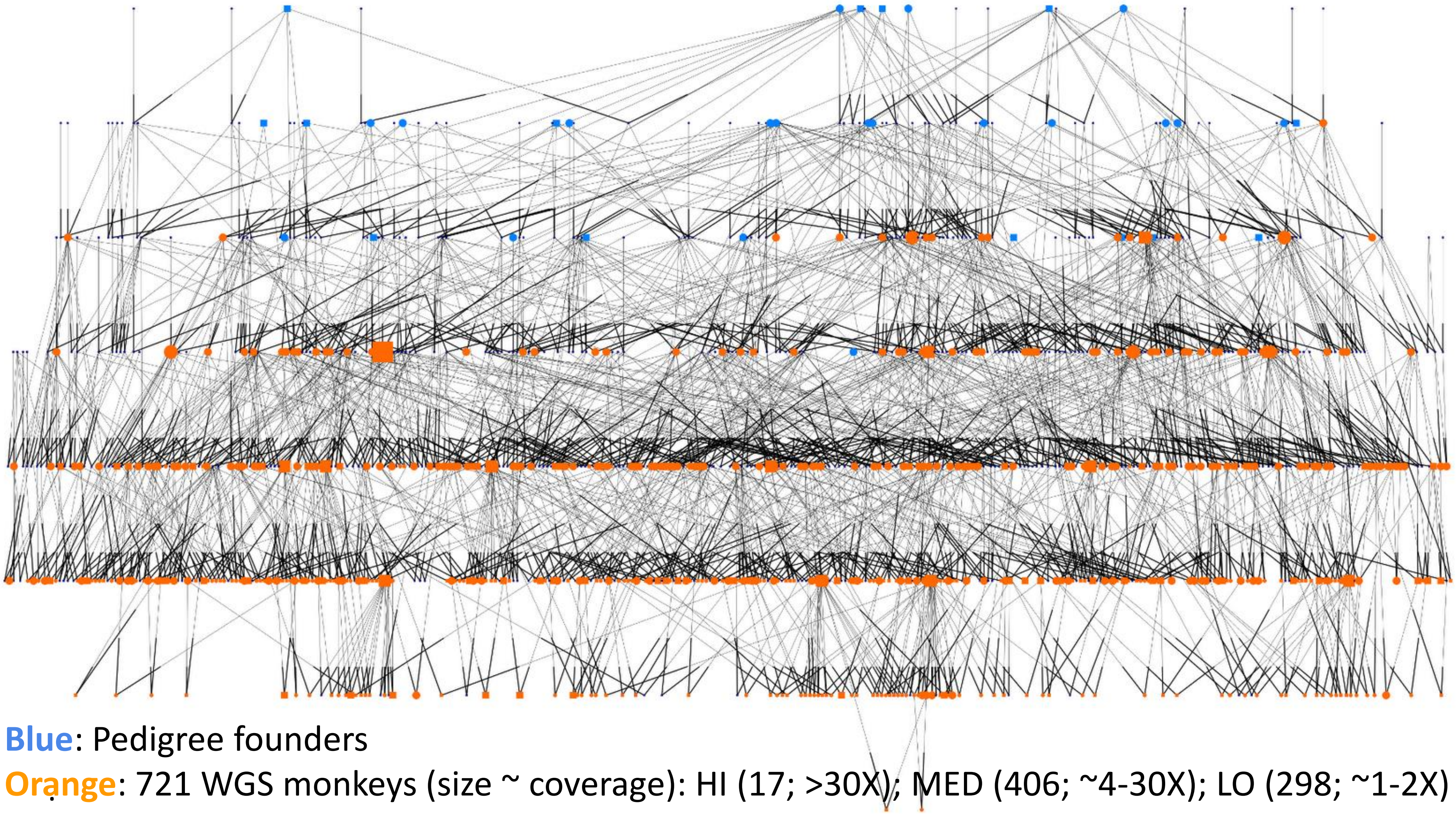


VRC Seq



VRC RNA

VRCseq: Comprehensive variant data on 721 monkeys



Blue: Pedigree founders

Orange: 721 WGS monkeys (size ~ coverage): HI (17; >30X); MED (406; ~4-30X); LO (298; ~1-2X)

Huang et al., 2015

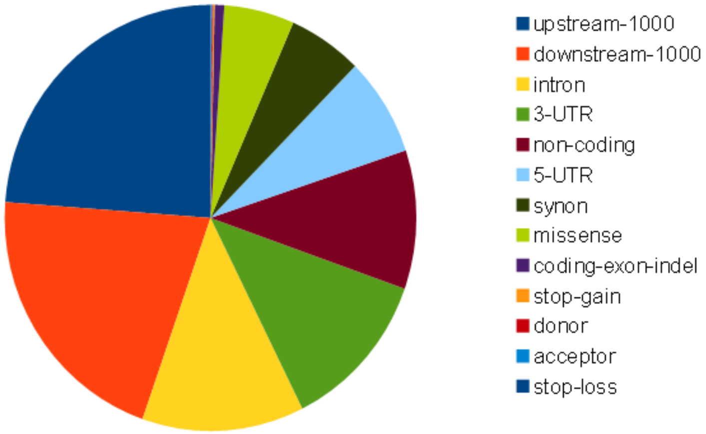
Sequencing in the VRC

-- WGS of >720 samples with varying coverage (1...30x)
-- Reference genome *C.sabaeus* 1.1: 29 + 2 chr
-- Gene sets: NCBI Annotation Release 100, ENSEMBL v.85

Workflow:
-- Raw variant calling with *GATK*, genotype refinement in trios
-- Postprocessing: genotype conflicts, Mendelian errors, low qual
-- Phasing in 99 = 82 HC + 17 LC samples with *Beagle*
-- Phasing and imputation in 620 LC, 99 as reference haplotypes
-- Postprocessing: Mendelian errors, QC, quality flags

-- Two independent call sets: 16.7 mln SNVs genomewide,
1.3 mln extended exome SNVs and indels

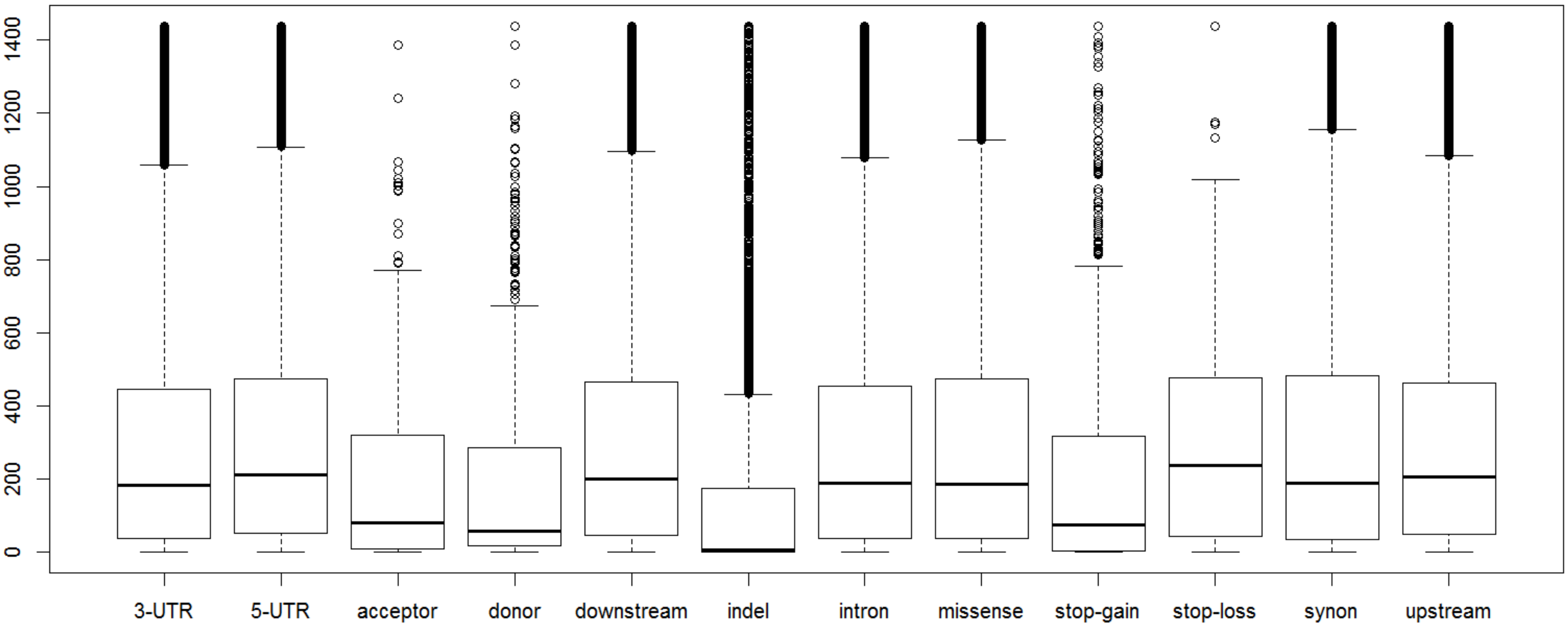
Exonic variant annotation in VRC



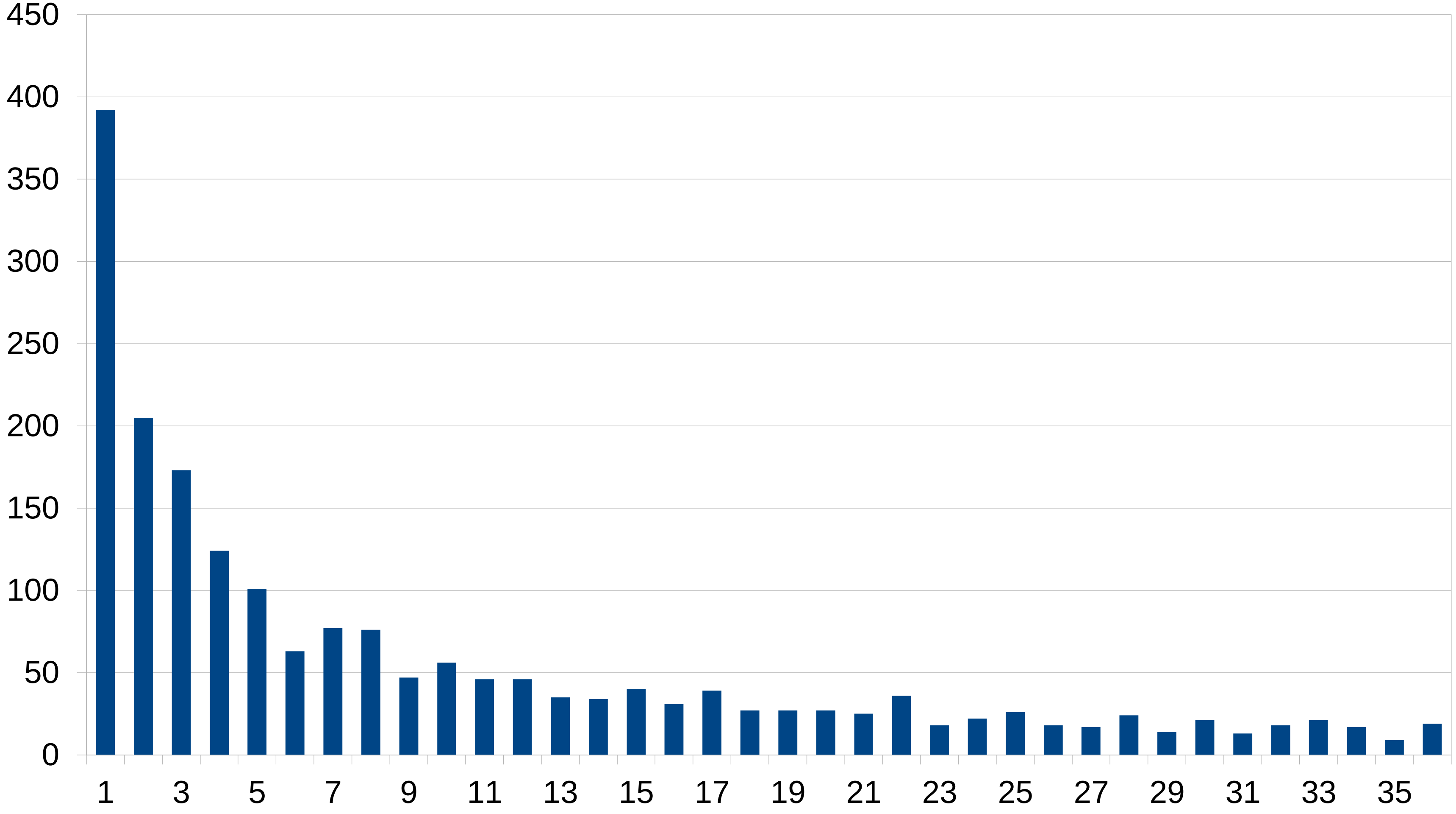
COMPLEX	47,964	3.7
DEL	128,967	10.0
INS	64,717	5.0
SNV	1,051,886	81.3

Type	**SNVs**	**Indels**	**Total**	**%**
Acceptor	520	604	1,124	0.1
Coding-exon-indel	N/A	9,930	9,930	0.8
Donor	800	507	1,307	0.1
Downstream-1000	212,581	56,487	269,068	20.8
Intron	130,781	34,320	165,101	12.8
Missense	70,212	N/A	70,212	5.4
Non-coding	113,244	22,665	135,909	10.5
Stop-gain	1,416	N/A	1,416	0.1
Stop-loss	131	N/A	131	0.0
Synon	76,469	N/A	76,469	5.9
Upstream-1000	242,610	65,035	307,645	23.8
3-UTR	122,669	35,804	158,473	12.3
5-UTR	80,453	16,296	96,749	7.5
Total:	**1,051,886**	**241,648**	**1,293,534**	

Alternative allele counts in VRC



hLoF occurrence in VRC



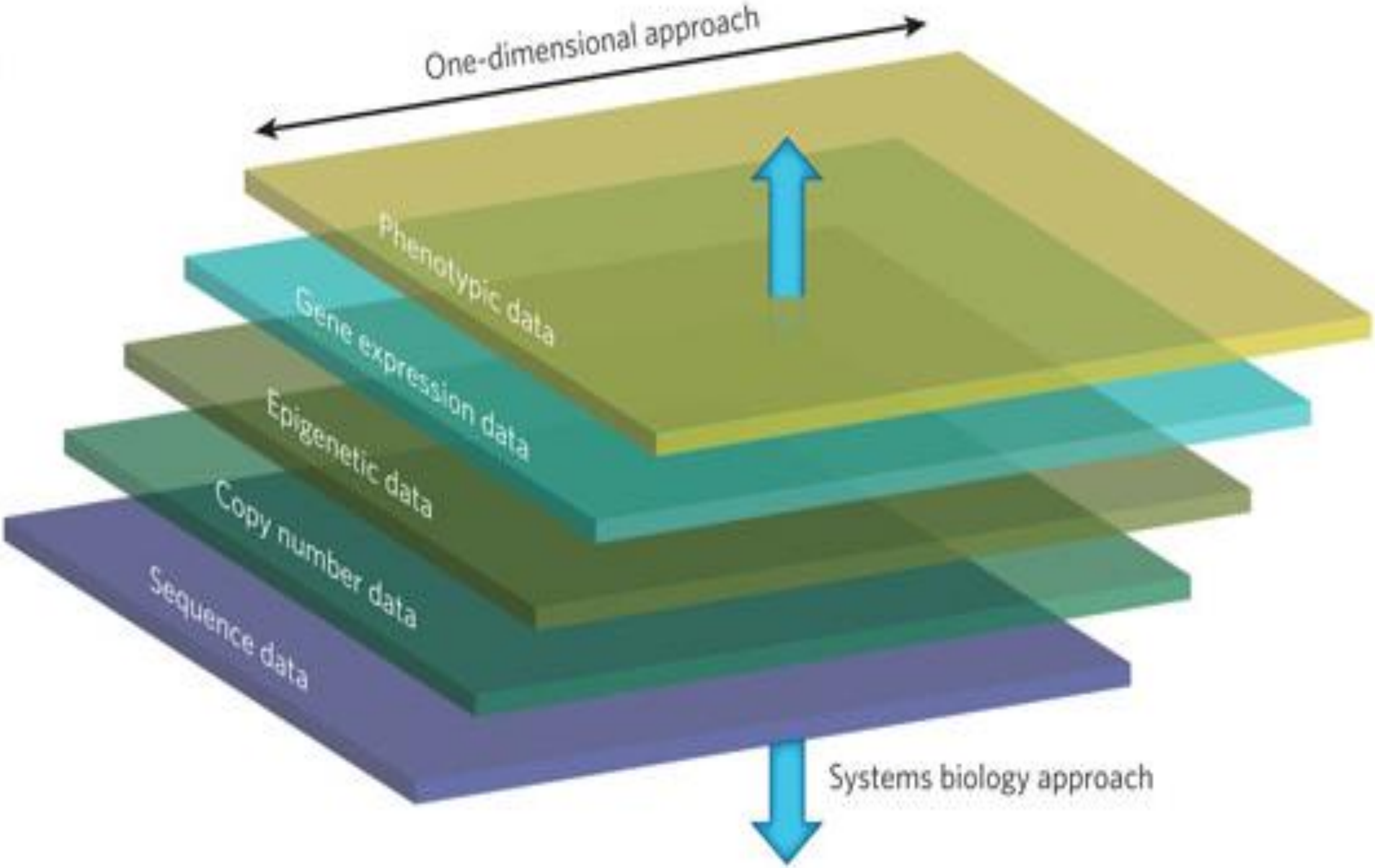
Example phenotype: Early infant mortality

- 320 female monkeys with >1 birth
- Overall neonatal/1st month mortality rate: 0.323
- 18 monkeys with significant excess EIM
  - 16 “unique” high-quality hLOFs
  - Most seen in only one mother
  - Stop-gain variant in *ARPC3*
    - In 2 mothers (each w 4/4 EIM) + one female with no live births

Vervet common variants for genetic studies

- Focused on high-quality, highly polymorphic, sites in animals sequenced at 30x: *4.2M sites*
- Called these sites + additional QC in 721 animals: *3.4M sites*
- “LD-pruned” to obtain “GWAS” set: *500K SNPs*
    - Average density: 200 SNPs/Mb
    - Average heterozygosity: 0.45

Vervet systems biology to elucidate phenotypes: illustration from gene expression data



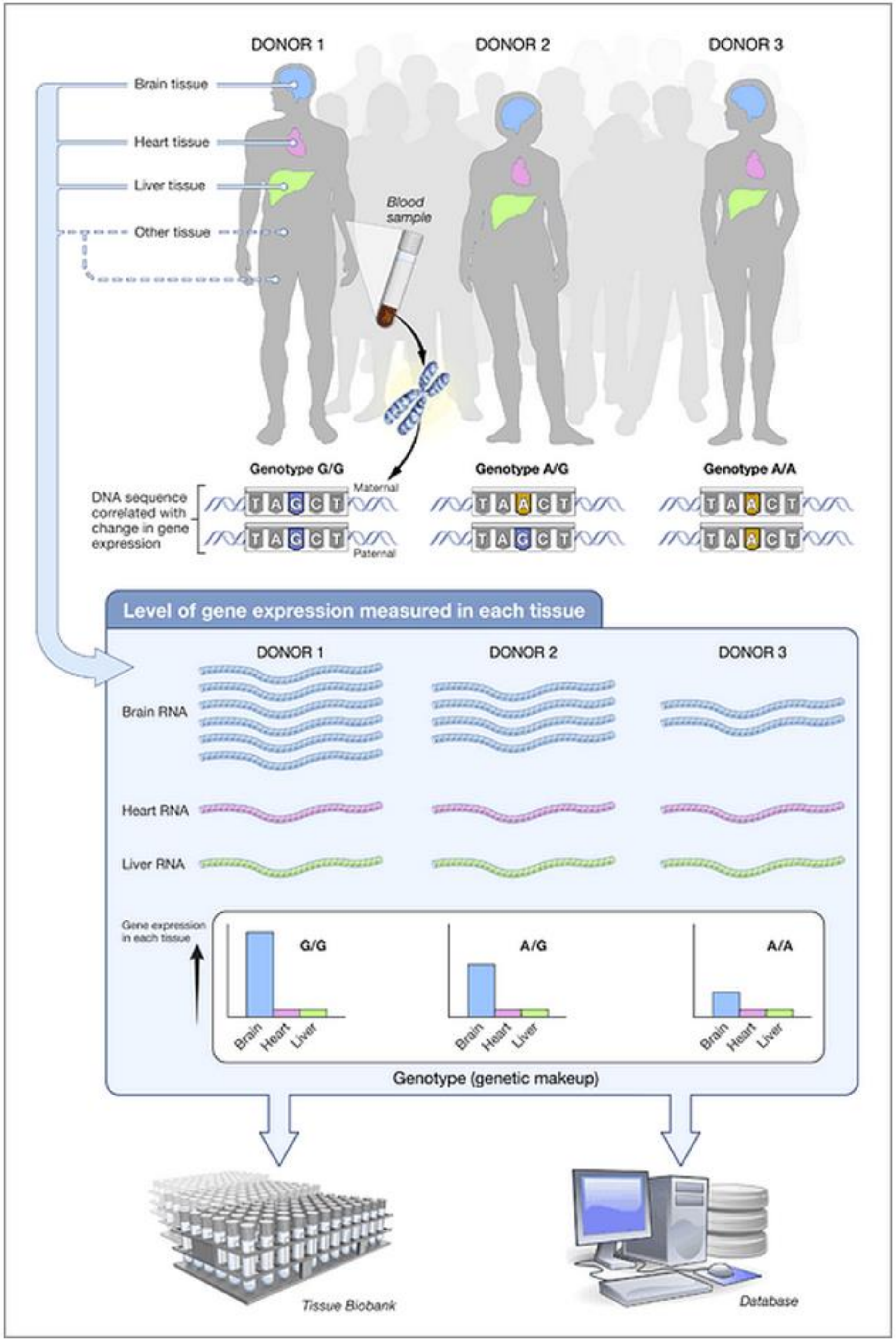
Geschwind & Konopka Nature, 2009

Genetic regulation of variation in gene expression (*expression quantitative trait loci, eQTL*)

- Trait-associated SNPs in human GWAS are highly enriched for eQTL (Nicolae *et al.*, 2010)

- Tissue-specific eQTL help identify causal SNPS at GWAS loci (Gibson *et al.* 2015)

- Genotype- Tissue Expression Project (GTEx) established to identify tissue specific eQTL



GTEx Features:

NIH funded with a very large team

- Large N (> 500)
- Mostly > 50 y.o.
- Ethnically heterogeneous
- Heterogeneous exposures
- Little phenotype data

Vervet “GTEx”:

Developmental Cross-Tissue RNA-Sequencing to Elucidate Phenotype and Genotype Data

90 vervets at 15 developmental stages: (birth to adulthood)

3 males

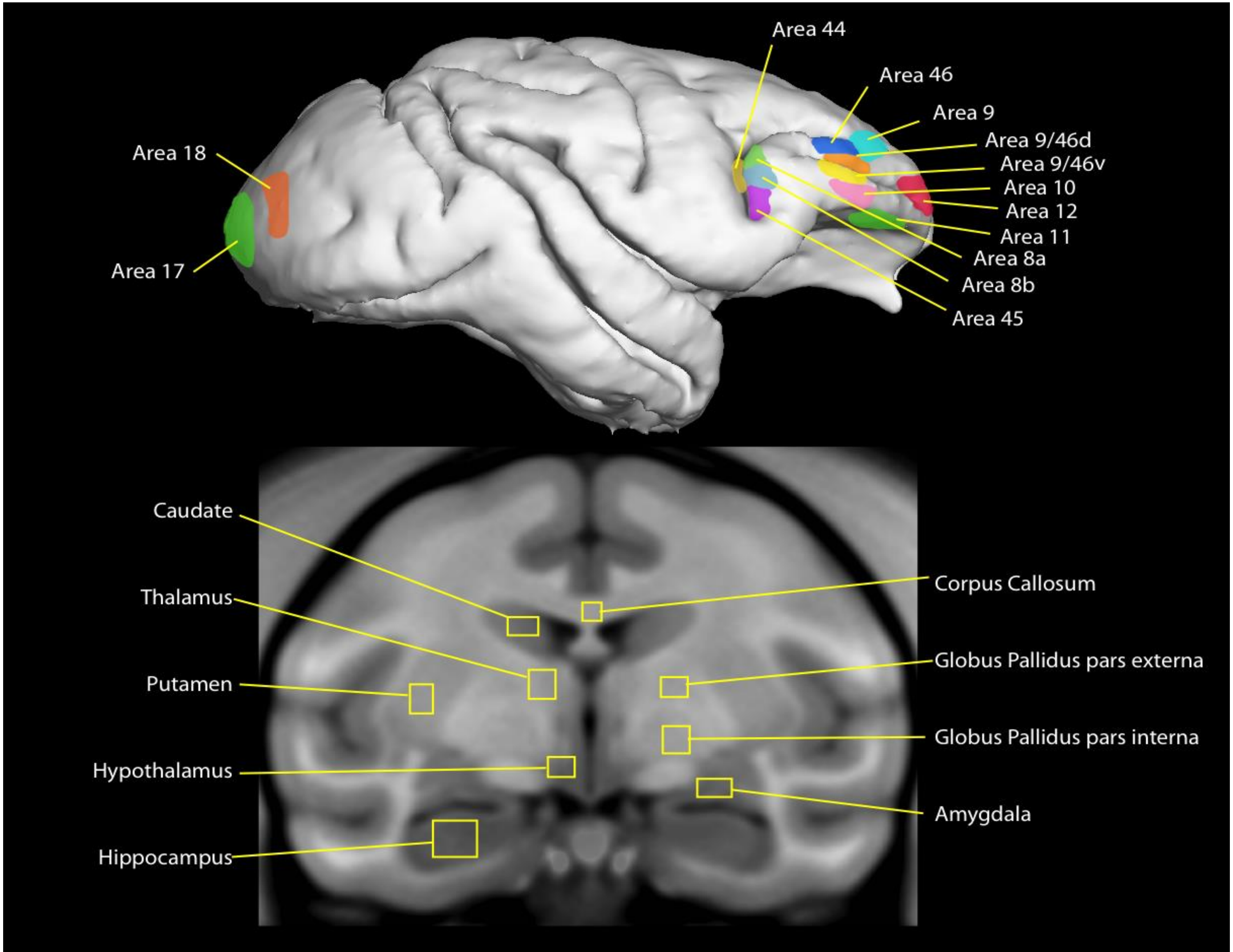
3 females

Tissues

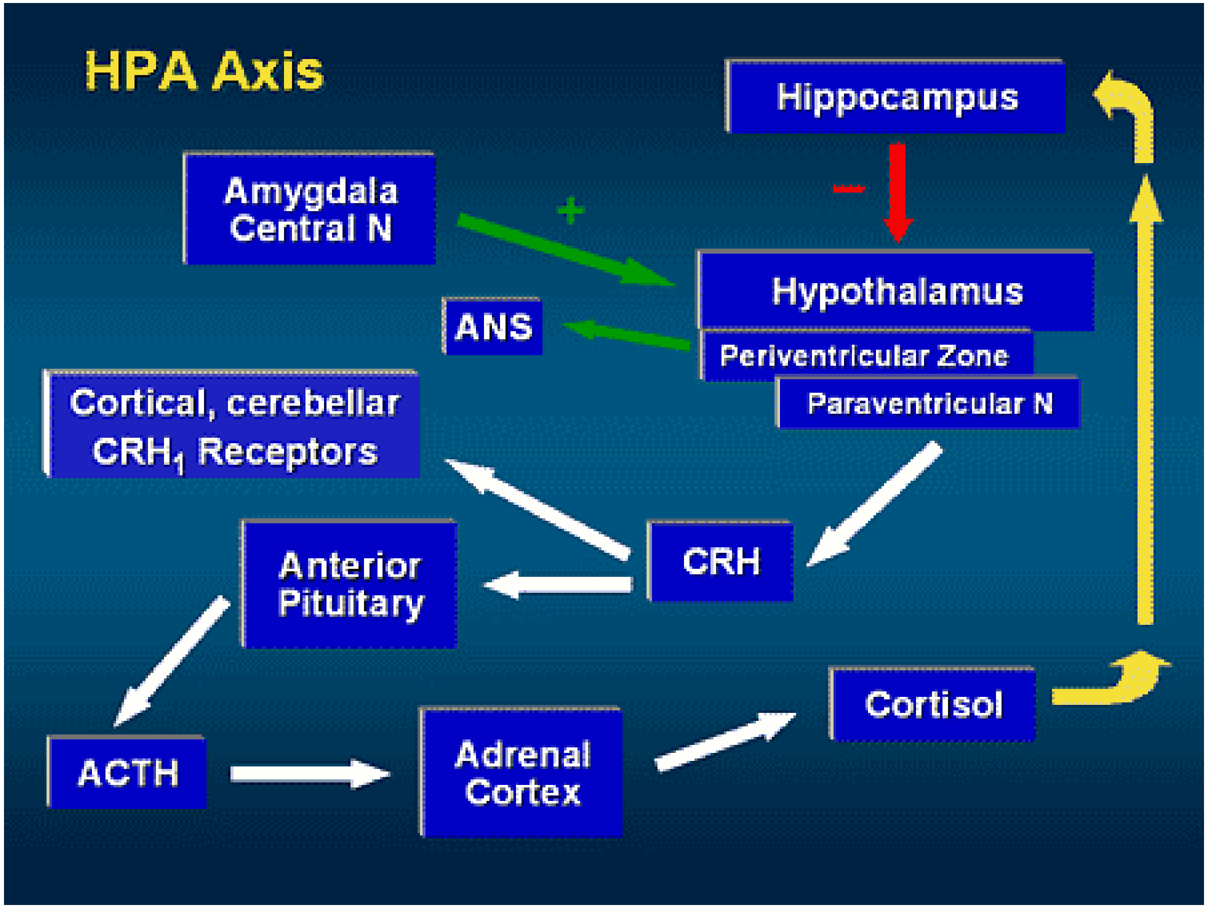
87 tissue specimens per animal

- •61 from brain
- •26 from periphery

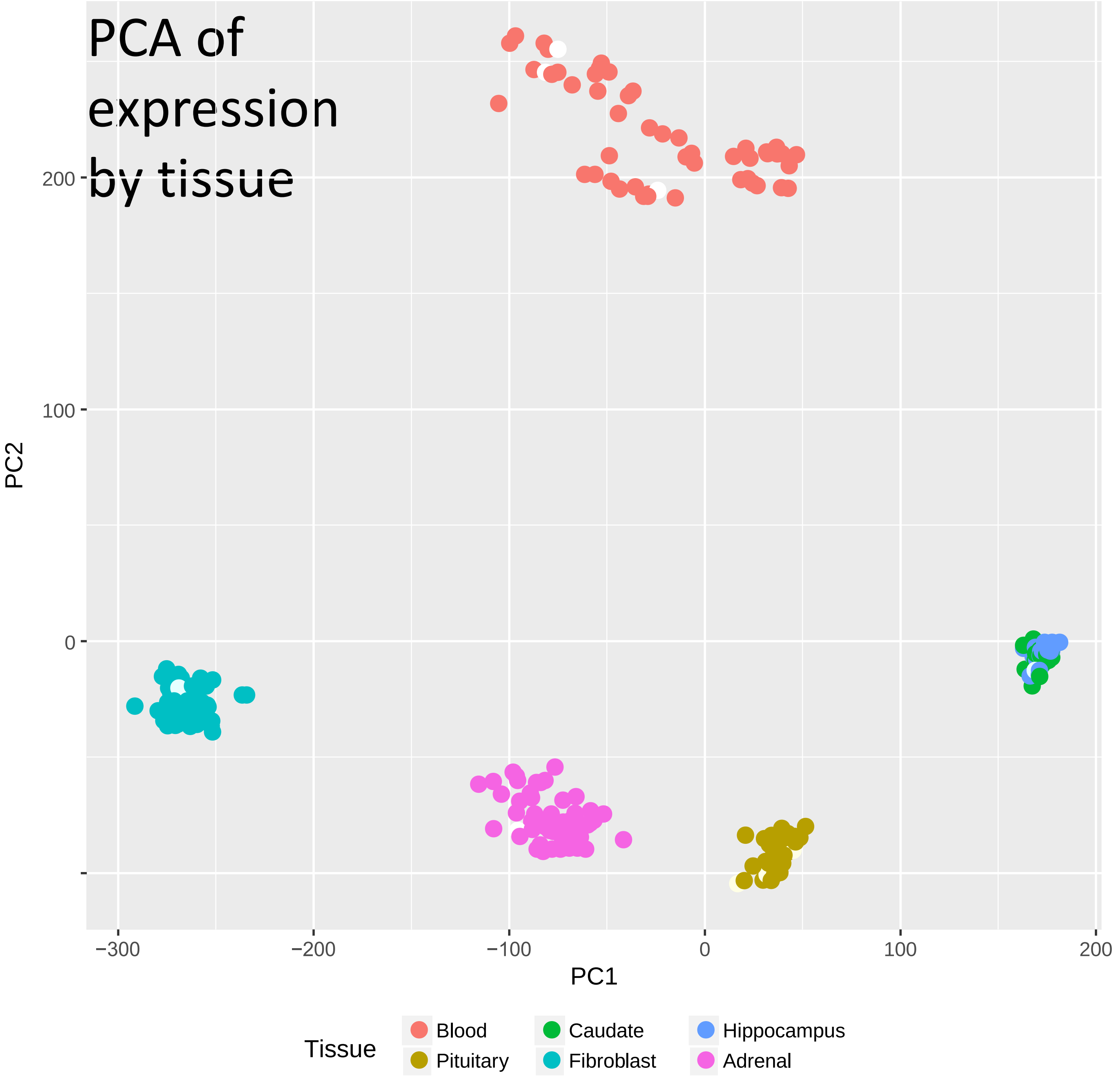
Vervet brain samples for transcriptome analysis



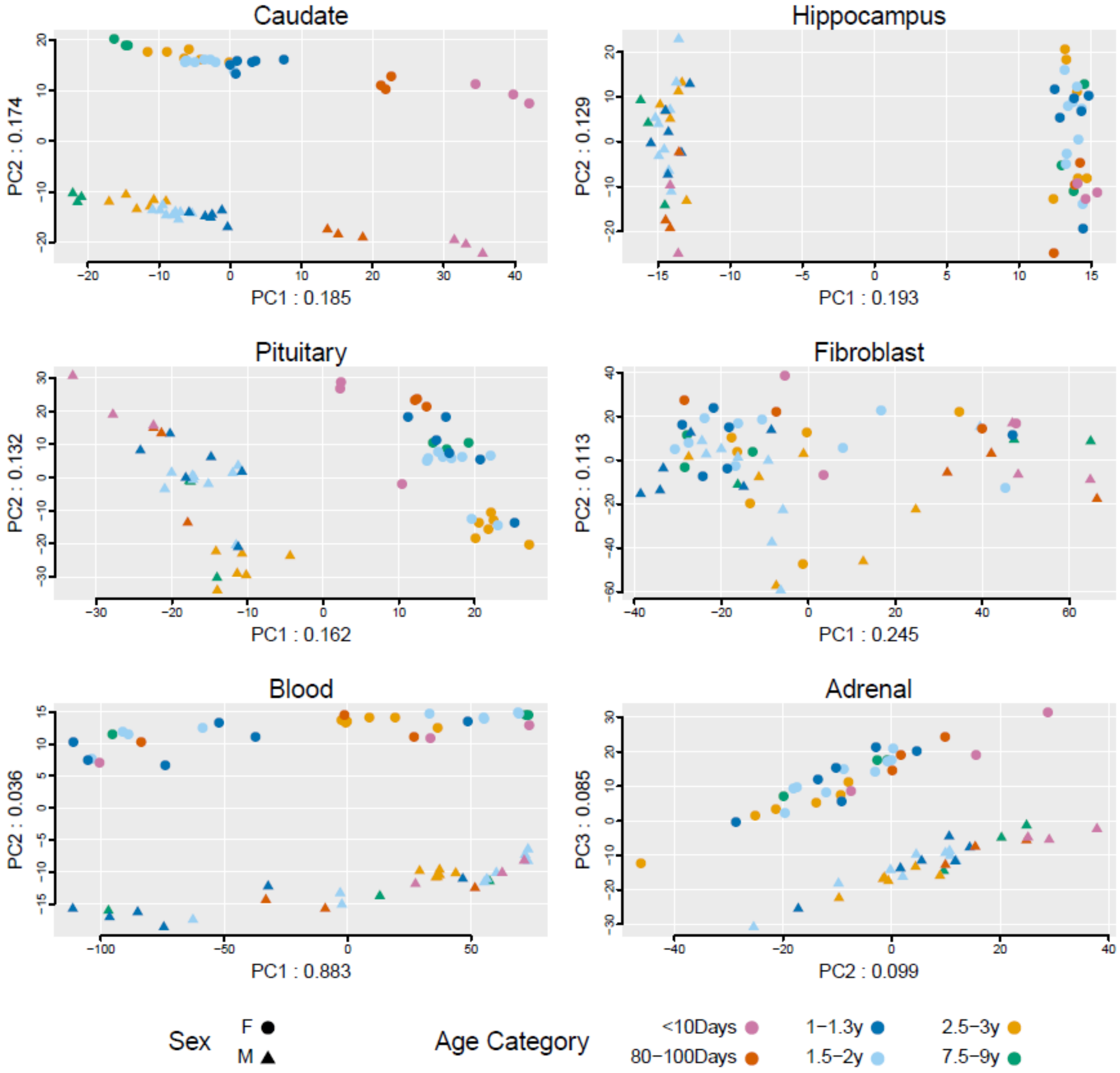
Initial Focus on Tissues Relevant to Hypothalamo-Pituitary-Adrenal (HPA) Axis (*Stress Response and Neurobehavioral Disorders*)



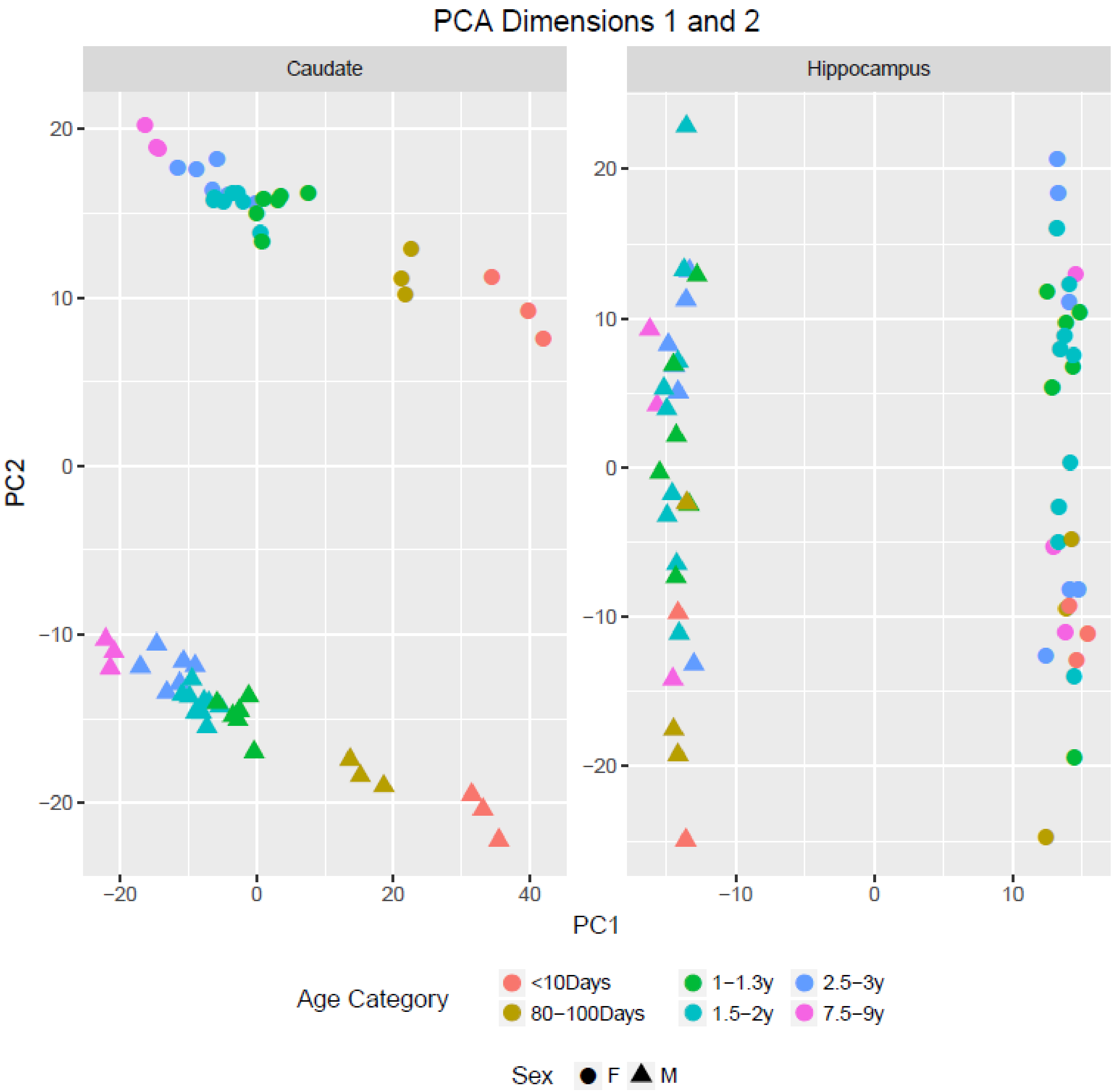
PCA of expression by tissue

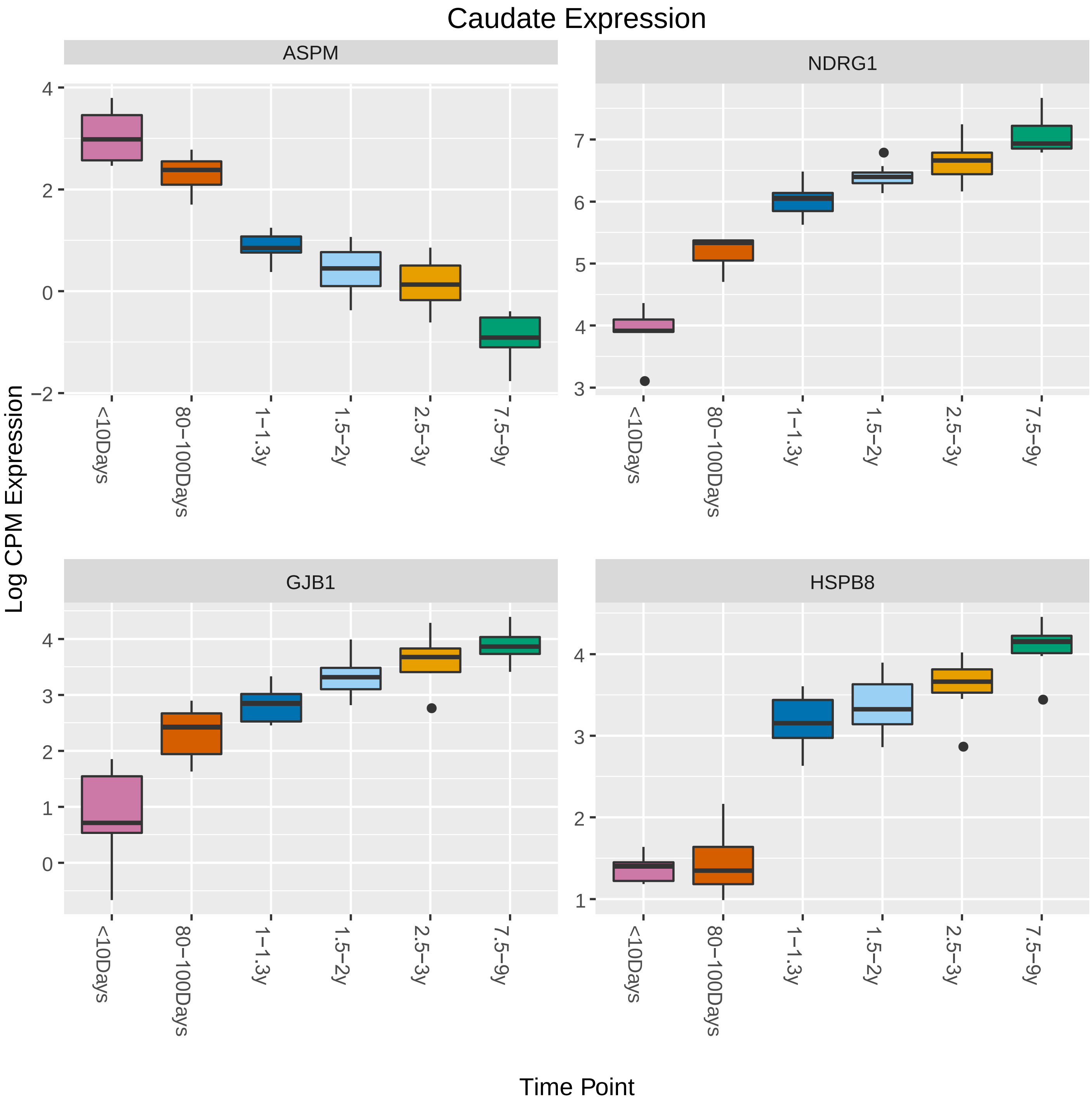


PCA of expression of most variable genes by sex & age-point

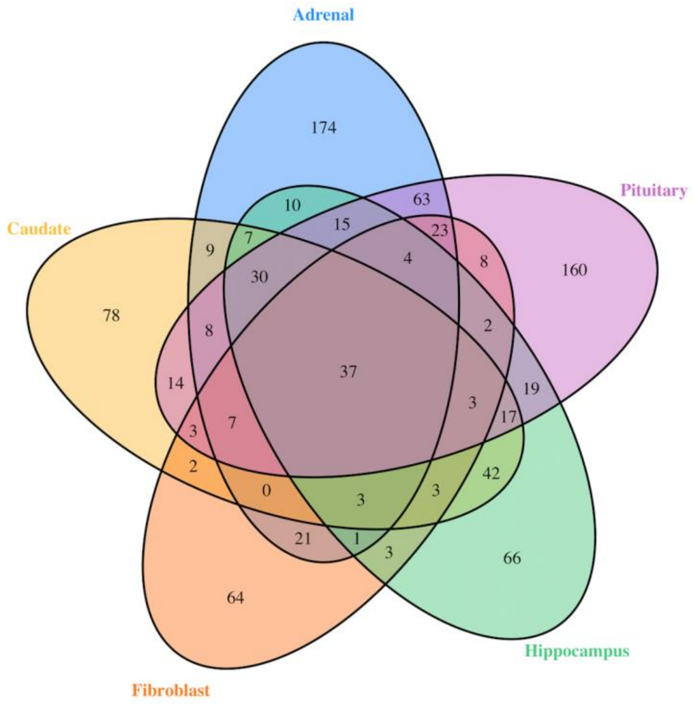


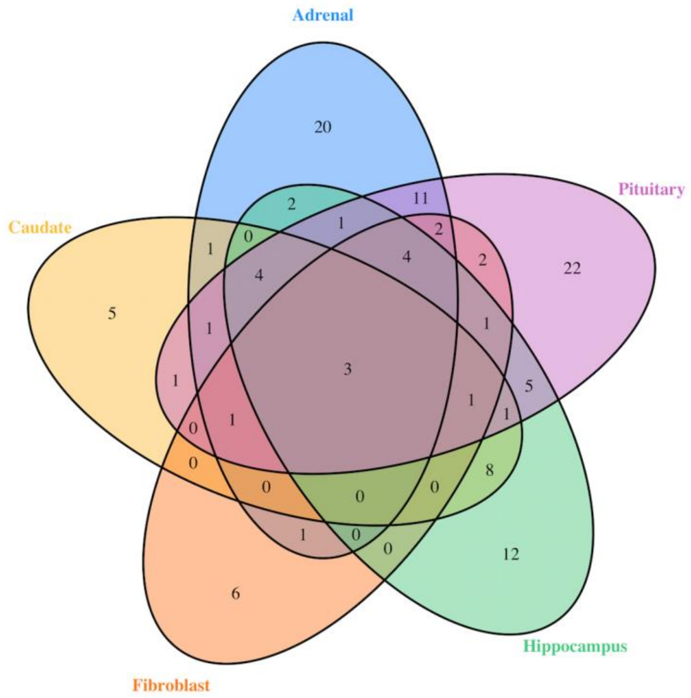
Caudate differential expression: by age/sex
Hippocampus differential expression: by sex only




Caudate Expression
ASPM
NDRG1
GJB1
HSPB8
Log CPM Expression
Time Point
4
2
0
−2
7
6
5
4
3
4
3
2
1
0
4
3
2
1
<10Days
80−100Days
1−1.3y
1.5−2y
2.5−3y
7.5−9y

Vervet genes regulated by local (left) and distant (right) eQTL across tissues

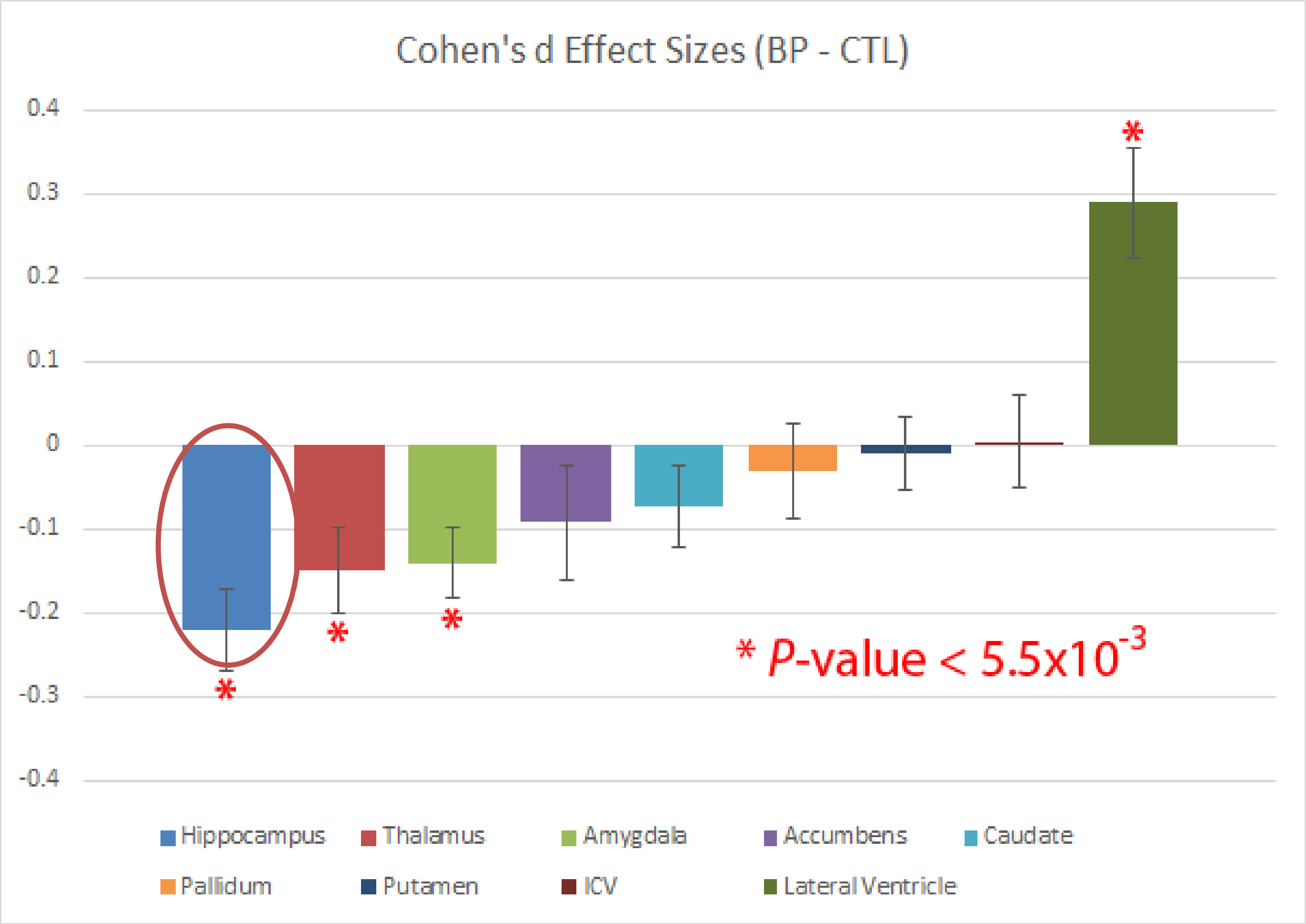




Vervet eQTL to identify causal variants for vervet phenotypes:

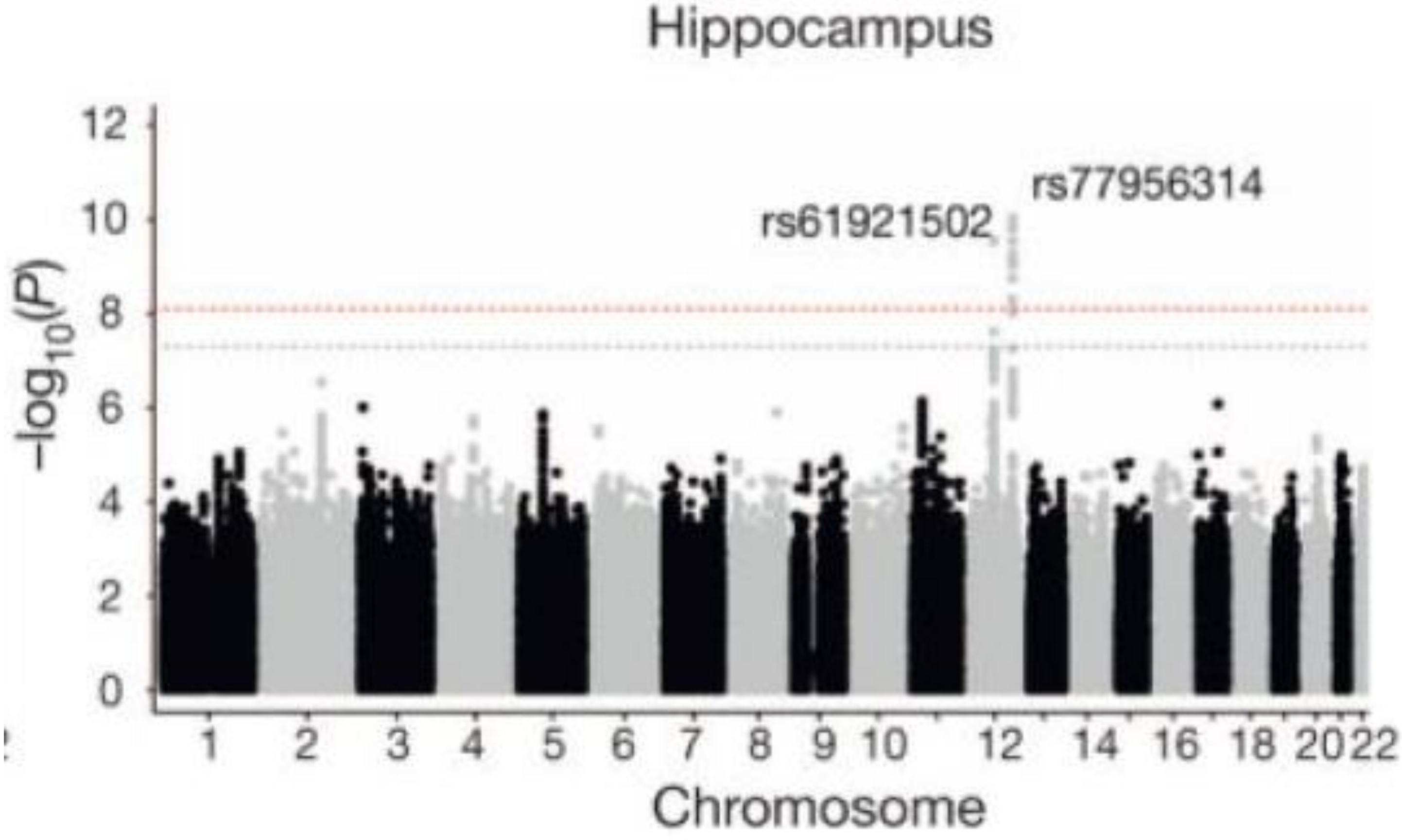
Hippocampus-specific eQTL and hippocampal volume

Hippocampal Volume is Highly Associated with Neuropsychiatric Diseases (e.g. *Bipolar Disorder*)



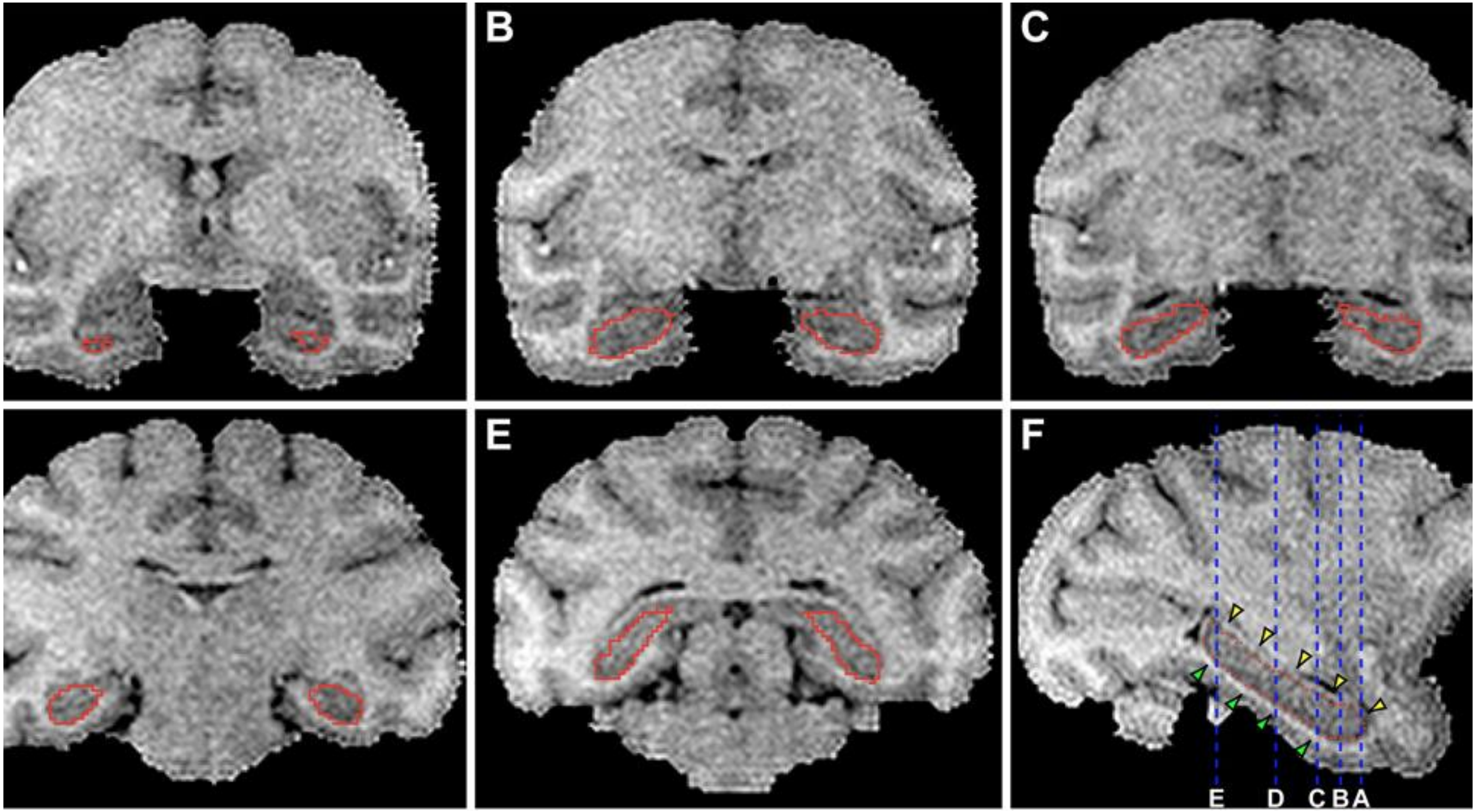
D. Hibar *et al*/ENIGMA Consortium, unpublished data

Hippocampal volume GWAS: significant associations in N > 30,000 people



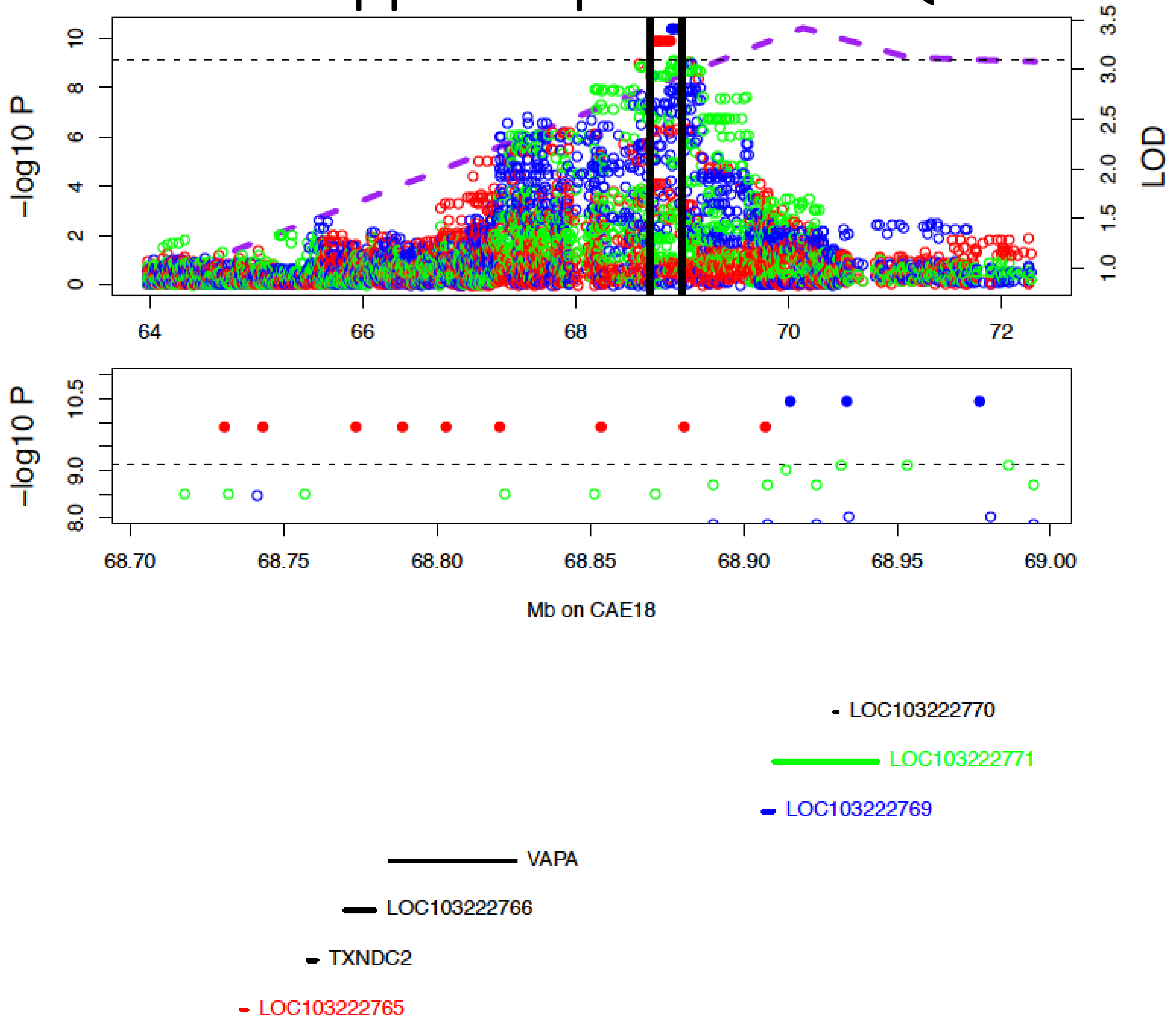
Hibar et al., 2015

Hippocampus volume
by MRI of > 300 monkeys in VRC pedigree

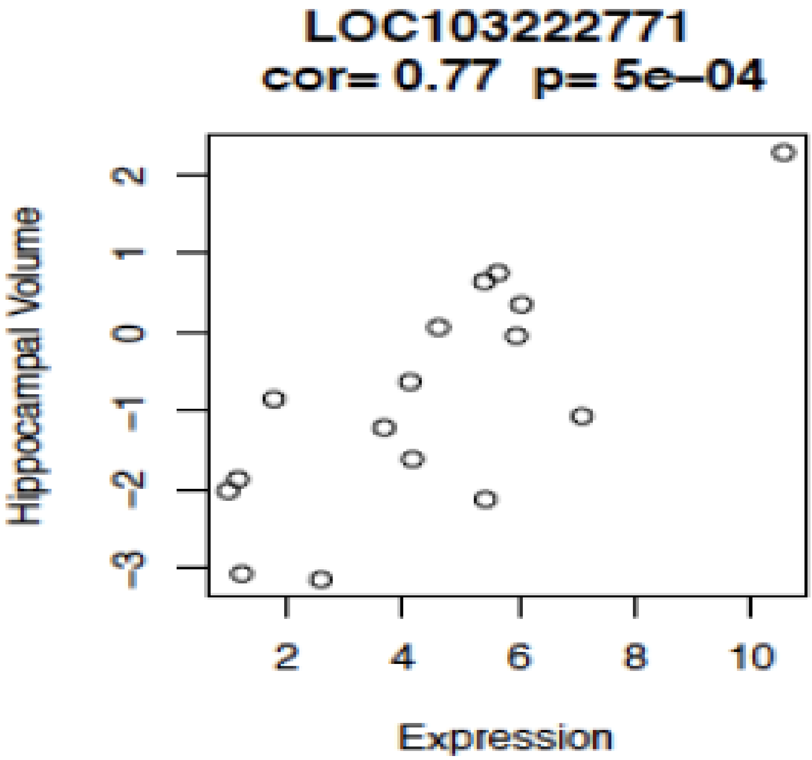


Heritability = 0.96   Peak lod score: 3.42 on Chr. 18

Hippocampal specific vervet eQTL for three long-noncoding RNAs lie within vervet hippocampal volume QTL



lncRNA LOC103222071 expression is highly correlated to hippocampus volume (in independent samples)

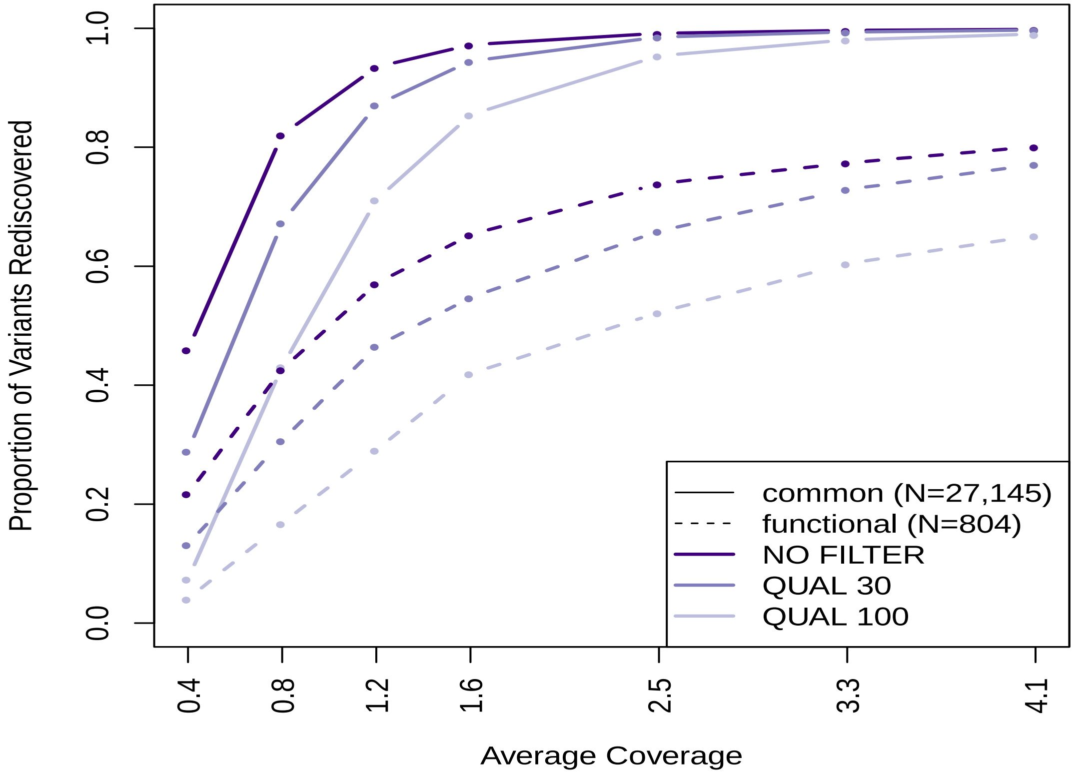


Caribbean vervet populations (N=50,000) will support large-scale association studies

Minor Allele Frequencies at Hippocampus eQTL

	VRC	SKN
LOC103222765	0.43	0.48
LOC103222769	0.46	0.34
LOC103222771	0.31	0.27

Recovery of VRC variants in SKN: Down-sampling of WGS data in 34 independent vervets



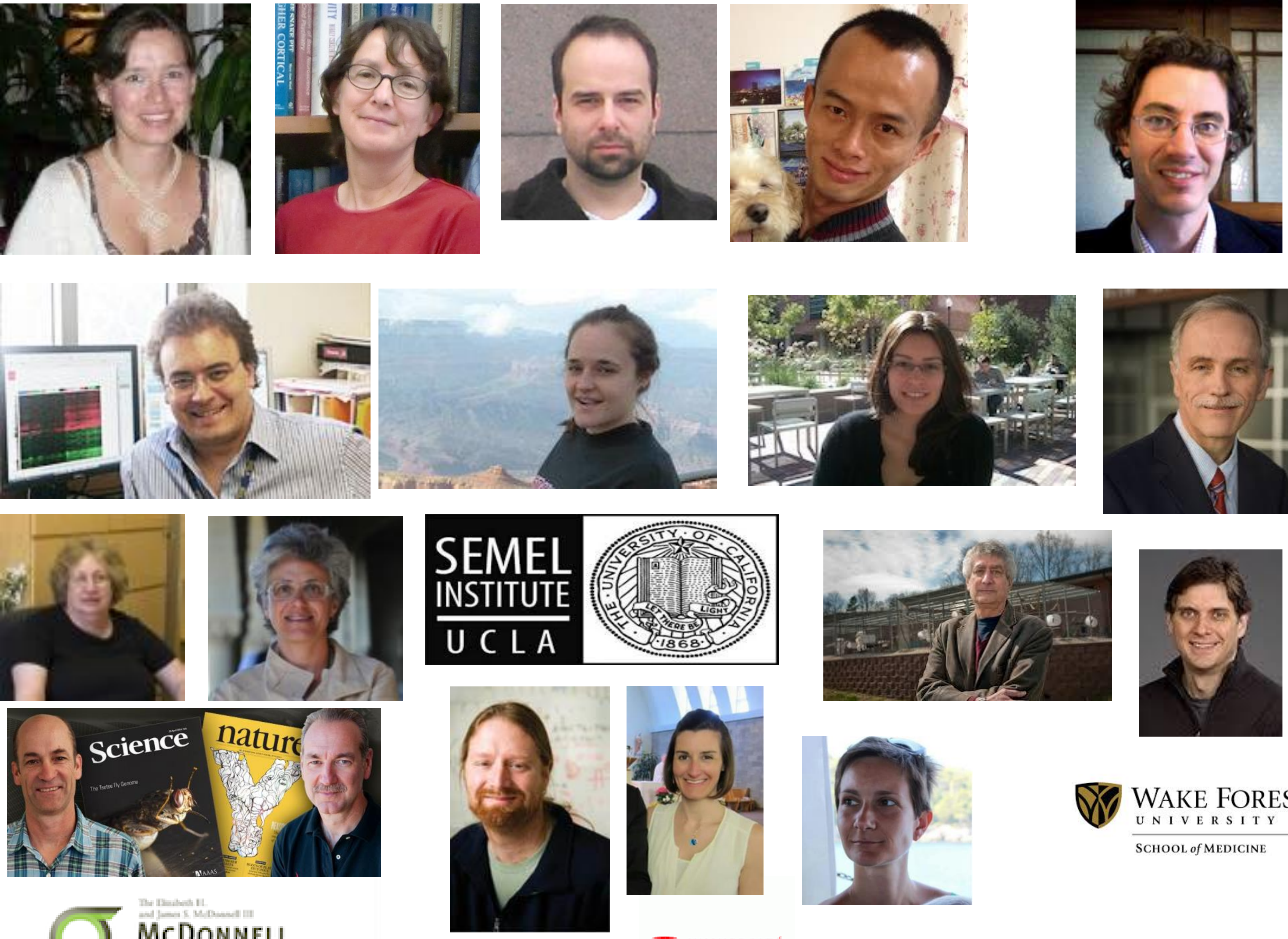
Planned studies of hippocampus and HPA axis in SKN population samples:

- Neurocognitive + behavioral assessment in relation to hippocampal eQTL genotypes
- Association studies in stress-response experiment
  - Transfer from wild to SKBRF research colony
  - Assessment of anxiety behavior and neurocognition
  - Glycemic, metabolic, and inflammatory markers
- Functional genomics at higher degree of specificity (including hypothalamus)
  - Specific nuclei to single cells

Summary

- Cheap sequencing has transformed utility of NHPs as genetic model systems:
  - Large phenotyped samples from appropriate populations
  - Comprehensive genetic variation (WGS)
  - Multiple functional genomic strategies

- Value will be particularly great for modeling brain/behavior disorders



SEMEL INSTITUTE
UCLA

WAKE FOREST UNIVERSITY
SCHOOL *of* MEDICINE

The Elizabeth H. and James S. McDonnell III
McDONNELL GENOME INSTITUTE
at Washington University

UNIVERSITÉ McGill