

Big data approaches in cardiovascular genomic research

Randal Westrick

CDaS and Department of Biological Sciences

Dec. 1, 2016

Thrombosis is the #1 Cause of Morbidity and Mortality

- **Thrombosis**

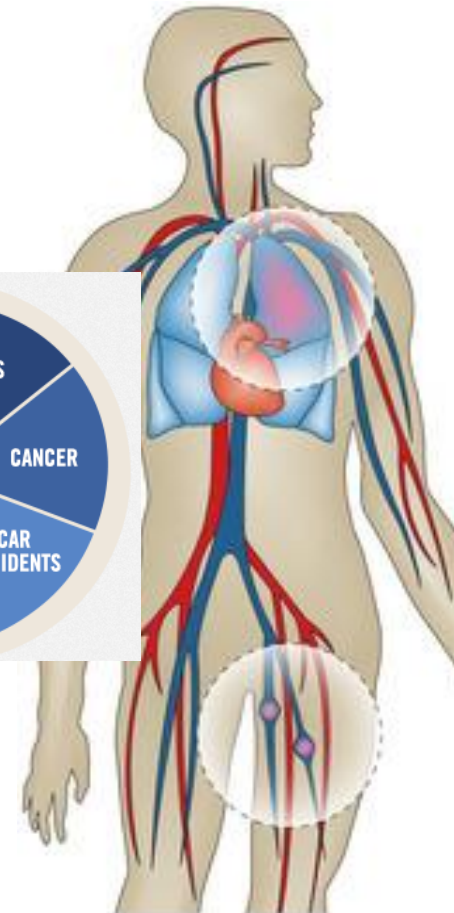
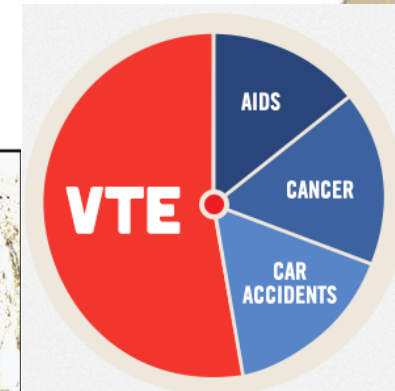
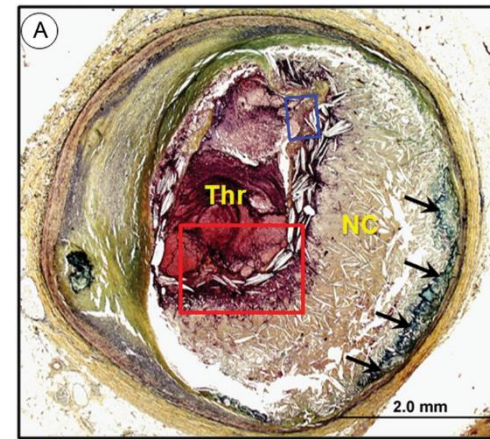
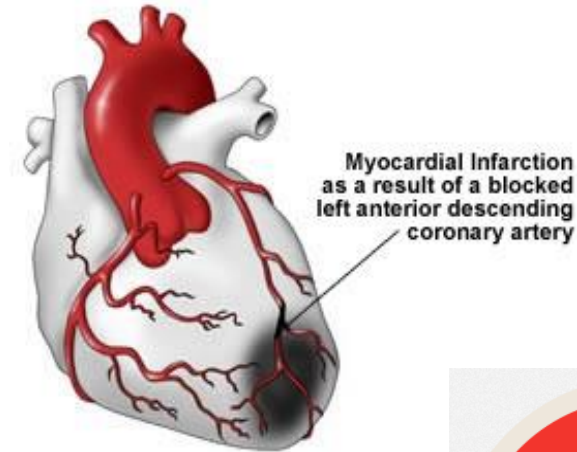
- Blood clotting **INSIDE** the blood vessels

- **Arterial Thrombosis**

- Atherothrombosis in arteries following plaque rupture causes **heart attacks and strokes**
 - No way to predict which plaques will rupture

- **Venous thrombosis**

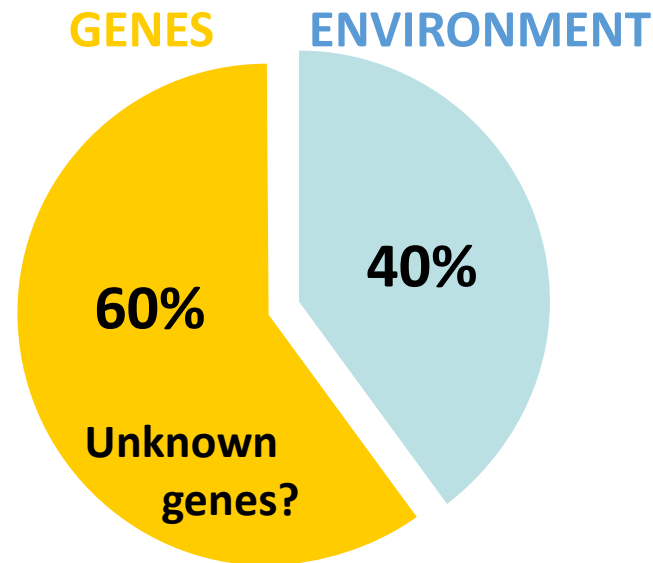
- Occurs in veins at sites of stasis, commonly veins of lower extremity
- **Venous thromboembolism (VTE)**
 - **Life threatening**
 - No way to predict when VTE will occur



What Are the Genetic and Environmental Risk Factors for Blood Clotting (Thrombosis)?

Venous thrombosis

- 60% heritable
- Factor V Leiden (F5^L)
 - Accounts for 25% of genetic risk
 - **Not everybody inheriting F5^L will get venous thrombosis**
 - Tissue Factor Pathway Inhibitor?



Arterial thrombosis

- 60% heritable
- Risk factors?
 - Plasminogen activator inhibitor 1 (PAI-1)
 - Causal?

Improved human and animal studies of arterial and venous thrombosis are essential for identifying thrombosis genes and environmental triggers

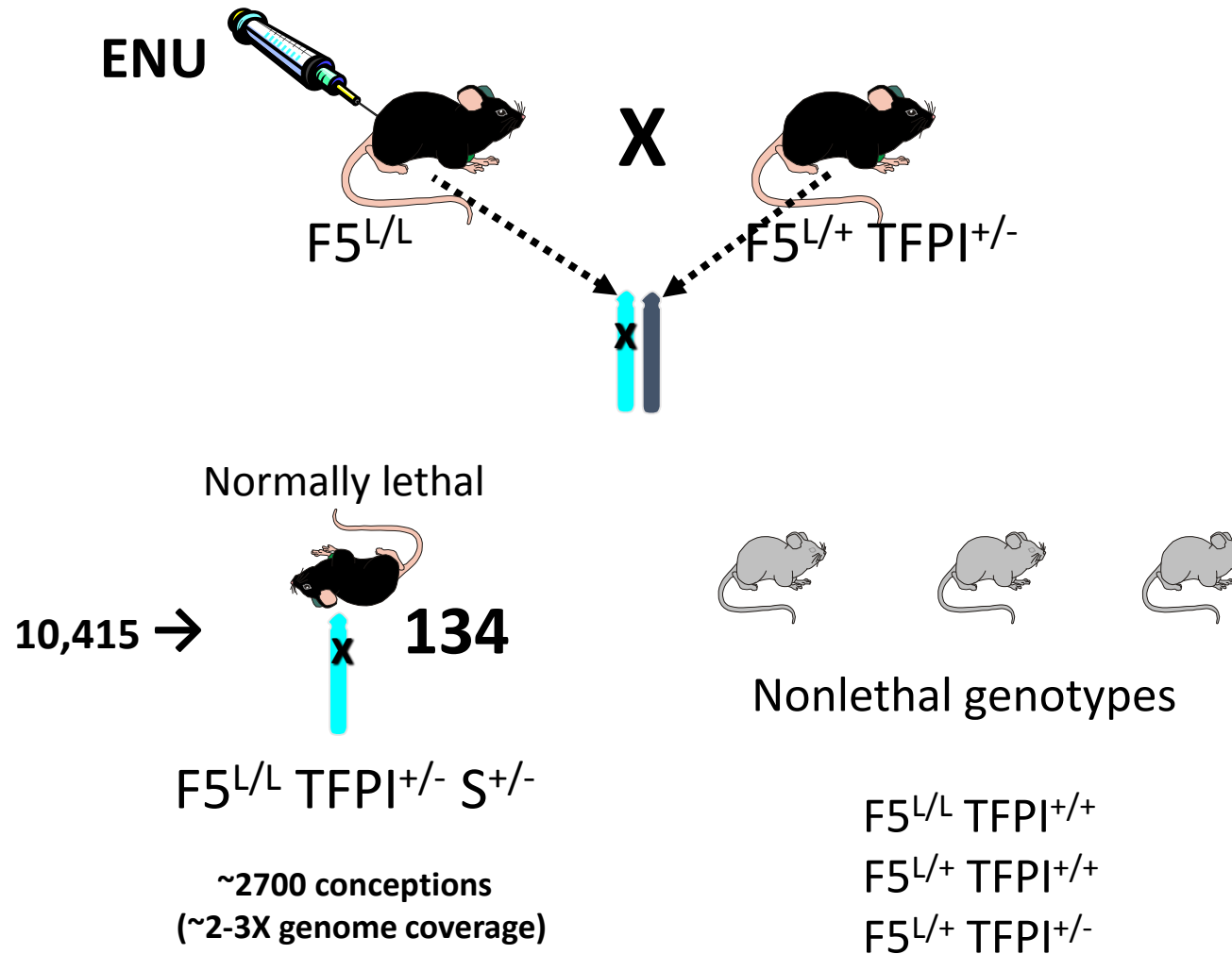
The Present Role of Big Data in Cardiovascular Thrombotic Disease

- Researchers are harnessing Big Data Genomic Analyses to:
 - Identify disease susceptibility even before illness appears
 - Facilitate preventive treatment
 - Improve diagnosis
 - Tailor treatment to diagnosis
 - Facilitate treatment prescription with minimum toxicity and maximum efficacy (pharmacogenomics)
- Efforts are centered on genotyping and sequencing hundreds of thousands of patient and control genomes
- Analyses have not yet identified big effect genes (disregarding environment???)



Genome Wide ENU Mutagenesis Suppressor Screen for Thrombosis Genes

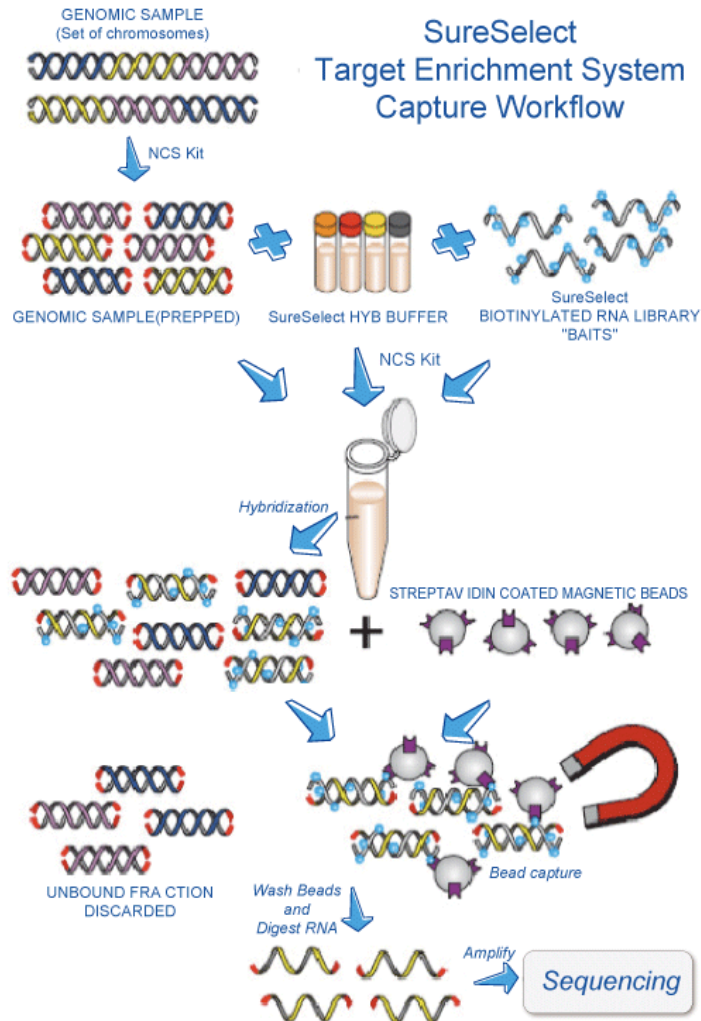
ENU induces ~30
random mutations
throughout the
genome per offspring





Identification of Thrombosis Genes by Whole Exome Sequencing

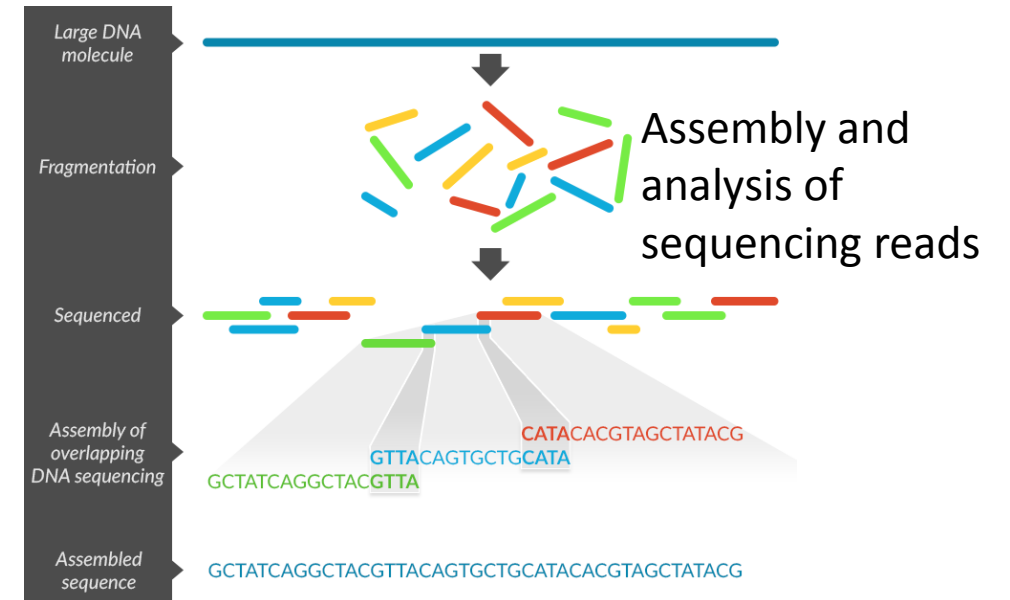
Agilent SureSelect System



Pull out all coding sequence (exome) from genomic DNA samples



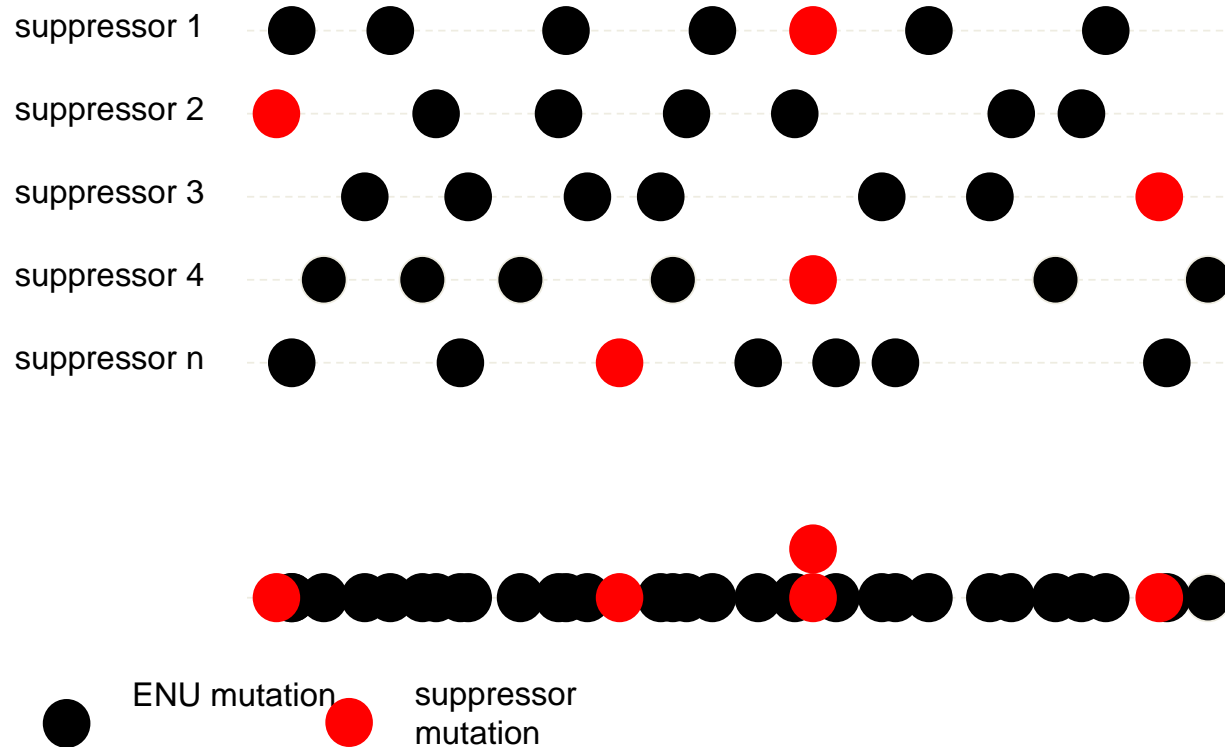
100x genome coverage per 50 megabase exome



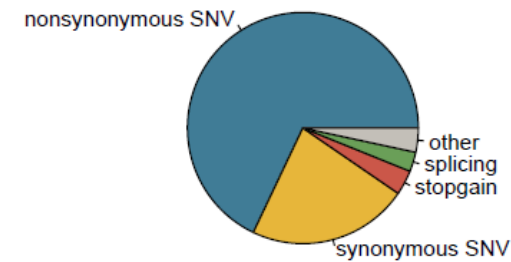
Mutation Burden Testing and Distributions of ENU Induced Variants in 114 F5^{L/L} TFPI^{+/-} Mice

NHLBI resequencing program

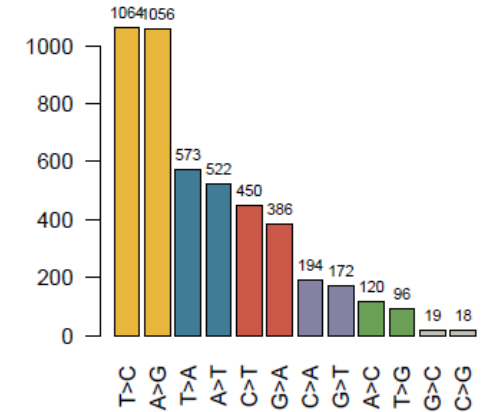
Mutation Burden



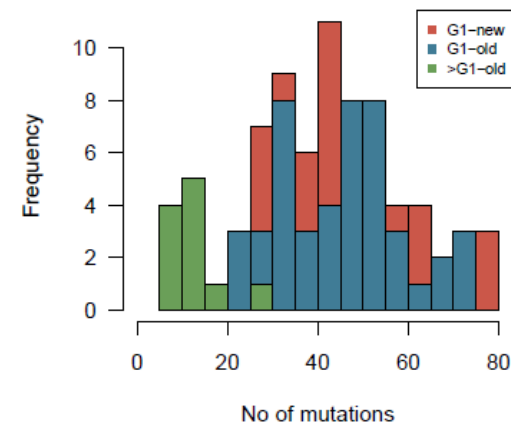
ENU induced variants (n=4688)



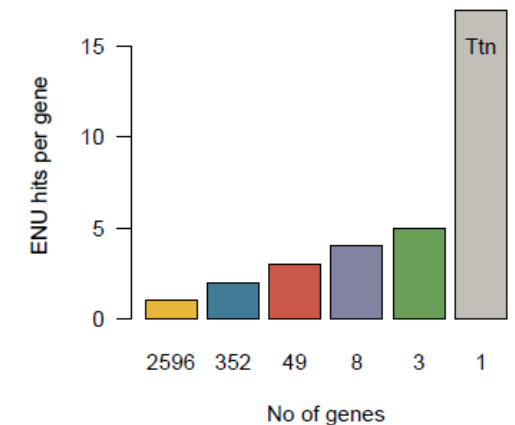
SNV types



Mutation burden

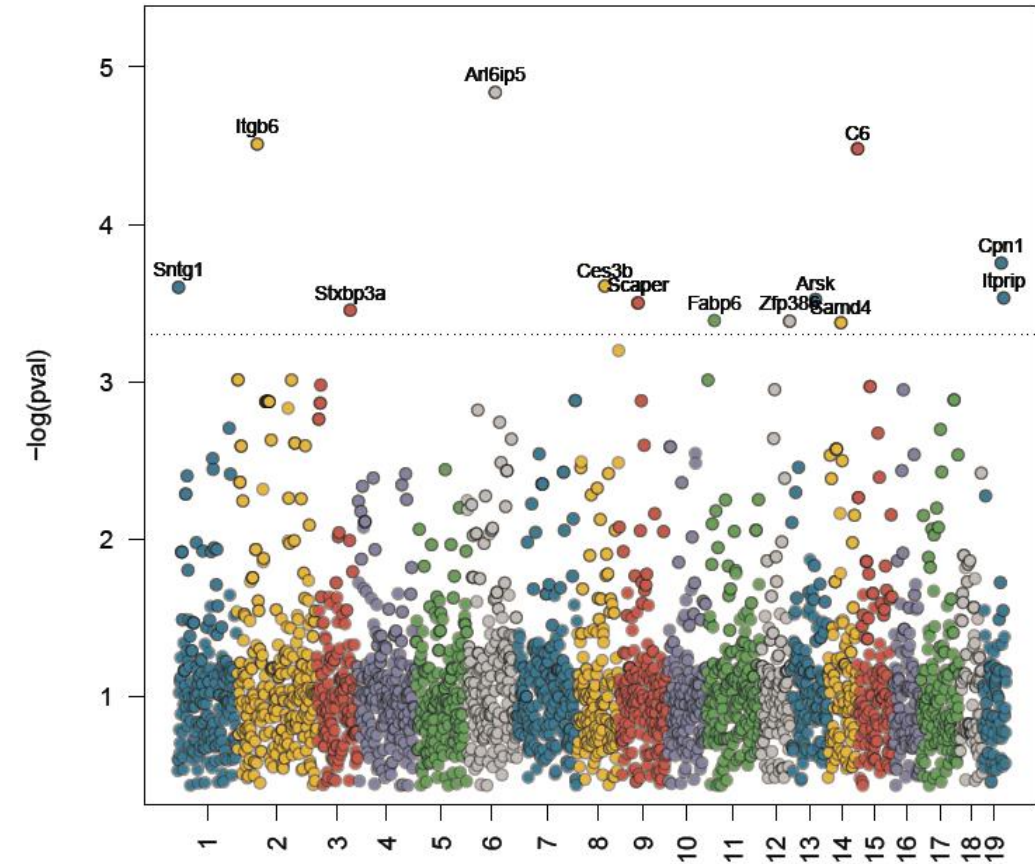
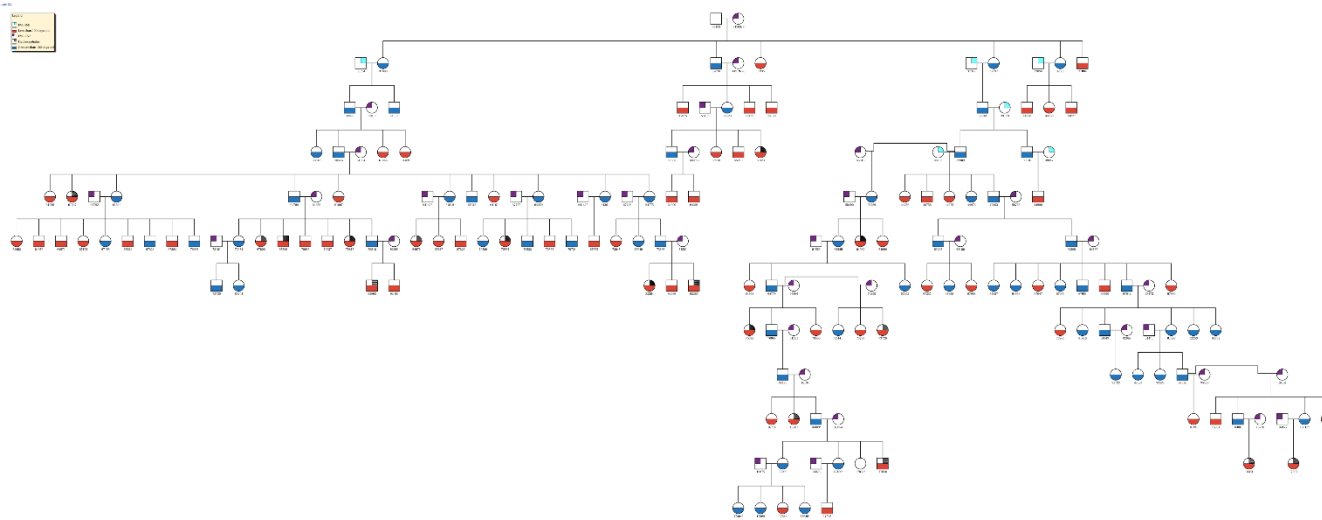


ENU hits distribution



A Subset of Novel Thrombosis Genes Identified by Whole Exome Sequencing

- Several thrombosis genes have been found through this method
- The majority have not been found
- Whole genome sequencing for the remainder



ENU SNVs by CHR
p-value calculated by 10,000,000 permutations (normalized to gene size)

Cardiovascular Thrombotic Disease Research at CDaS

- Leverage Center expertise to analyze and identify environmental thrombosis triggers from lifestyle data (sleep, activity, stress...) using Biosensors
 - In our genetically susceptible rodents (proof of principal)
 - Humans (in partnership with Translational medicine researchers and companies)
- Integrate Genome and Envirome information to inform us about how particular genomes interact with their environments
- Graduate and Undergraduate Students will acquire research skills including Big Data Genome Analysis and Big Data Sensor/Scanning analyses
- Develop novel preventive or pre-disease therapeutic strategies for thrombosis/cardiovascular disease

