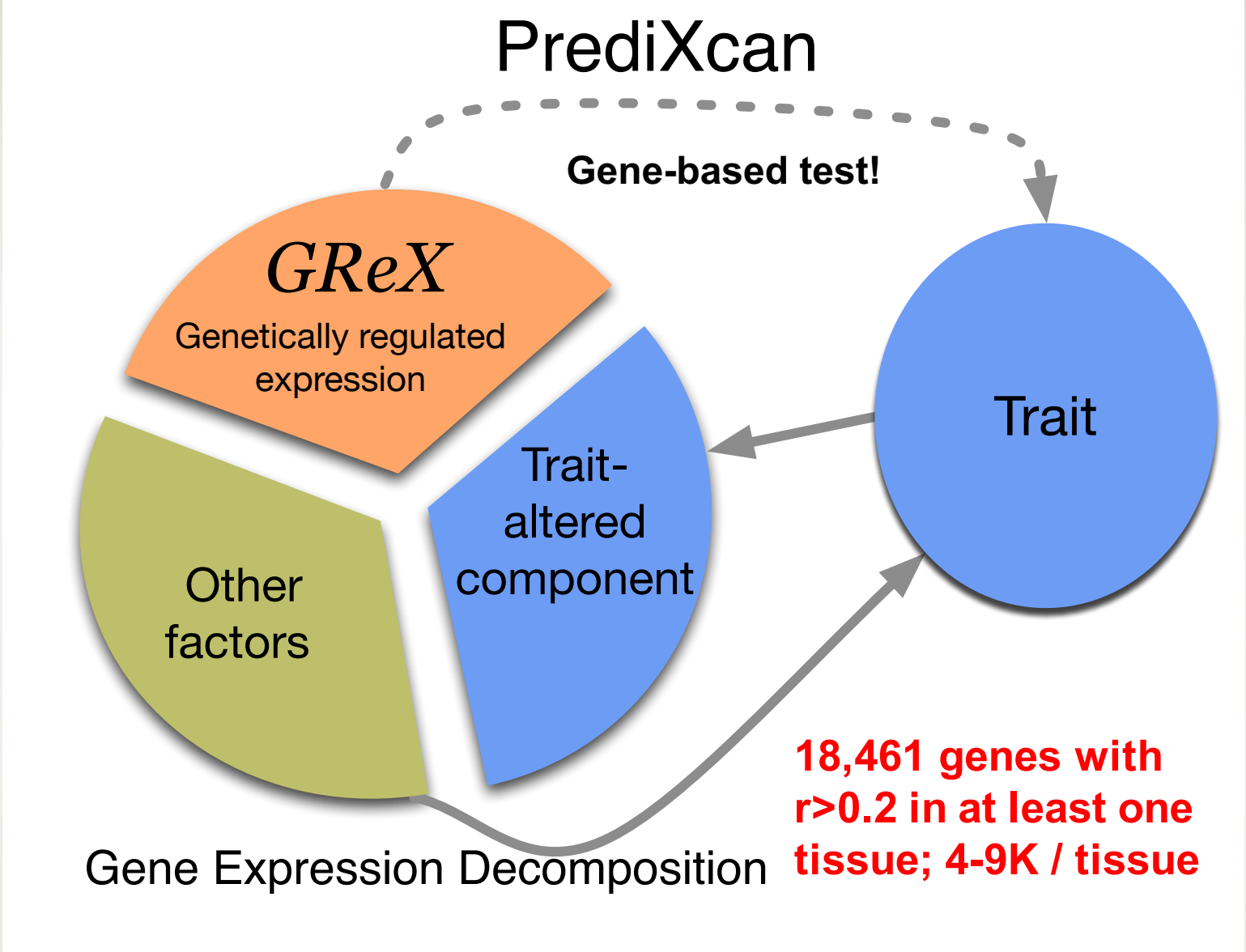
VANDERBILT UNIVERSITY
MEDICAL CENTER

Data Integration: Genome x Transcriptome x EMR

Building a Catalog of Gene to Medical Phenome

anderbilt Genetics Institute

Nancy J. Cox



Gamazon et al (2015) A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics 47(9):1091-8. PMC4552594

VANDERBILT UNIVERSITY MEDICAL CENTER

https://github.com/hakyimlab/PrediXcan

Resources for EMR-based research at Vanderbilt

The Synthetic Derivative

A de-identified and continuously-updated image of the EMR: 2,500,000 subjects

BioVU

Subjects with DNA: >215,000

- Dense (GWAS-level) genotypes: ~20,000
- Exome chip data: 36,000

Resources for EMR-based research at Vanderbilt 2017

The Synthetic Derivative

A de-identified and continuously-updated image of the EMR: 3,000,000 subjects

BioVU

Subjects with DNA: >225,000

- Dense genotypes: >120,000
- Whole genome or exome sequencing: ~1000's

BioVU X PrediXcan: Gene-based PheWAS

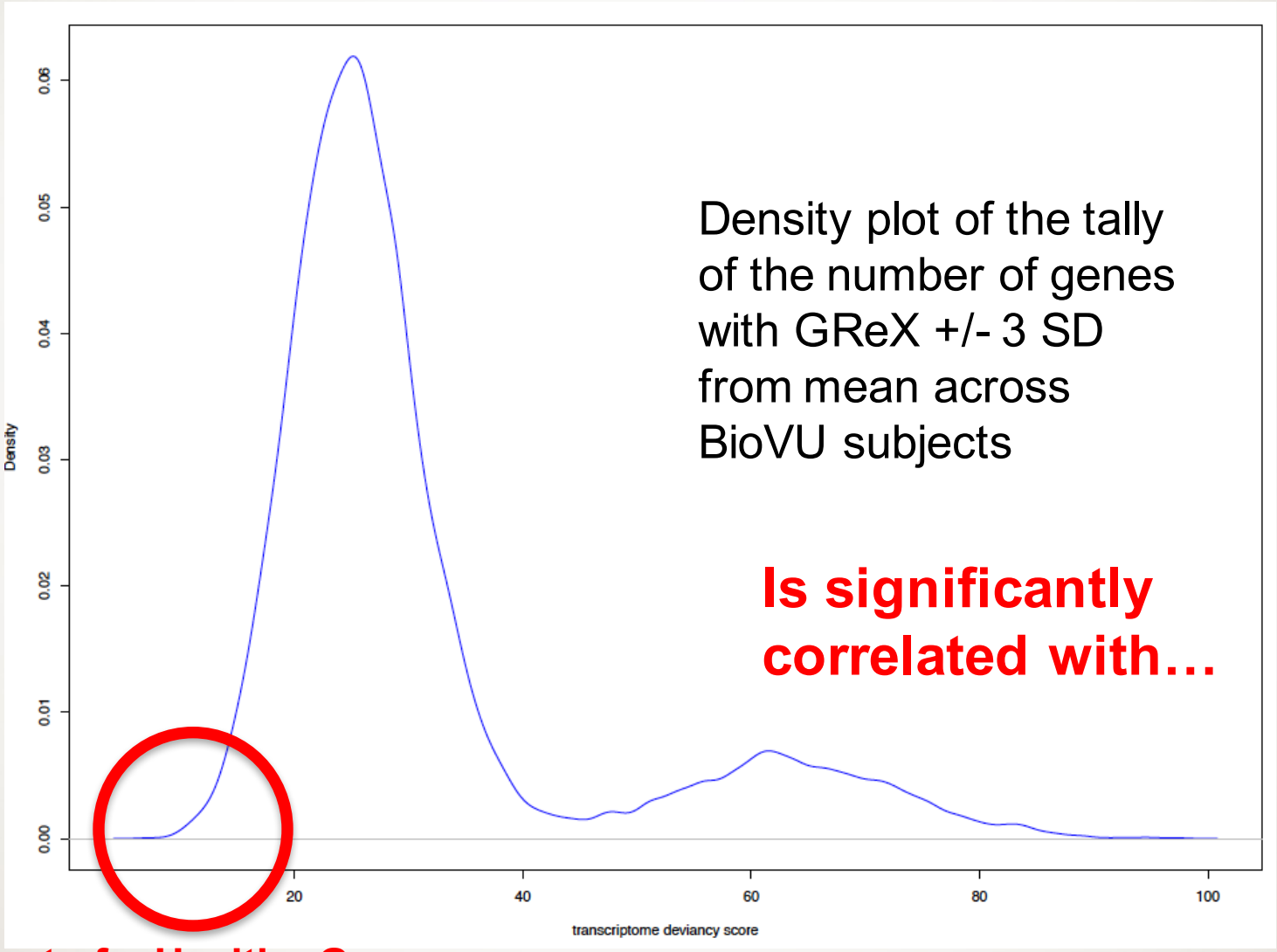
Comprehensive Gene X Medical Phenome Catalog

Knock-down each gene in each tissue and read out consequences across the medical phenome

Up-regulate each gene in each tissue and read out consequences across the medical phenome

An *in silico* Discovery Engine

Deviance of the Transcriptome



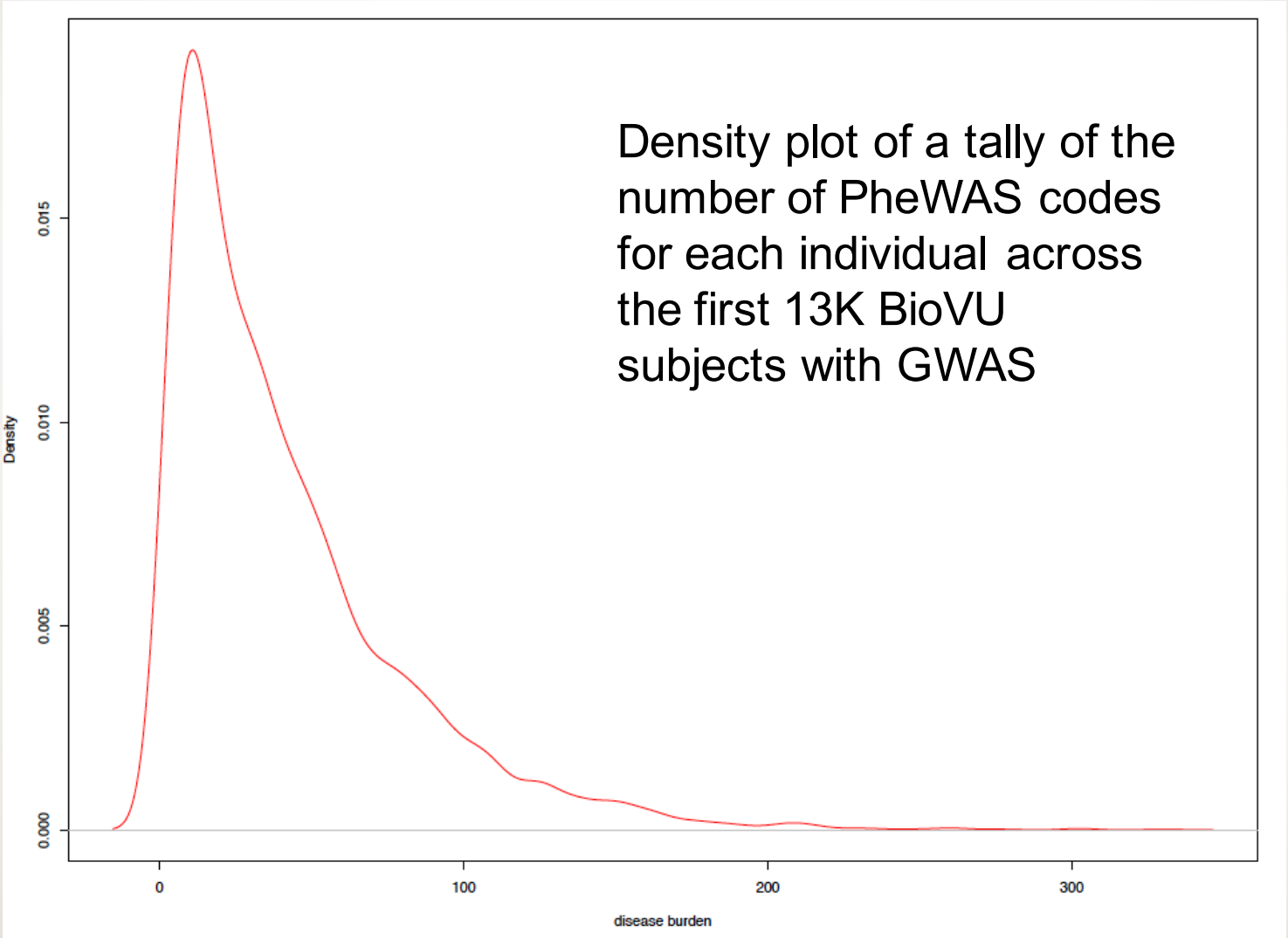
Density plot of the tally of the number of genes with GReX +/- 3 SD from mean across BioVU subjects

Is significantly correlated with…

Part of a Healthy Genome

VANDERBILT UNIVERSITY MEDICAL CENTER

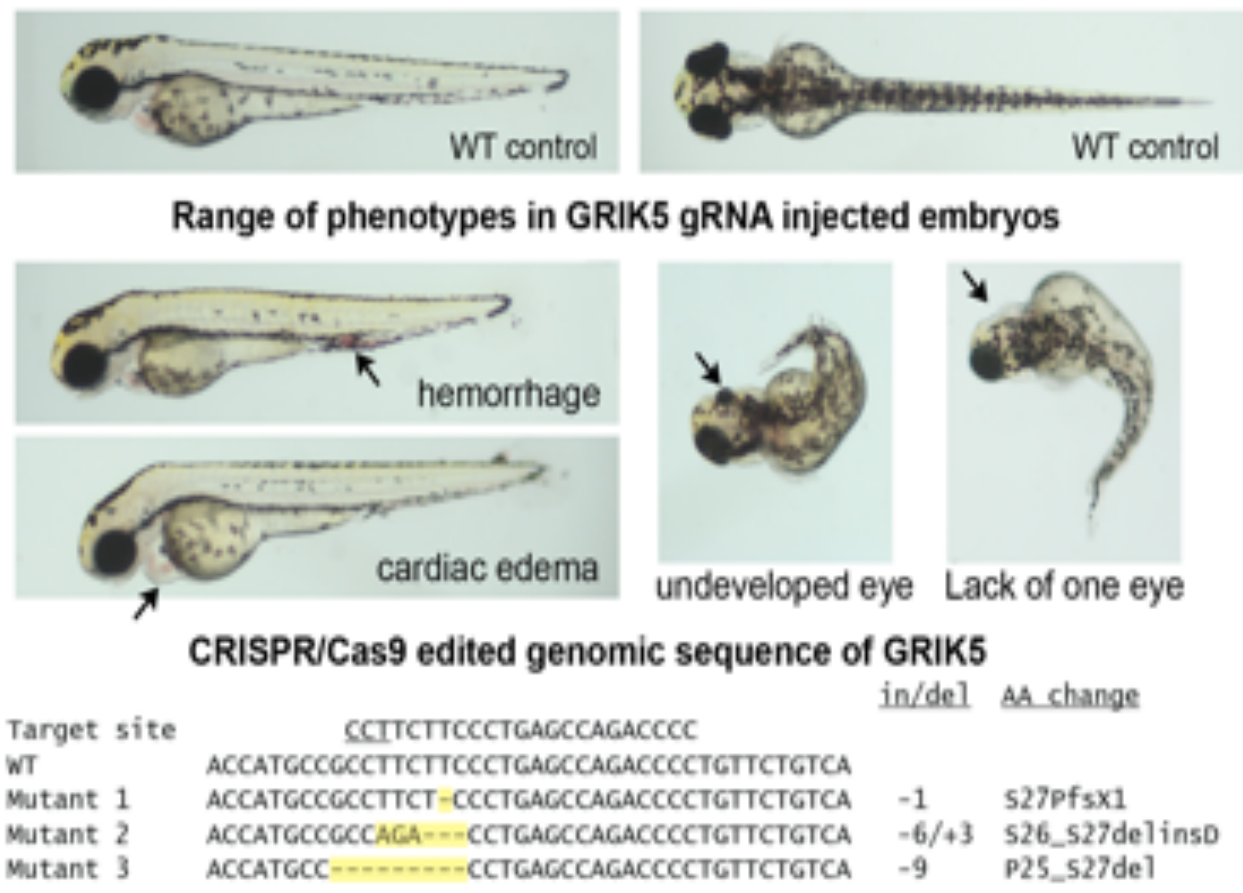
The Burden of Medical Disease



Reduced Predicted Expression *GRIK5*

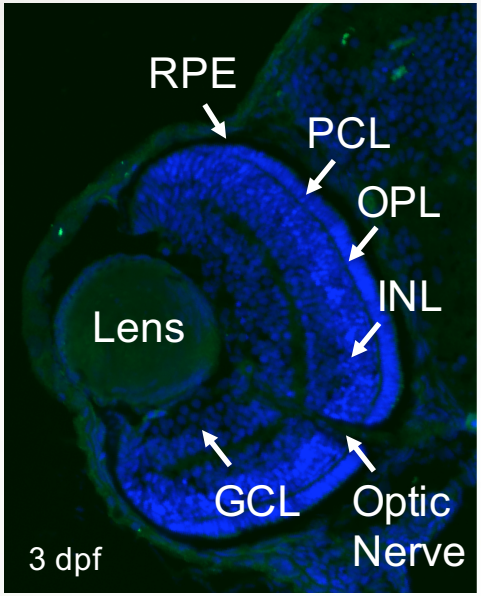
361	Retinal detachments and defects	54	0.000629
366	Cataract	629	0.000642
365	Glaucoma	219	0.00105
379	Other disorders of eye	233	0.00131
250.6	Polyneuropathy in diabetes	276	0.0014
365.11	Primary open angle glaucoma	72	0.00153
365.1	Open-angle glaucoma	150	0.00226
79	Viral infection	246	0.00379
627	Menopausal and postmenopausal disorders	365	0.00401
250.3	Insulin pump user	449	0.00422
530.1	Esophagitis, GERD and related diseases	1408	0.00455
366.2	Senile cataract	530	0.00507
627.2	Symptomatic menopause	235	0.0052
476	Allergic rhinitis	527	0.00525
379.2	Disorders of vitreous body	188	0.00627
530	Diseases of esophagus	1551	0.00636
763	Thoracic or lumbosacral neuritis or radiculitis, unspecified	134	0.00649
362	Other retinal disorders	321	0.00739
613	Other nonmalignant breast conditions	99	0.00752
577.3	Cyst and pseudocyst of pancreas	40	0.00756
530.11	GERD	1268	0.00812
514.2	Solitary pulmonary nodule	20	0.00831

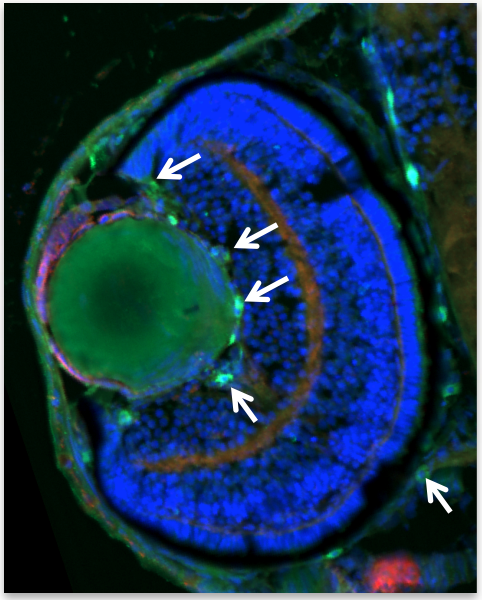
An Eye Super Gene?



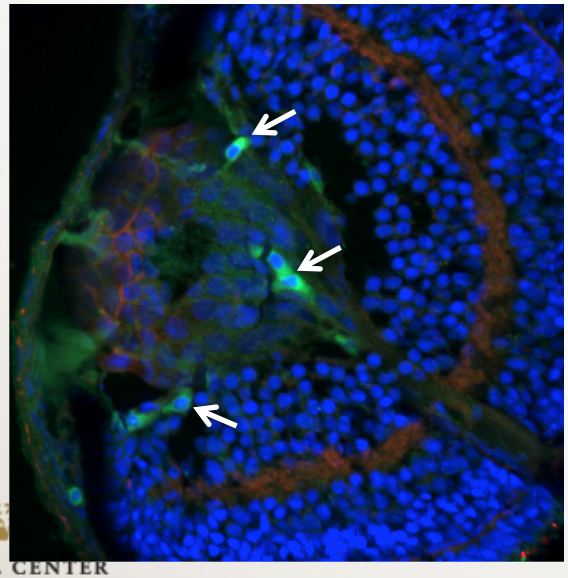
Zebrafish studies conducted in the Zebrafish Aquatic Facility by Ela Knapik, and students Daniel Levin, Gokhan Unlu, and Jessica Brown

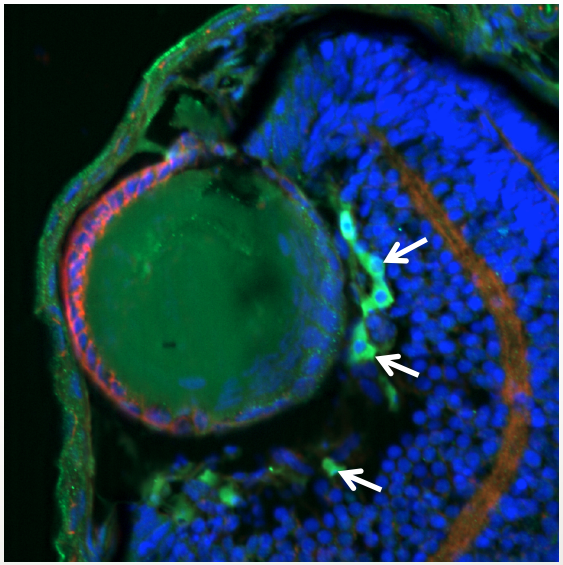
GRIK5 protein antibody staining in the zebrafish eye





RPE – Retinal Pigment Epithelium
PCL – Photoreceptor Cell Layer
OPL – Outer Plexiform Layer
INL – Inner Nuclear Layer
GCL – Ganglion Cell Layer





Nuclei (DAPI)
GRIK5

VANDERBILT MEDICAL CENTER

Continuum from Mendelian to Complex

Continuum from LOF to deleterious to ↓ expression

Nuclear Factor I, X-type – *NFIX*

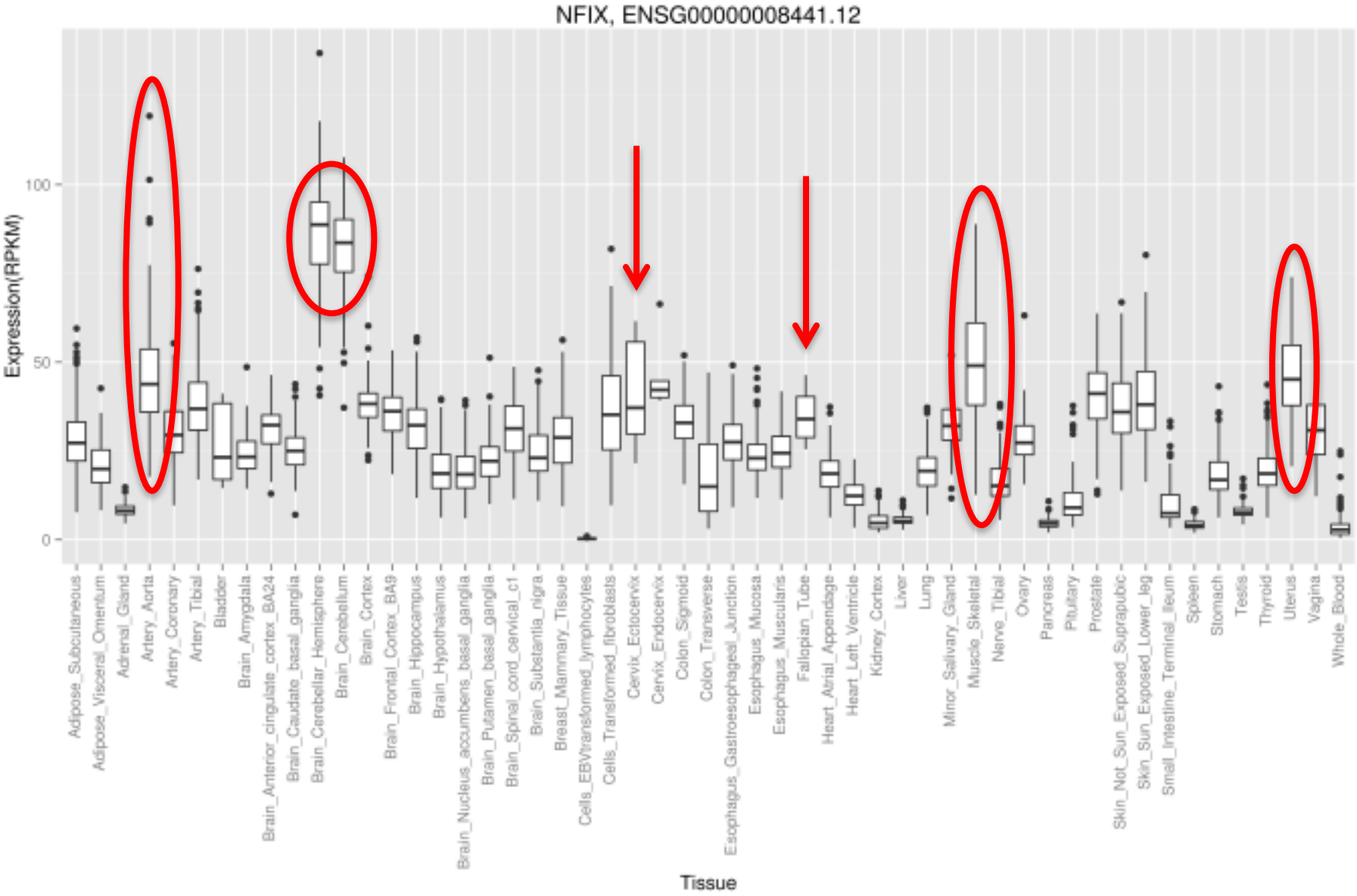
Mutations associatated with autosomal dominant diseases:

Marshall-Smith Syndrome

- Accelerated bone formation in hands and feet; fracture
- Diminished muscle tone
- Breathing difficulties; larynx and trachea "floppy"
- Facial features, blue sclera
- Mental and motor delays
  - Speech absent/abnormal
  - Intellectual disability / impairment

Sotos Syndrome 2

- Overgrowth in childhood; curvature/scoliosis, facial
- Muscle weakness
- Abnormalities of kidney, heart, eyes, ears, deafness
- Benign tumors, low-grade malignancies; seizures
- ID, behavior problems, speech/language disease
  - ADHD, OCD, etc
  - Stuttering, speech/language



VANDERBILT UNIVERSITY MEDICAL CENTER

NFIX Reduced Predicted Expression (blood)

Inflammatory diseases of uterus, except cervix			3.74E-19
Sialolithiasis			1.17E-12
Congenital anomalies of esophagus	In other tissues:		7.05E-12
Protozoan infection	Facial weakness	5.30E-10	1.42E-11
Pelvic inflammatory disease, NOS	Pneumonia due to fungus	5.46E-08	4.04E-11
Giant cell arteritis	Diseases of larynx and vocal cords	1.18E-06	1.11E-10
Acute inflammatory pelvic disease	Symptoms of respiratory system	6.11E-06	4.29E-10
Major puerperal infection	Symbolic dysfunction	8.00E-06	5.54E-10
Complications in administration of anesthetic / other sedation in labor delivery	Speech and language disorder	8.57E-06	5.81E-10
Cervical intraepithelial neoplasia [CIN] [Cervical dysplasia]			3.29E-09
Cardiac and circulatory congenital anomalies			7.04E-09
Aphakia and other disorders of lens	Disorders of tympanic membrane	2.21E-14	1.37E-08
Hypotony of eye	Neural tube defects	6.07E-06	3.06E-08
Pemphigus and pemphigoid	Kidney anomalies, disease	range	3.32E-08
Pelvic inflammatory disease (PID)	Fractures (ankle, foot, patella…)	range	3.43E-08
Hemarthrosis	Seizures, convulsions, epilepsy	range	8.65E-08
Viral infection			7.05E-07
Congenital anomalies of posterior segment of eye			1.09E-06
Cardiac congenital anomalies			1.40E-06

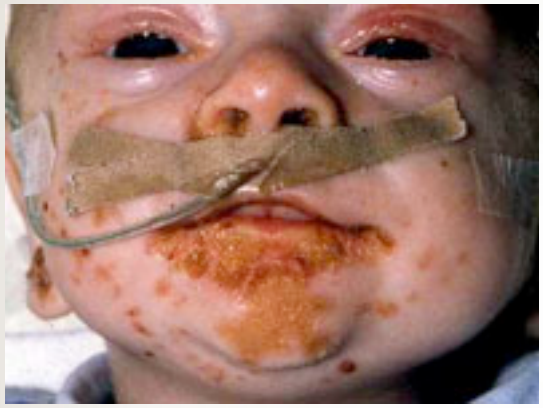
Dominant mutations: Sotos syndrome 2; Marshall-Smith Syndrome

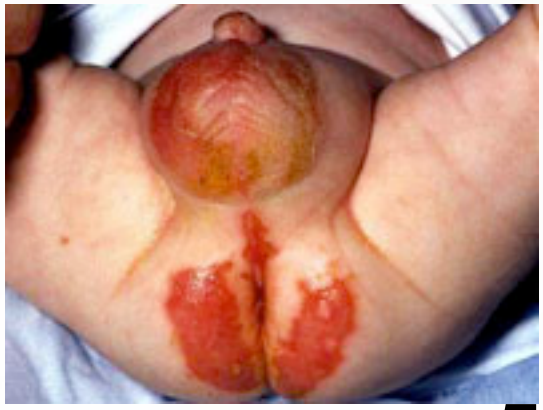
VANDERBILT UNIVERSITY MEDICAL CENTER

What we are doing…

- **Database of Mendelian disease genes and associated phenotypes** Improve diagnosis, Cycle back to phenotyping
  - Rare disease characterized in just a few patients; "data-driven models for range of clinical features"
  - Need for "OUTCOMES" as patients live longer
- **Creating a database of "Mendelian genes in waiting"**
  - Genes not (yet) characterized as Mendelian, but have multiple congenital anomalies and ID (and other really bad phenotypes)
  - One of the few ways to predict "de novo" phenotypes

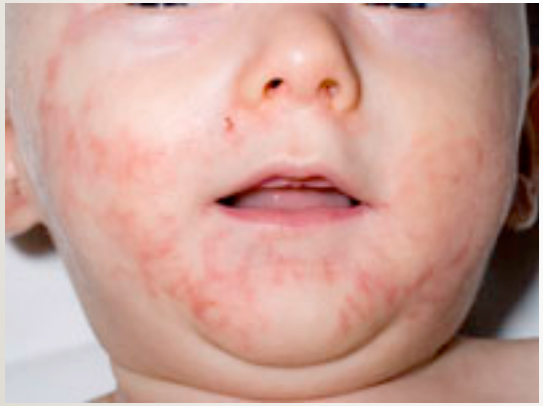
SLC39A4: Autosomal recessive acrodermatitis enteropathica

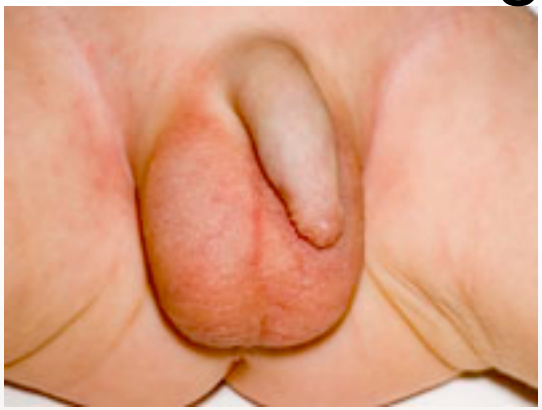


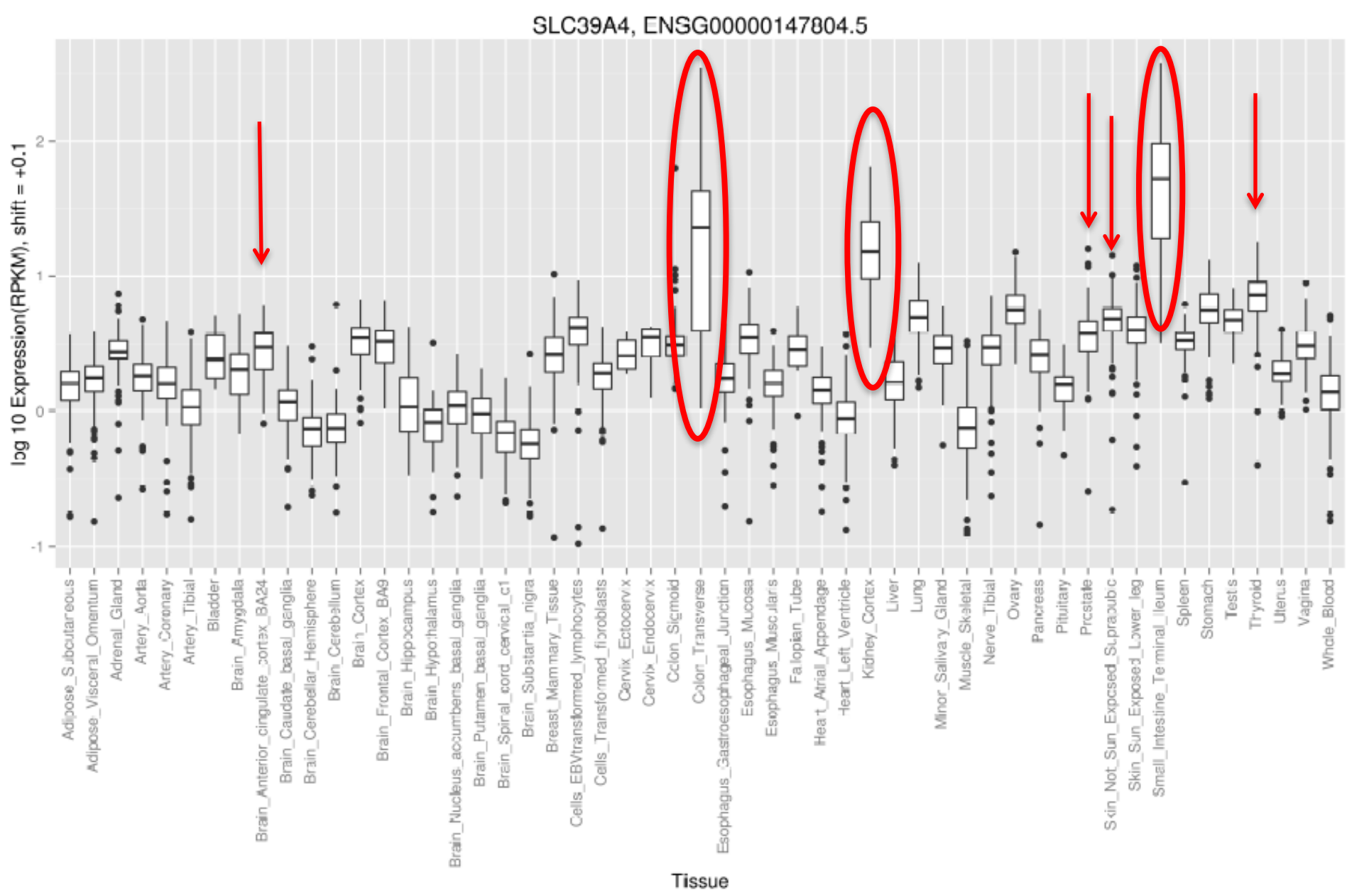


Other phenotypes: chronic diarrhea, gastritis, serious behavioral problems, anemia; fatal in early childhood

5 days after zinc supplementation





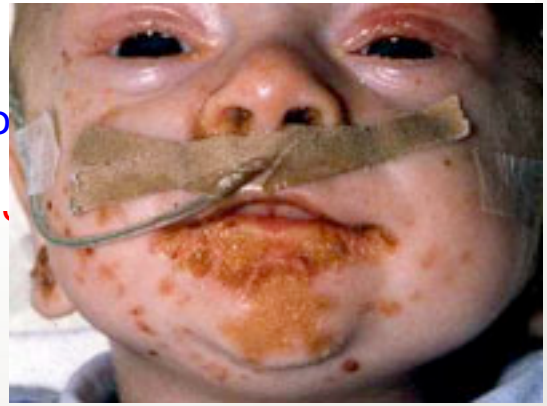

SLC39A4, ENSG00000147804.5
log 10 Expression(RPKM), shift = +0.1
2
1
0
-1
Adipose_Subcutaneous
Adipose_Visceral_Omentum
Adrenal_Gland
Artery_Aorta
Artery_Coronary
Artery_Tibial
Bladder
Brain_Amygdala
Brain_Anterior_cingulate_cortex_BA24
Brain_Caudate_basal_ganglia
Brain_Cerebellar_Hemisphere
Brain_Cerebellum
Brain_Cortex
Brain_Frontal_Cortex_BA9
Brain_Hippocampus
Brain_Hypothalamus
Brain_Nucleus_accumbens_basal_ganglia
Brain_Putamen_basal_ganglia
Brain_Spinal_cord_cervical_c1
Brain_Substantia_nigra
Breast_Mammary_Tissue
Cells_EBVtransformed_lymphocytes
Cells_Transformed_fibroblasts
Cervix_Ectocervix
Cervix_Endocervix
Colon_Sigmoid
Colon_Transverse
Esophagus_Gastroesophageal_Junction
Esophagus_Mucosa
Esophagus_Muscularis
Fallopian_Tube
Heart_Atrial_Appendage
Heart_Left_Ventricle
Kidney_Cortex
Liver
Lung
Minor_Salivary_Gland
Muscle_Skeletal
Nerve_Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin_Not_Sun_Exposed_Suprapubic
Skin_Sun_Exposed_Lower_leg
Small_Intestine_Terminal_Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole_Blood
Tissue

SLC39A4 Reduced Predicted Expression (blood)

Other hypertrophic cardiomyopathy	1.61E-16
Hereditary hemolytic anemias	1.24E-13
Benign neoplasm of other fema	1.45E-10
Schizophrenia	3.43E-09
Acquired deformities of hip	1.82E-07
Toxic effect of corrosive aroma	2.45E-07
Mineral deficiency NEC	2.23E-06
Kaschin-Beck disease	2.92E-06
Multiple gestation	3.09E-06
Iodine hypothyroidism	3.13E-06
Abnormal spermatozoa	3.56E-06
Cervical incompetence	4.76E-06
Intestinal disaccharidase defici[illegible]	6.78E-06
Gastritis and duodenitis, NOS	1.00E-05

In other tissues:

Impetigo	10.0E-22
Pilonidal cyst	1.31E-11
T2D	8.52E-11
Pruritis and related conditions	4.27E-10
Disorders of mineral metabolism	2.29E-08
Acute renal failure, …	3.70E-08
Primary pulmonary hypertension	8.69E-08
Suicidal ideation or attempt	1.64E-07
Other cerebral degeneration	9.26E-07
Bullous dermatitis	6.24E-06
Psoriasis	
Diarrhea, ...	
Gout and crystal arthro	



Recessive mutations: Acrodermatitis enteropathica

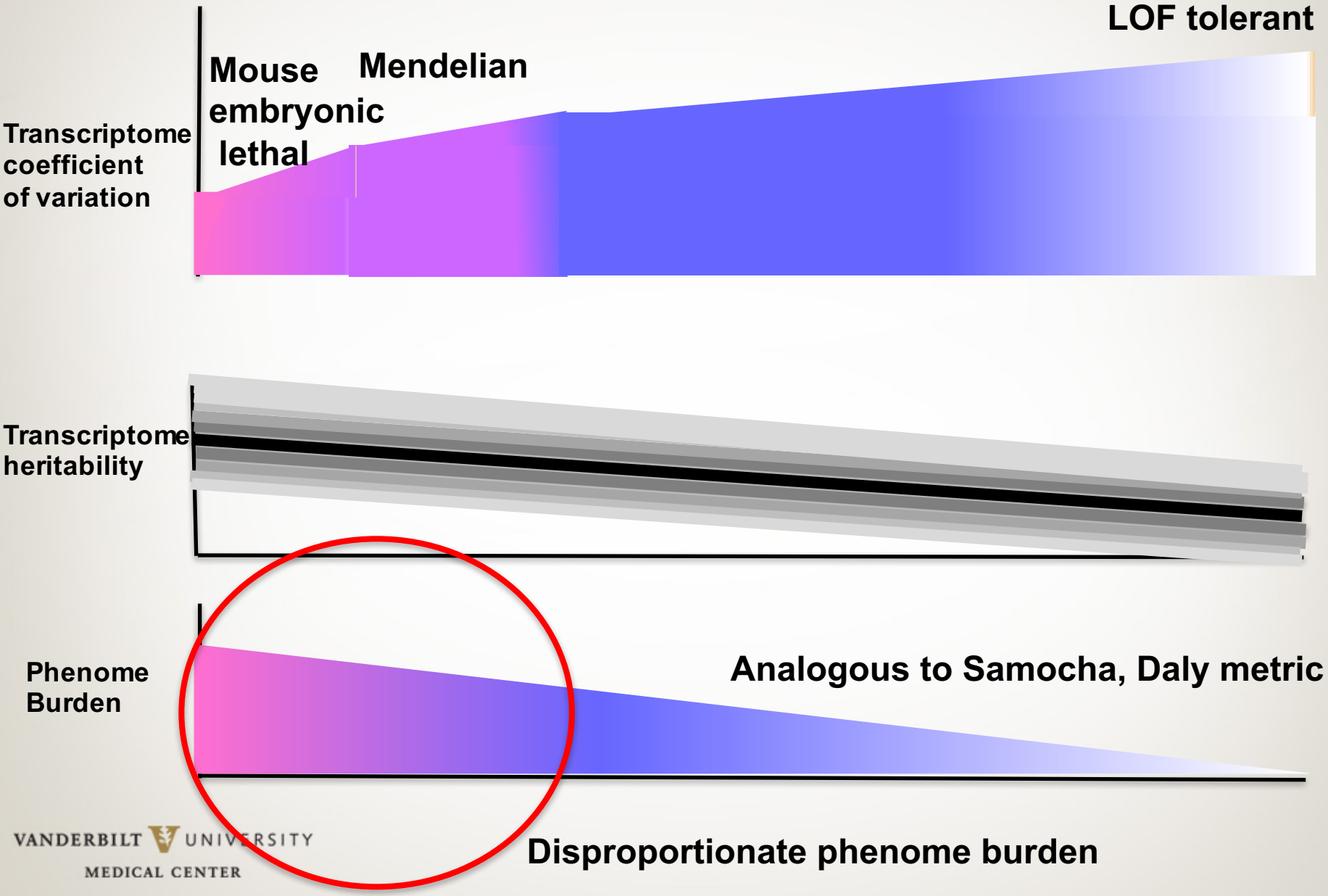
VANDERBILT UNIVERSITY MEDICAL CENTER

Continuum between Mendelian and Common Disease

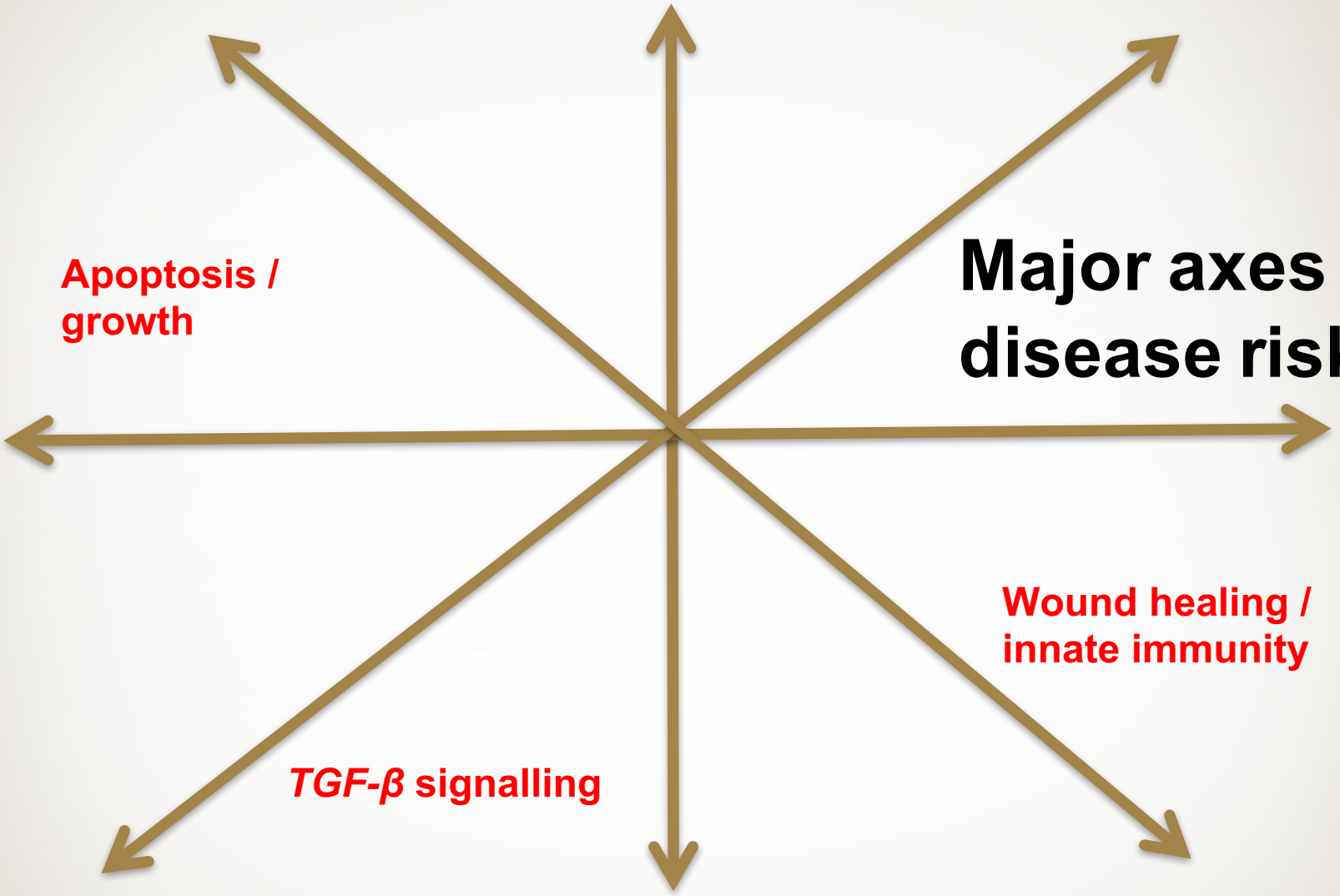
- There are dozens of Mendelian diseases that can be treated reasonably effectively with innocuous therapies – vitamin or mineral supplementation or dietary intervention
- There WILL be more people with increased risk of disease due to reduced expression of JUST these genes than there are people who have a Mendelian disease

Acrodermatitis enteropathica: 1 / 500,000 live births; none in BioVU now
>5000 patients in BioVU today are at high risk for worst sub-phenotypes
>300 patients in BioVU will have multiple of the worst sub-phenotypes

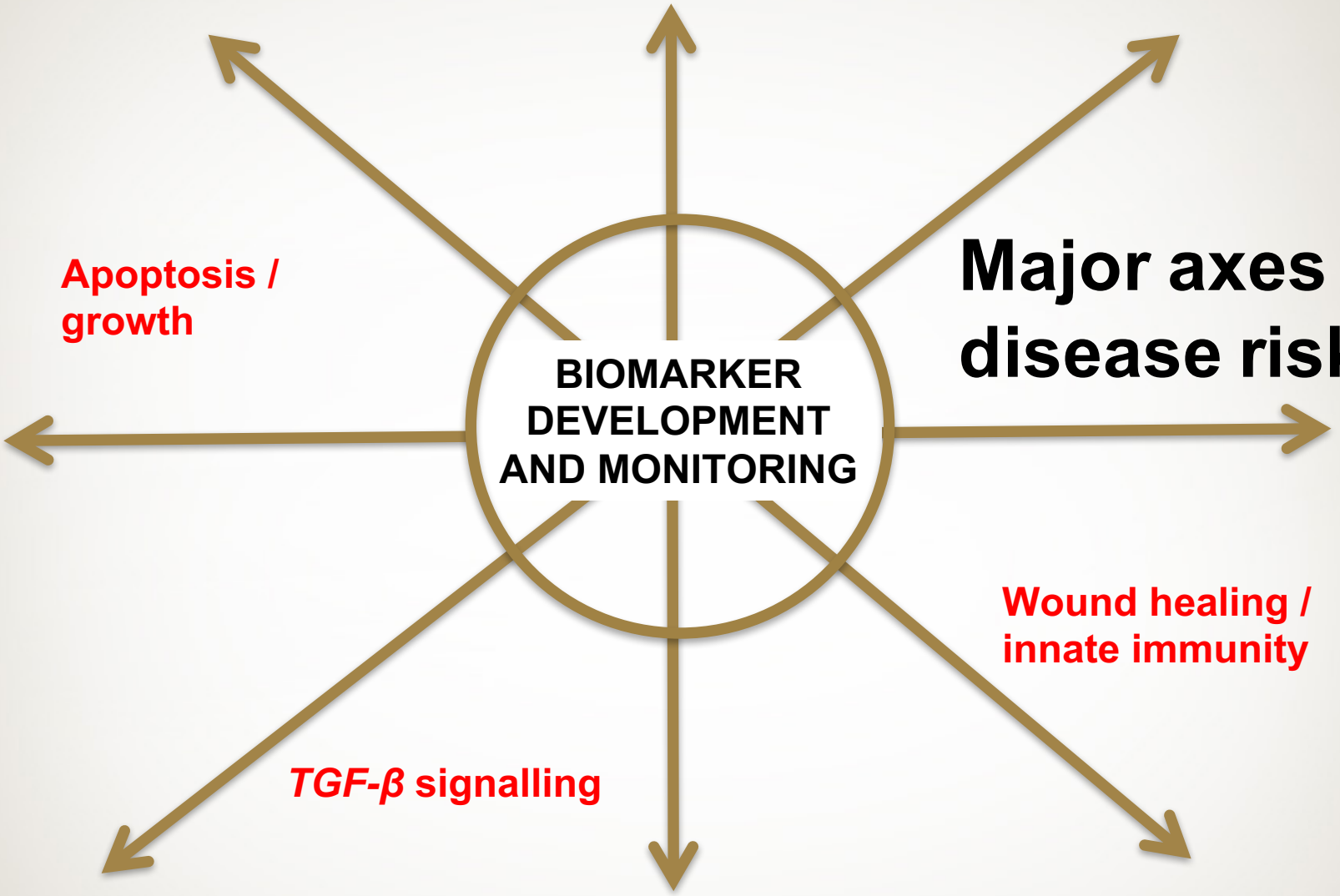
Big Picture …



VANDERBILT UNIVERSITY MEDICAL CENTER


Apoptosis / growth
Major axes of disease risk
Wound healing / innate immunity
TGF-β signalling

Was Aristotle Right?



VANDERBILT UNIVERSITY MEDICAL CENTER

Results on all genes in 18,000

Results on all genes in 36,000

Results in 72,000,

120,000+, …

Eric Gamazon

Lisa Basterache

Anuar Konkashbaev

VANDERBILT UNIVERSITY
MEDICAL CENTER

Vanderbilt Zebrafish Aquatic Facility

Ela Knapik

Gokhan Unlu

Jess Brown

Daniel Levic

VICTR – Vanderbilt Institute for Clinical and Translational Research

Gordon Bernard

Our GTEx Team at University of Chicago

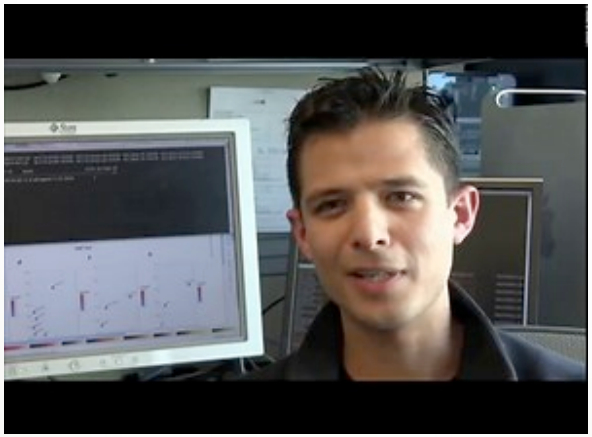
Dan Nicolae

Lin Chen

Hae Kyung (Haky) Im

Barbara Stranger

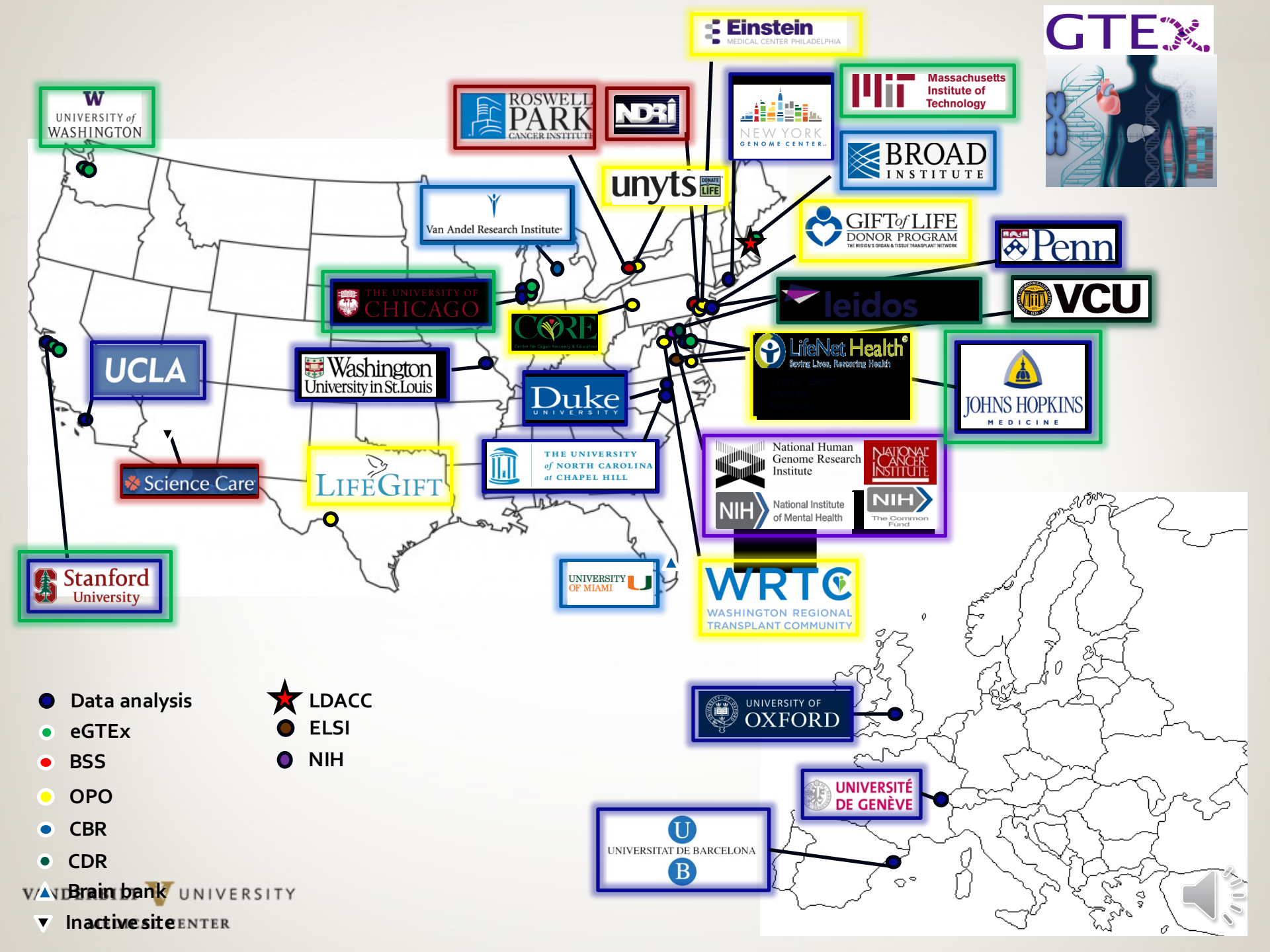
Kaanan Shah



Jason Torres

Keston Aquino-Michaels

VANDERBILT UNIVERSITY
MEDICAL CENTER


GTEx
Einstein
MEDICAL CENTER PHILADELPHIA
UNIVERSITY of WASHINGTON
ROSWELL PARK CANCER INSTITUTE
NDRI
NEW YORK GENOME CENTER
Massachusetts Institute of Technology
BROAD INSTITUTE
unyts
Van Andel Research Institute
GIFT of LIFE DONOR PROGRAM
Penn
THE UNIVERSITY OF CHICAGO
leidos
VCU
CORE
UCLA
Washington University in St.Louis
LifeNet Health
Saving Lives, Restoring Health
JOHNS HOPKINS MEDICINE
Duke UNIVERSITY
Science Care
LIFEGIFT
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
National Human Genome Research Institute
NATIONAL CANCER INSTITUTE
NIH
National Institute of Mental Health
NIH The Common Fund
Stanford University
UNIVERSITY OF MIAMI
WRTC
WASHINGTON REGIONAL TRANSPLANT COMMUNITY
UNIVERSITY OF OXFORD
UNIVERSITÉ DE GENÈVE
UNIVERSITAT DE BARCELONA
Data analysis
eGTEx
BSS
OPO
CBR
CDR
Brain bank
Inactive site
LDACC
ELSI
NIH
VANDERBILT UNIVERSITY
MEDICAL CENTER