

SNPs and Networks

Rotation talk by: Yao Fu
March 16, 2011

Outline

- SNPs & Network
- Positive Selection

SNPs & Network

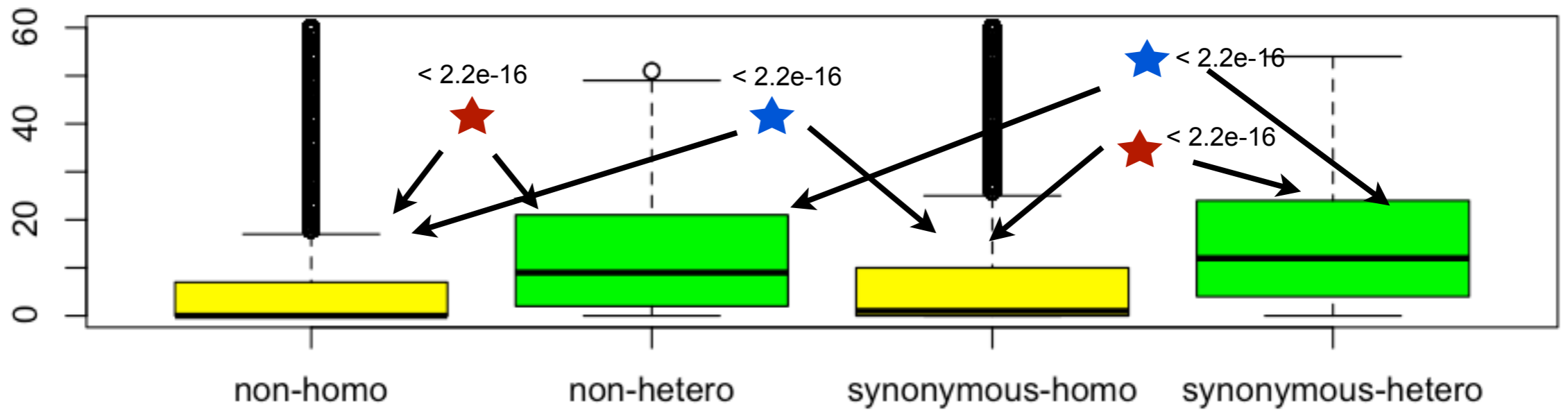
■ Data

- 1000 genome SNPs data, mainly non-synonymous mutations.

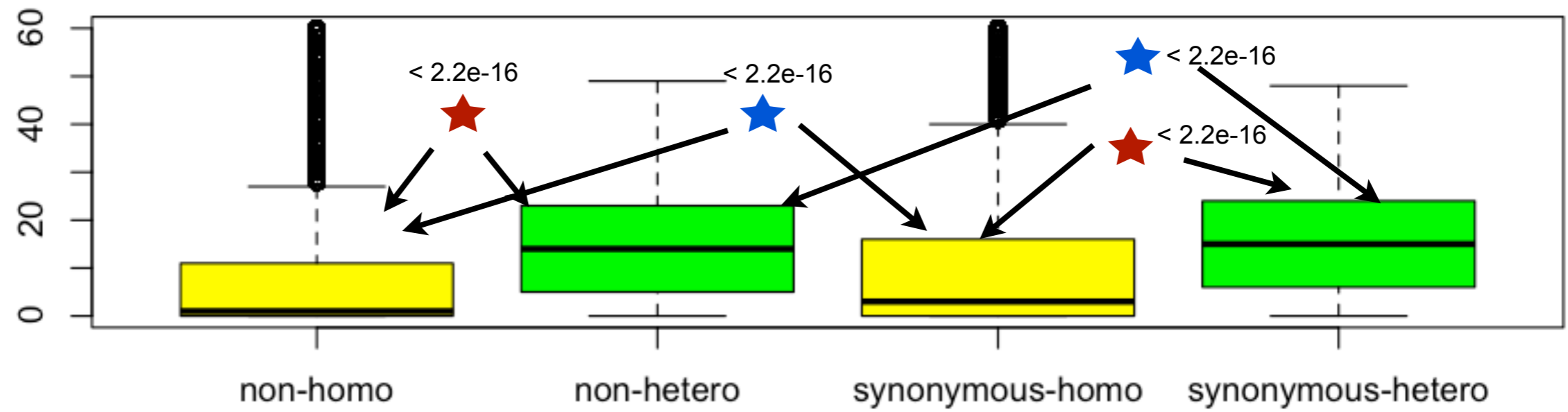
Category	#nonsynonymous SNP	#corresponding genes
CEU	28089	10777
CHBJPT	23255	9840
YRI	37796	11978

- Network: Human regulatory network; Human protein-protein interaction network; Human genetic interaction network.

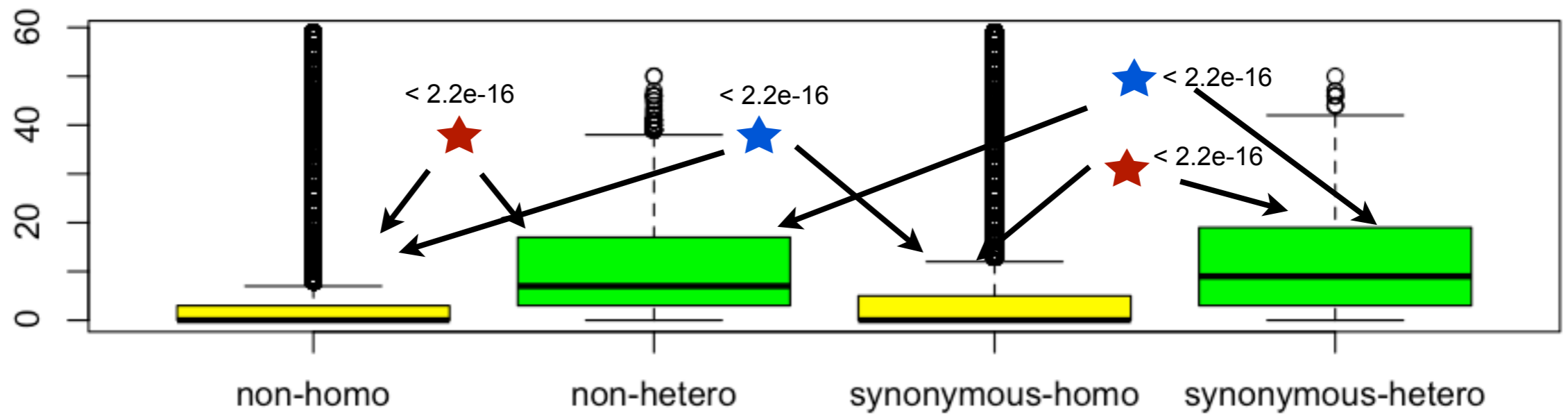
CEU



CHBJPT



YRI



1000 Genome & Human regulatory network

- Network from Nitin
- 6896 regulating relationships
- Nodes are classified into four layers, according to their relationships.

level	#number	function
Top	58	Regulators
Middle	52	Regulators
Bottom	27	Regulators
Bottom most	3030	Targets

- CEU non-synonymous mutated genes

level	# total	# found	Percentage
Top	58	18	0.31
Middle	52	11	0.21
Bottom	27	7	0.26
Bottom most	3030	1215	0.40

- Chi-square test

category	p -value	no significance
CEU non-synonymous	0.07133	
CHBJPT non-synonymous	0.2854	
YRI non-synonymous	0.4813	

- CEU non-synonymous mutated genes

- Chi-square test

category	p -value	no significance
CEU non-synonymous	0.07133	
CHBJPT non-synonymous	0.2854	
YRI non-synonymous	0.4813	

- CEU non-synonymous mutated genes

- CEU non-synonymous mutated genes

level	# total	# found	Percentage
Top three layers	137	36	0.26
Bottom most	3030	1215	0.40

- CEU non-synonymous mutated genes

level	# total	# found	Percentage
Top three layers	137	36	0.26
Bottom most	3030	1215	0.40

- Fisher test

- CEU non-synonymous mutated genes

level	# total	# found	Percentage
Top three layers	137	36	0.26
Bottom most	3030	1215	0.40

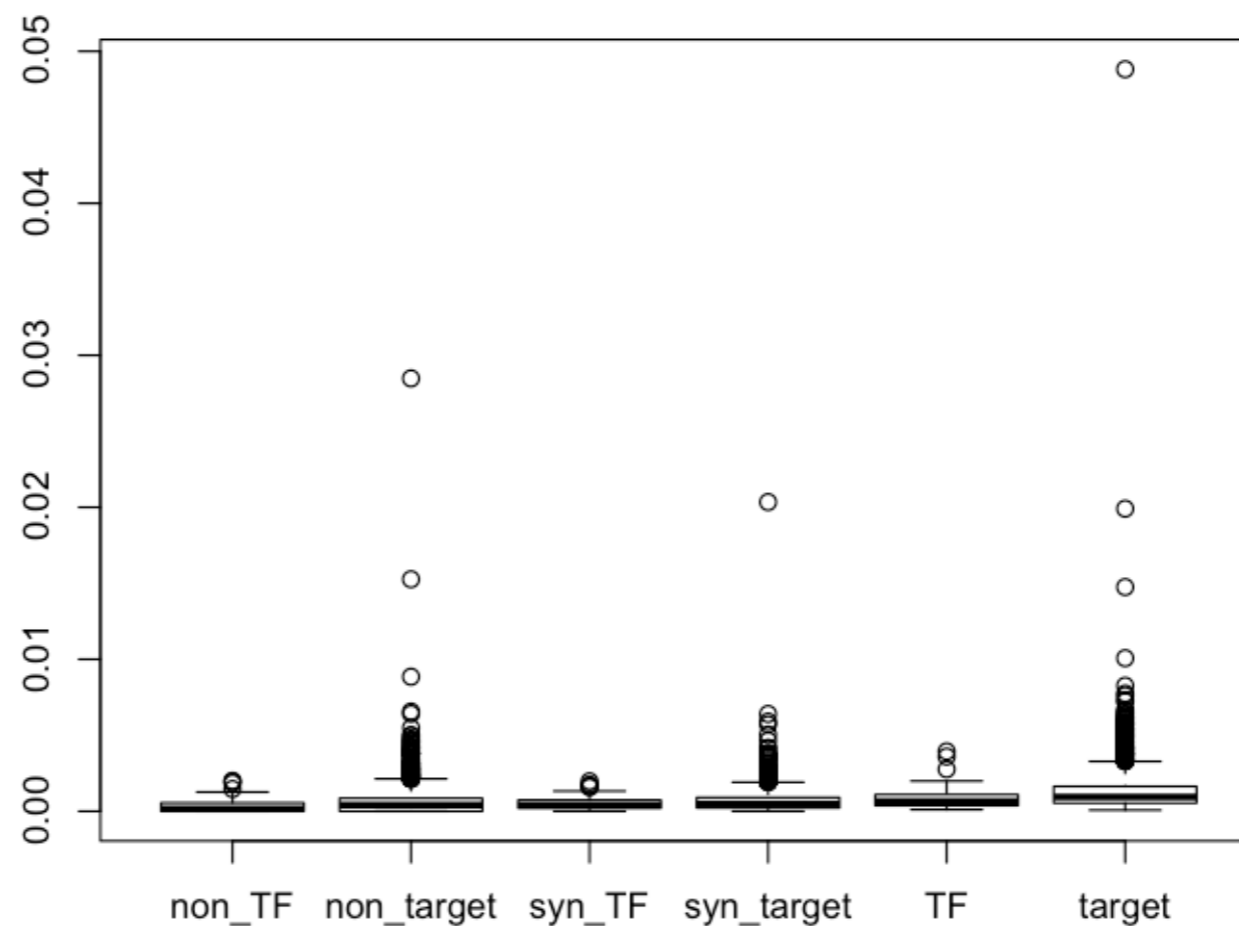
- Fisher test

category	p -value
CEU non-synonymous	0.014
CHBJPT non-synonymous	0.056
YRI non-synonymous	0.116

Regulatory network & Number of SNP per gene

$$\frac{\# \text{SNP}}{\text{Length}(\text{Gene})}$$

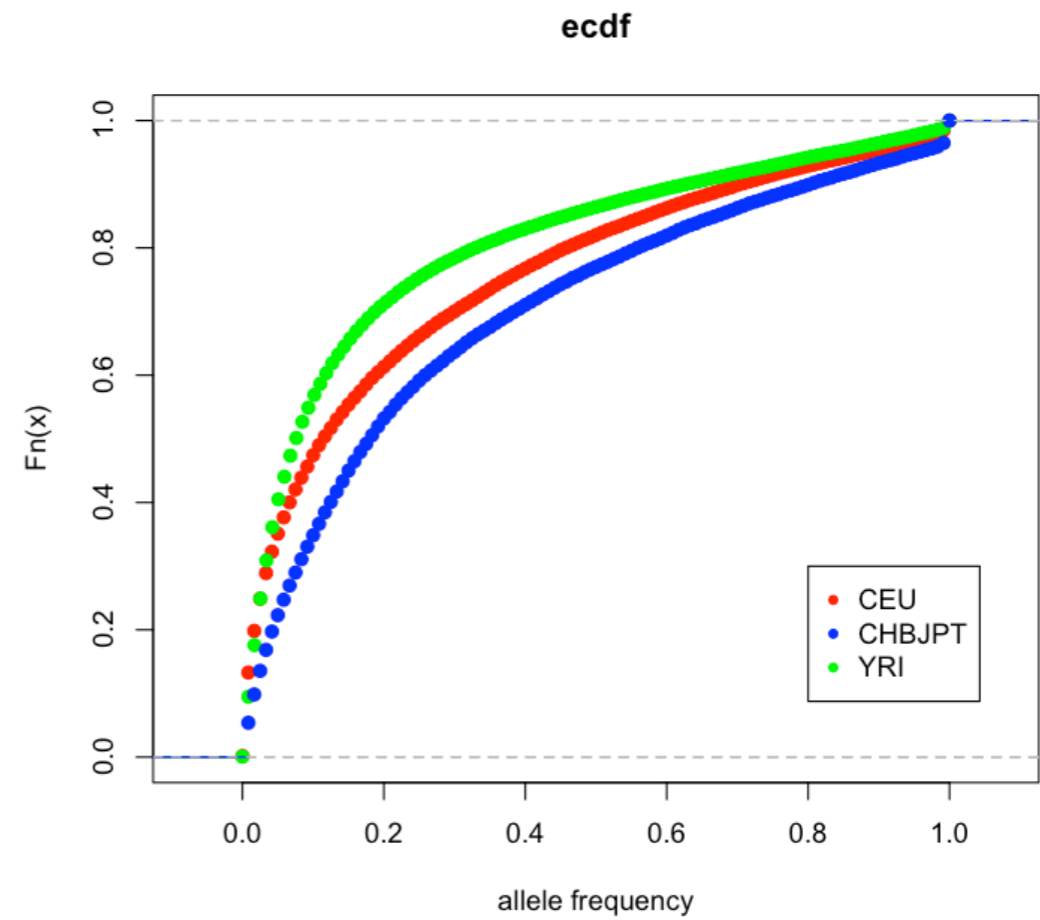
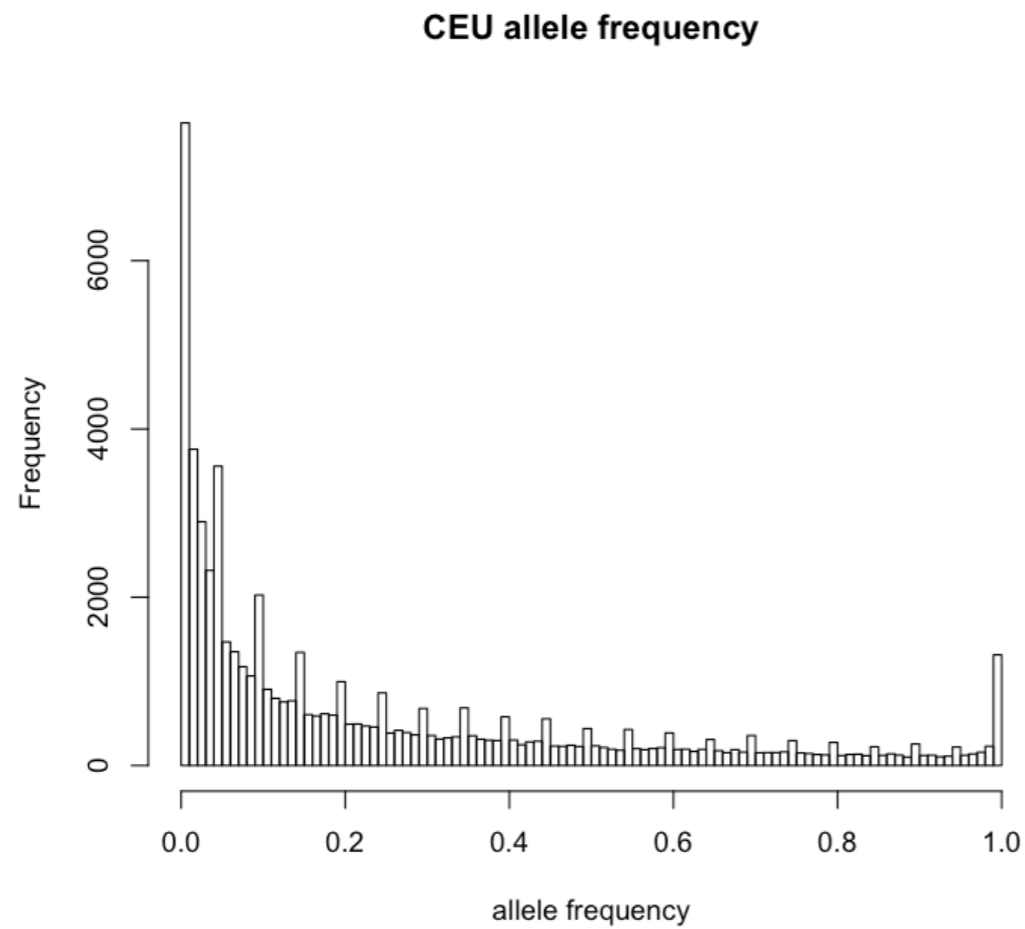
TF Vs. Target



Wilcoxon rank sum test:

Category	CEU	CHBJPT	YRI
non-synonymous	0.001124	0.001195	5.76E-05
synonymous	0.2030	0.01299	3.55E-05
all-SNPs	0.0002714	2.31E-06	4.44E-09

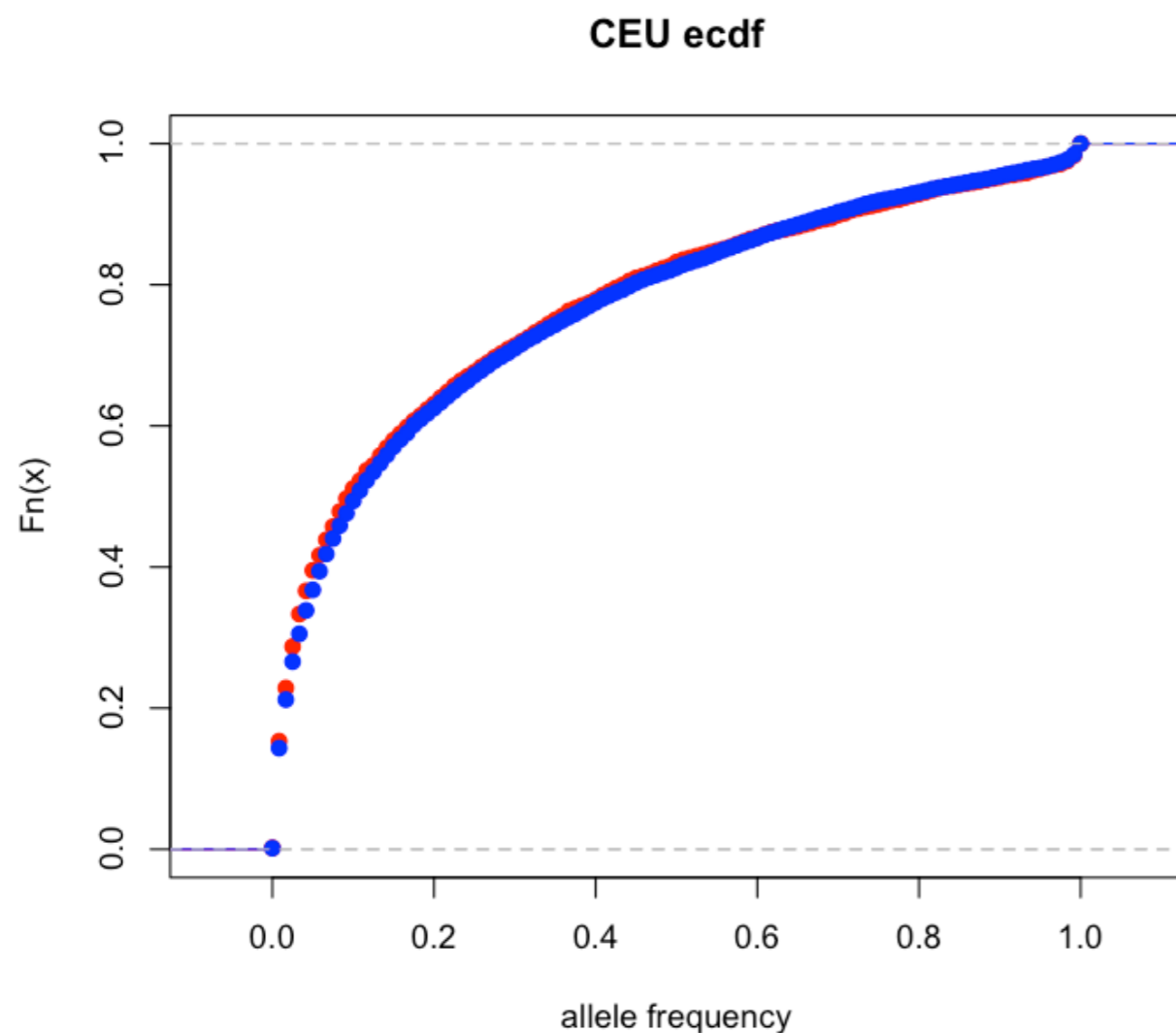
Allele frequency



Ks test: P value $< 2.2e-16$

Allele frequency & Human PPI

- BIOGRID recent release -- BIOGRID-ORGANISM-3.1.73.tab.zip (homo sapiens) 10168 genes, 36852 interactions.
- 20% genes with highest degree centrality are classified as “core set”; others “periphery set”.

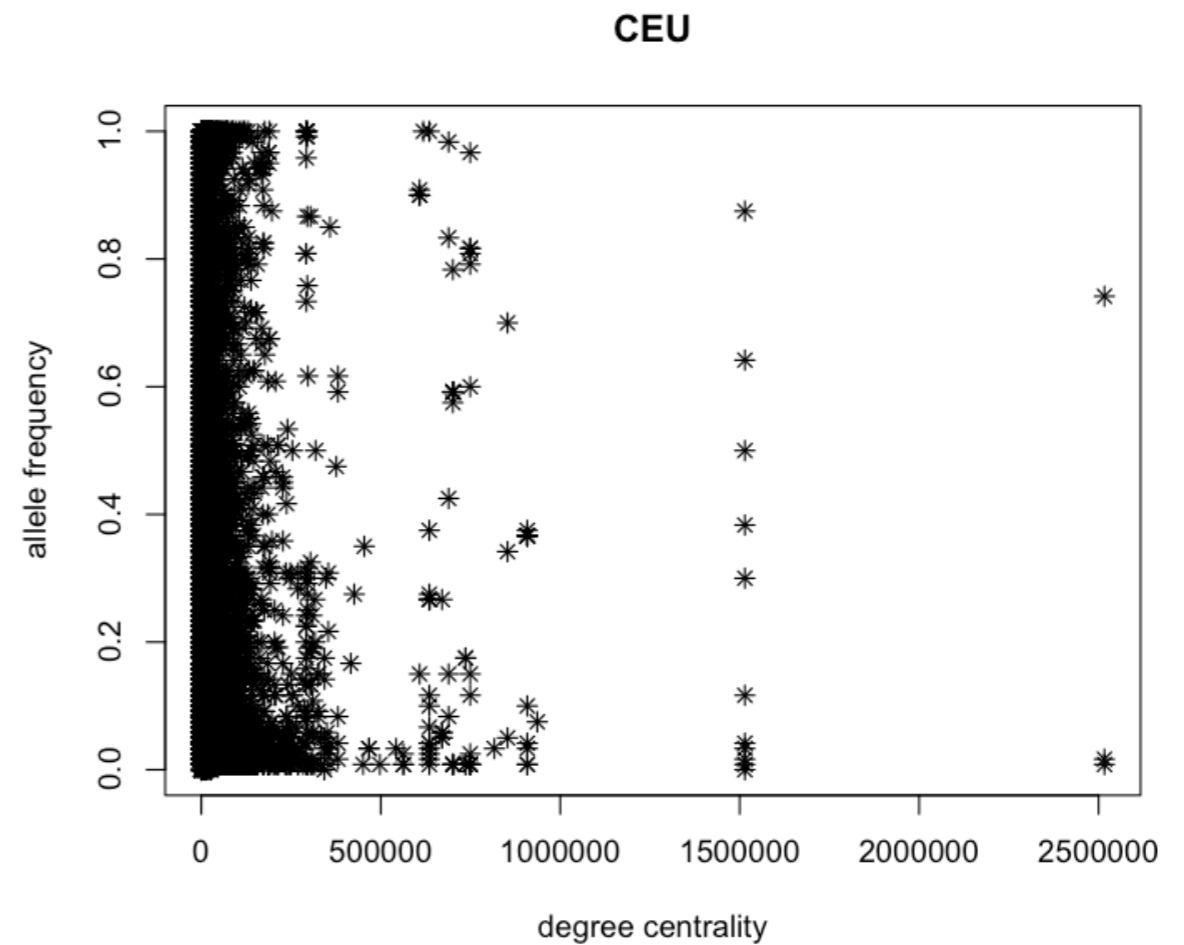
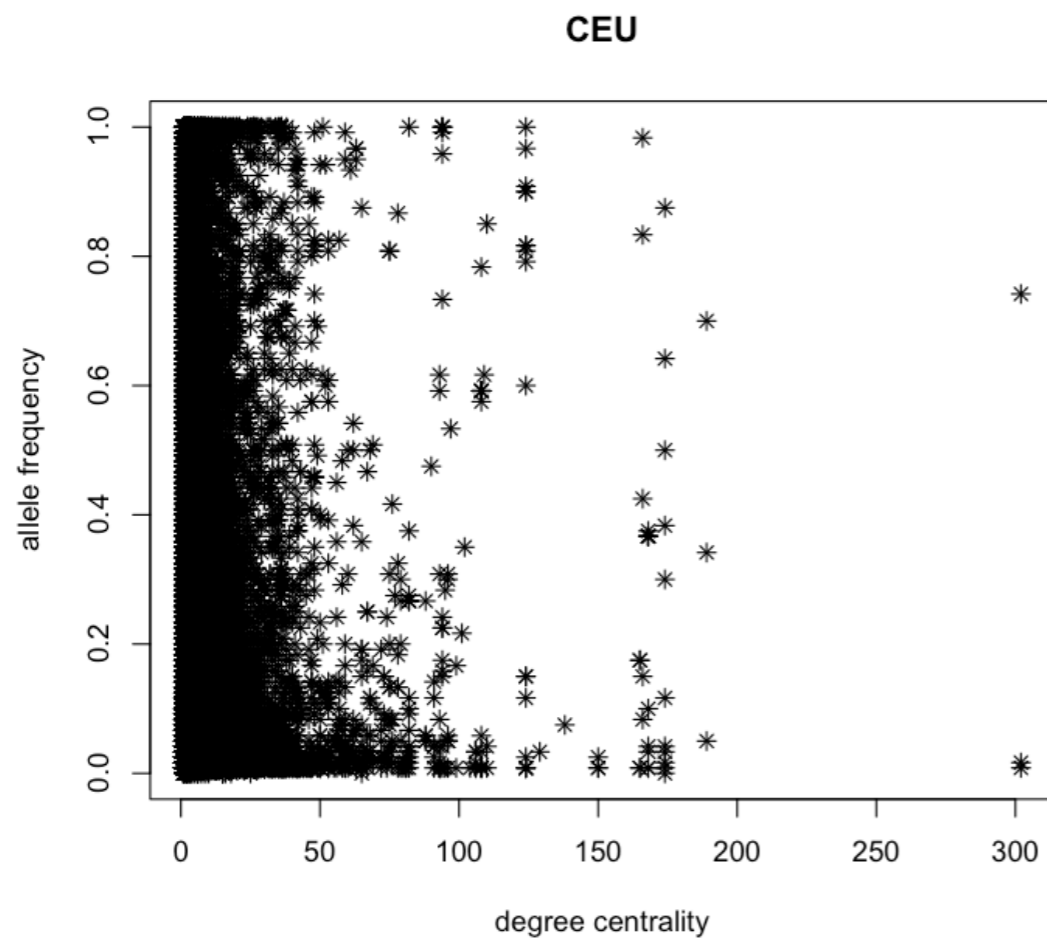


Population	KS test
CEU	$D = 0.0281$, p-value = 0.003698
CHBJPT	$D = 0.0101$, p-value = 0.8869
YRI	$D = 0.0175$, p-value = 0.0683

Allele Frequency & Centrality of Gene (the SNP is in)

□ Degree centrality

□ Betweenness centrality



Spearman Correlation

CEU rho=-0.01608384, p-value = 0.01146

CHBJPT rho=-0.001660419, p-value = 0.8123

YRI rho=-0.009100765, p-value = 0.0921

Spearman Correlation

CEU rho=-0.01416267, p-value = 0.02599

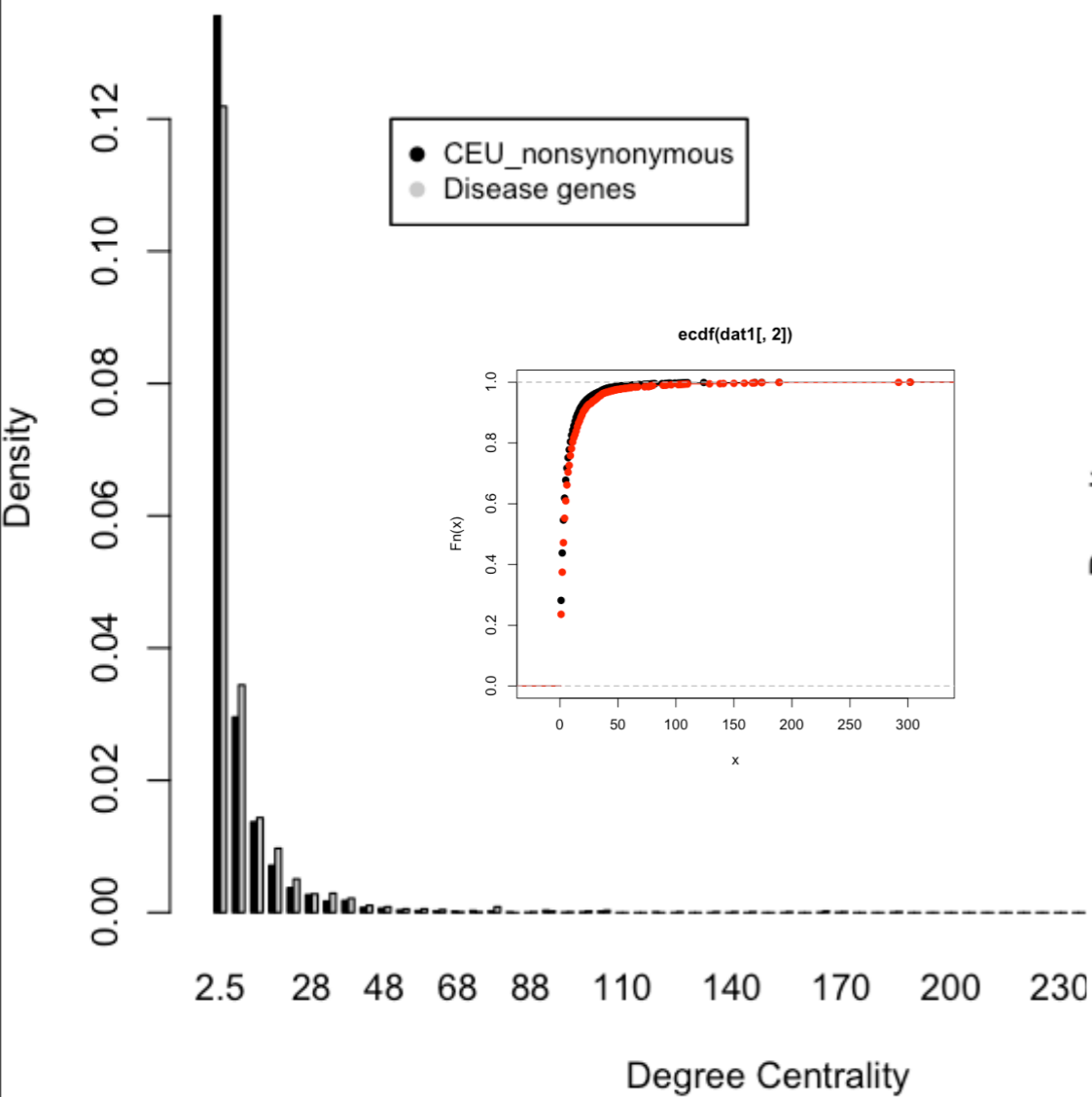
CHBJPT rho=-0.002593034, p-value = 0.7107

YRI rho=-0.003510808, p-value = 0.5158

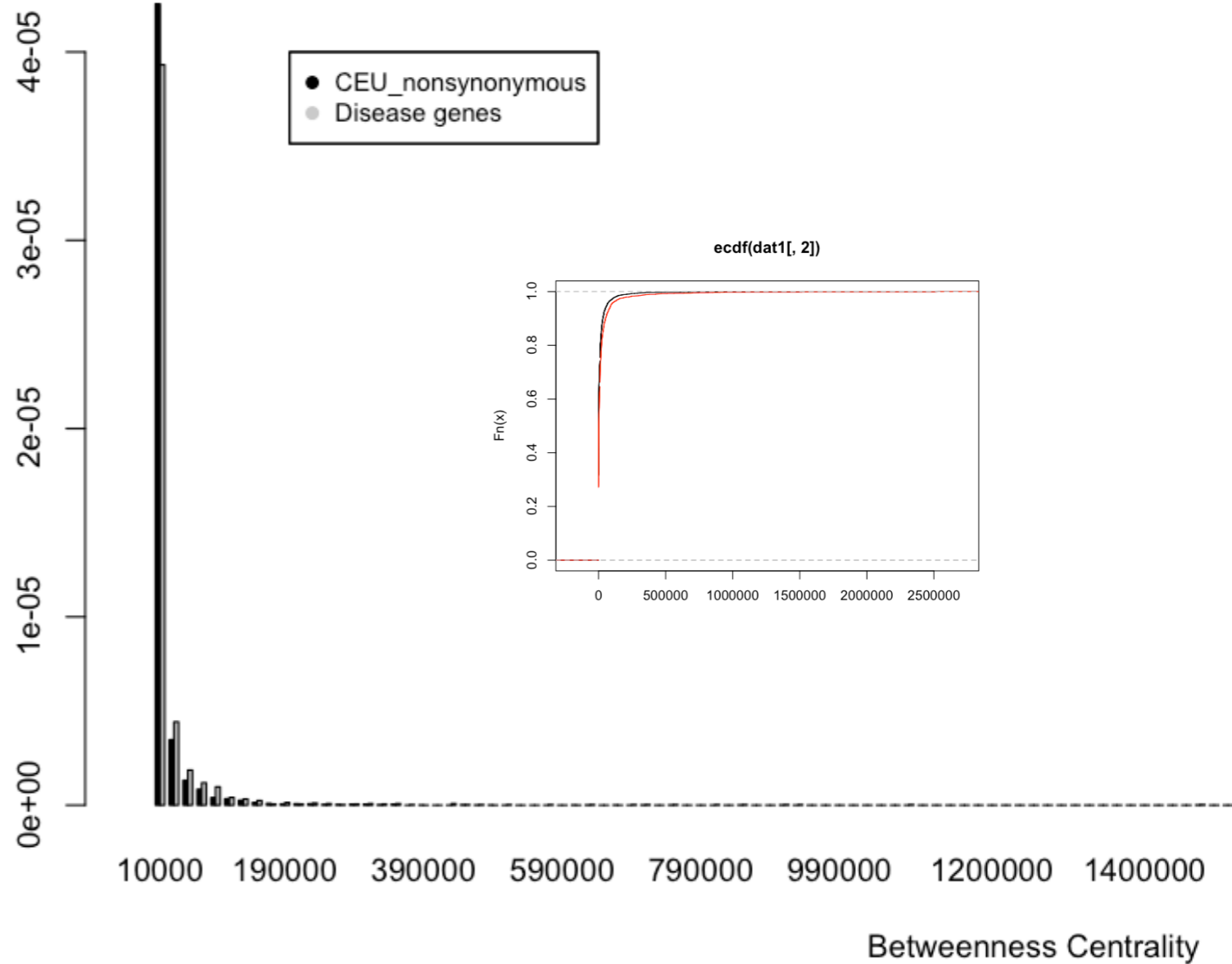
Disease SNPs Vs. 1000 genome SNPs

- **1000 genomes non-synonymous SNPs Data**
- **Disease SNPs Data:** HGMD. Disease related non-synonymous mutations. 54076 SNPs involved in 3014 genes.

Non-synonymous mutated genes Vs. Disease genes in human PPI



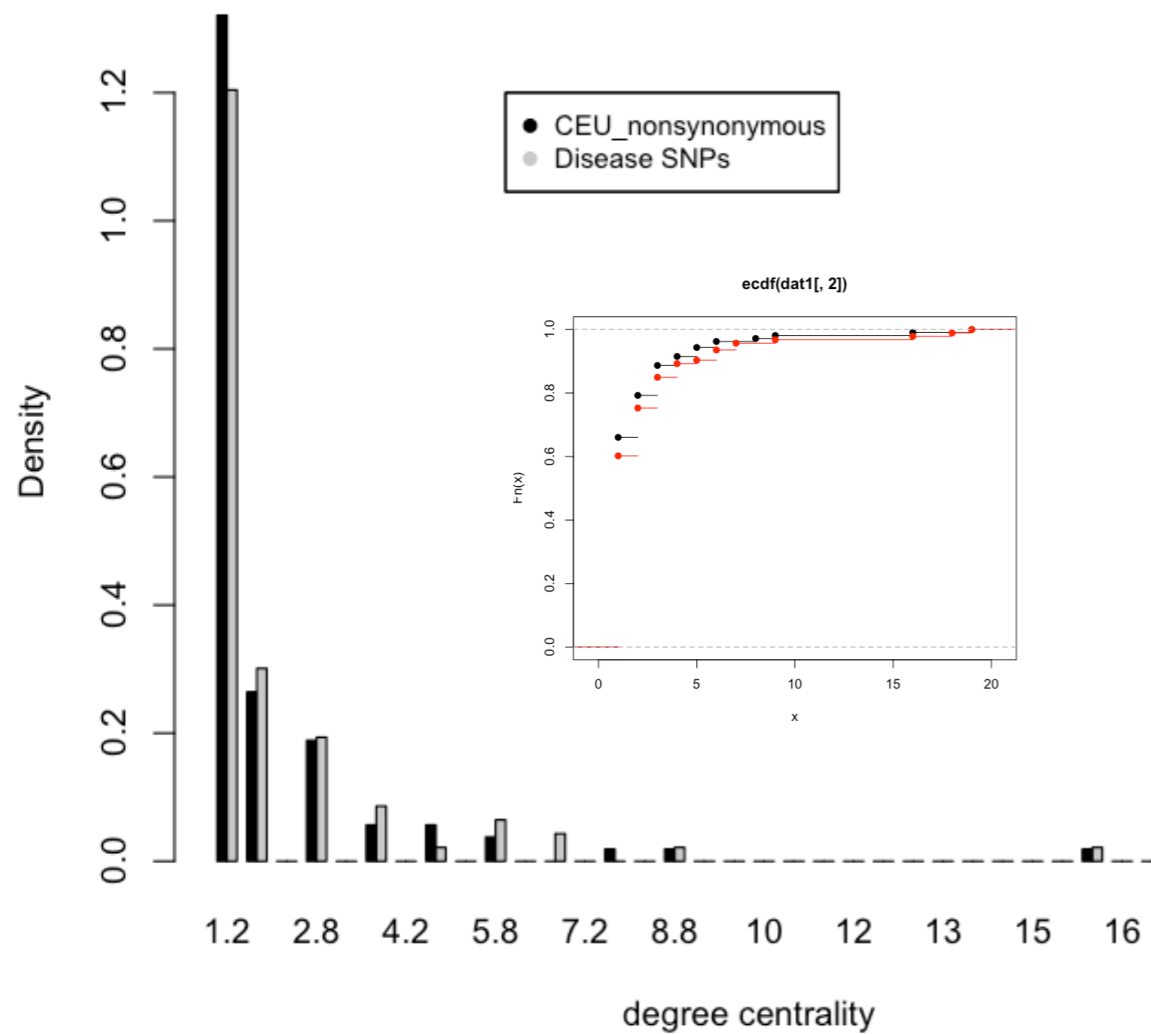
Ks test: CEU: p-value = $1.178e-06$, $D=0.0744$
 CHBJPT: p-value = $3.259e-06$, $D=0.0727$
 YRI: p-value = $2.938e-06$, $D=0.071$



Ks test: CEU: p-value = $8.227e-14$, $D=0.1091$
 CHBJPT: p-value = $2.034e-13$, $D=0.1089$
 YRI: p-value = $8.067e-13$, $D=0.1035$

Non-synonymous mutated genes Vs. Disease genes in Human genetic interaction network

Human genetic interaction

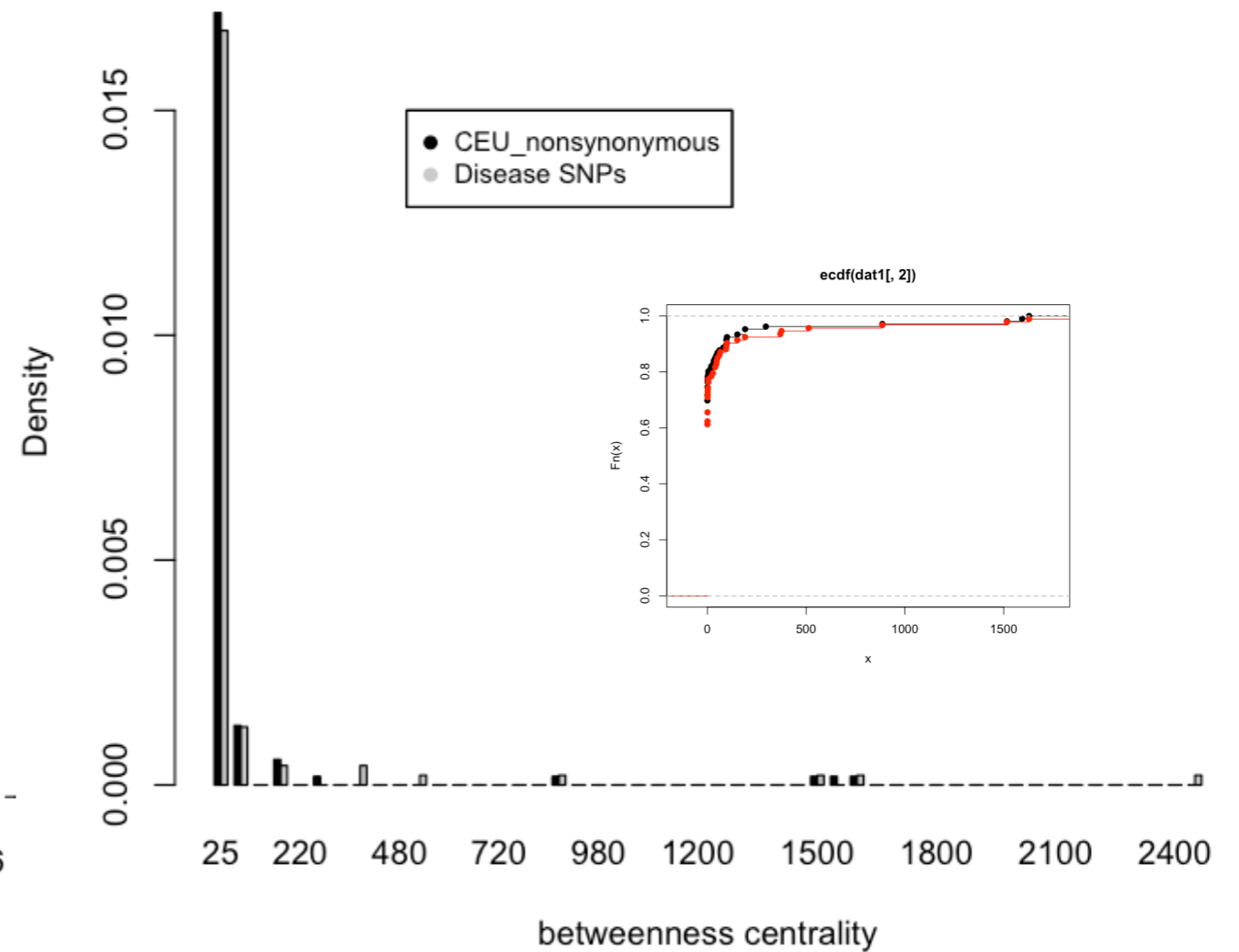


Ks test: CEU: p-value =0.996,D=0.0582

CHBJPT: p-value = 1,D=0.0453

YRI: p-value = 0.986,D=0.0619

Human genetic interaction



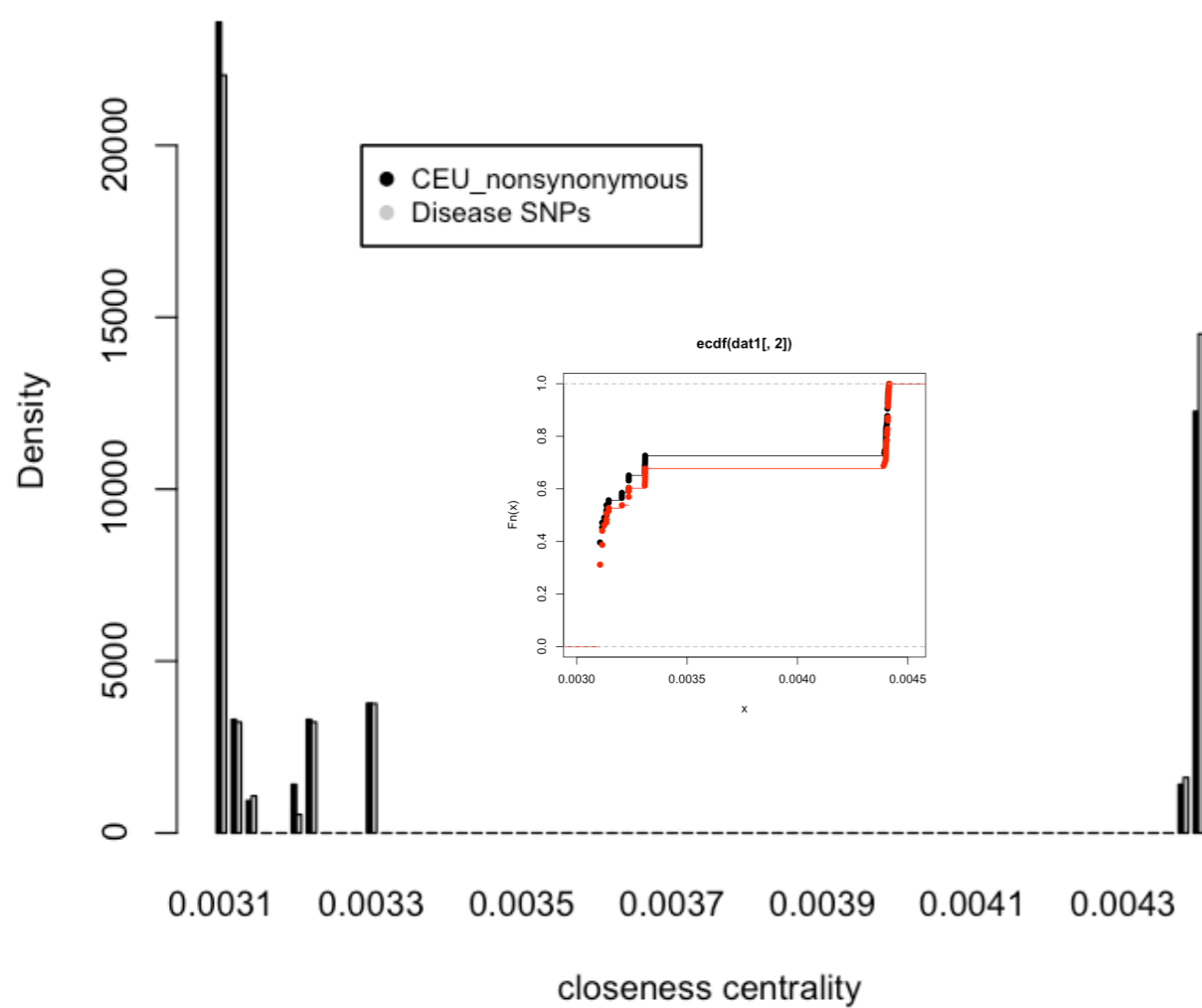
Ks test: CEU: p-value =0.7813,D=0.0933

CHBJPT: p-value = 1, D=0.0289

YRI: p-value = 0.773, D=0.0902

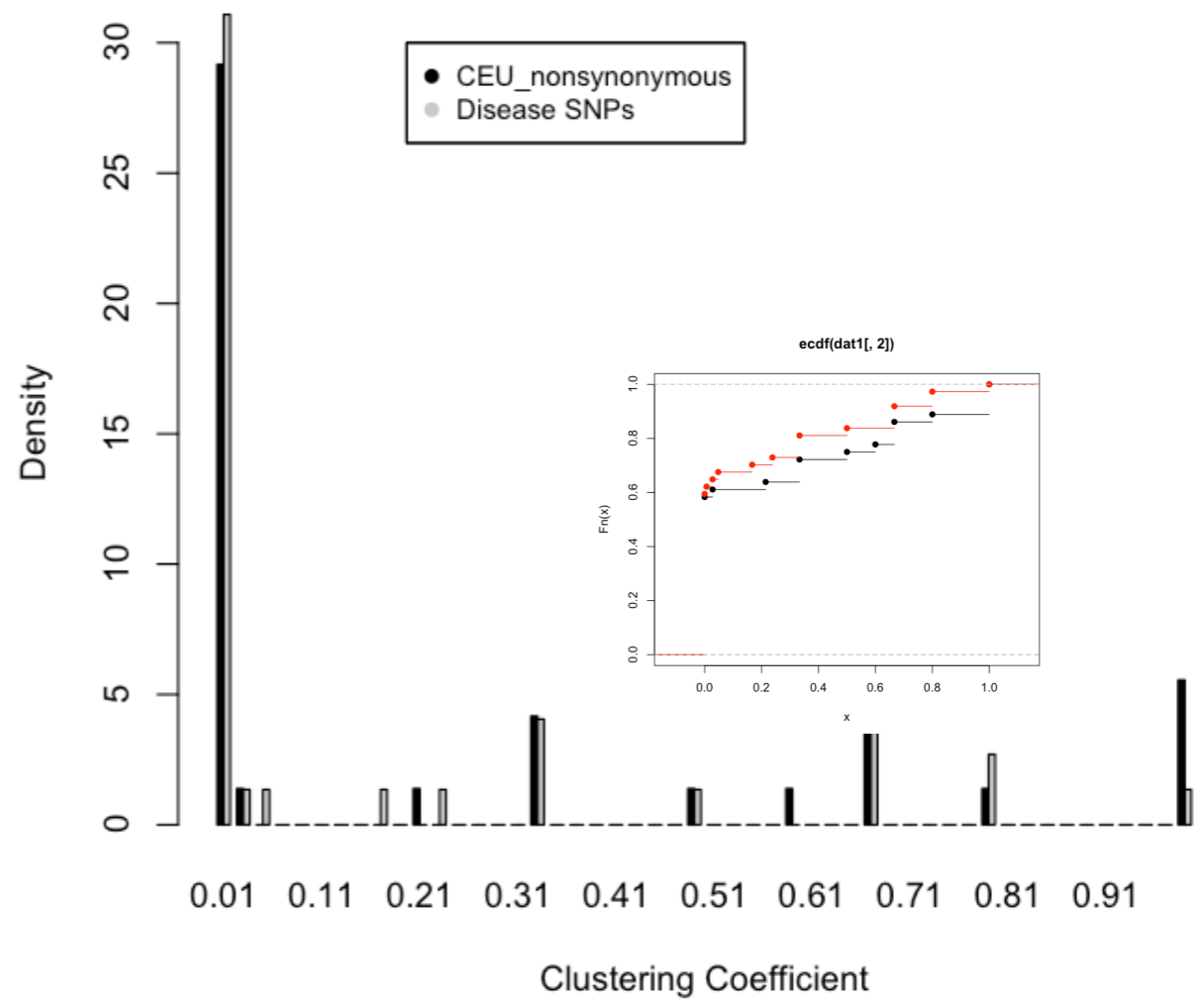
Non-synonymous mutated genes Vs. Disease genes in Human genetic interaction network

Human genetic interaction



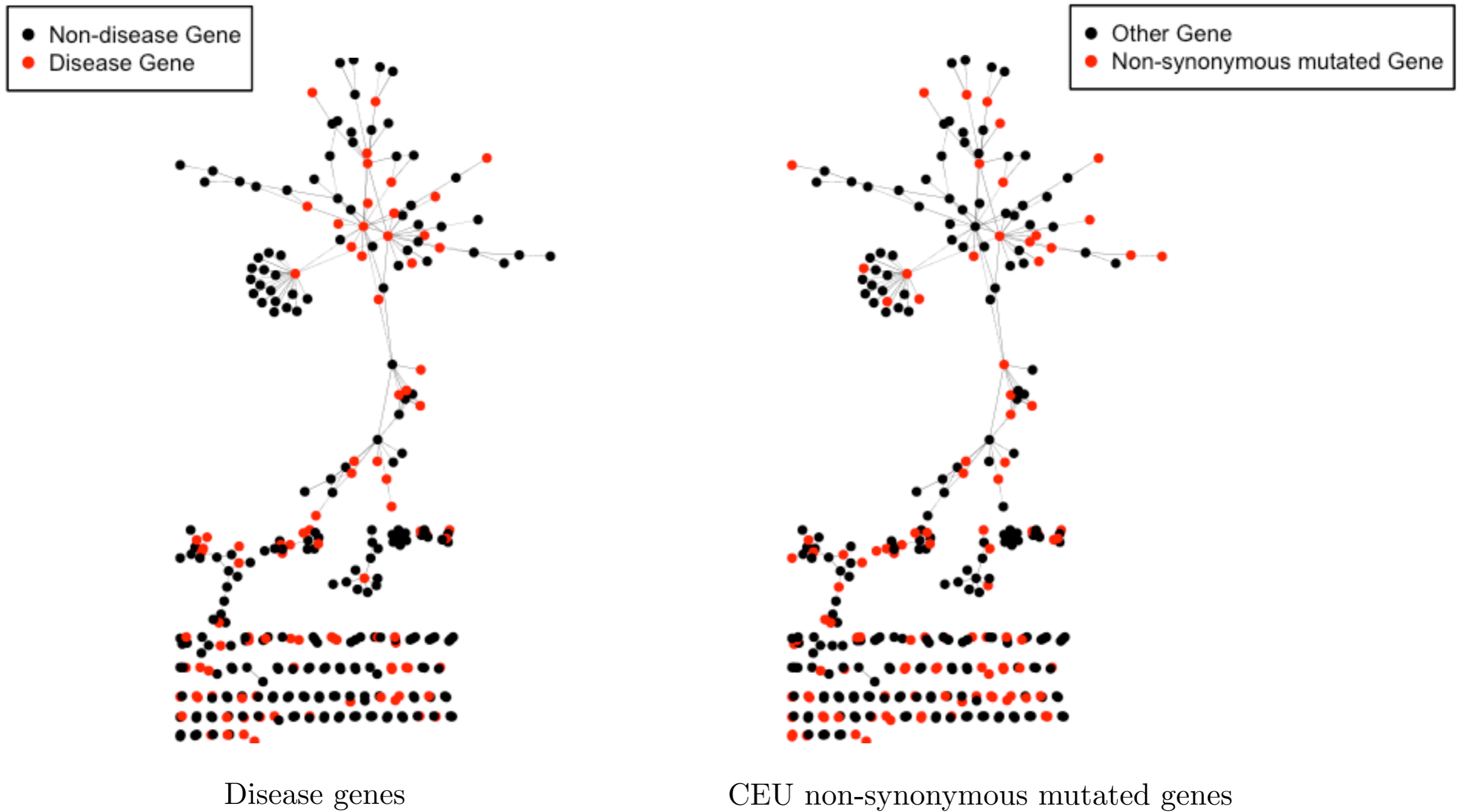
Ks test: CEU: p-value = 0.8721
 CHBJPT: p-value = 0.9984
 YRI: p-value = 0.9518

Human genetic interaction



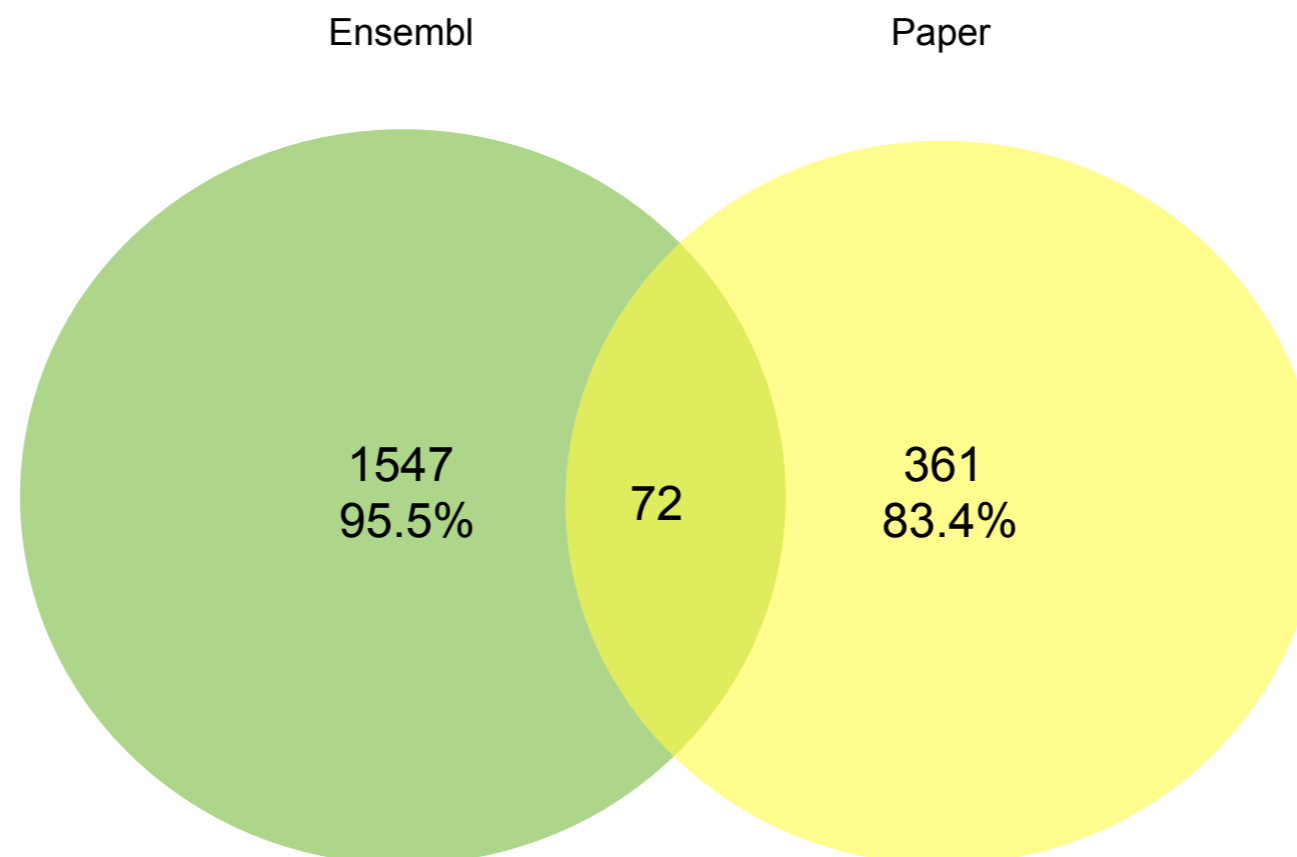
Ks test: CEU: p-value = 0.998
 CHBJPT: p-value = 0.9997
 YRI: p-value = 0.9974

Non-synonymous mutated genes Vs. Disease genes in Human genetic interaction network



Positive Selection

- Positive selection information from Ensembl -- dn and ds between chimp and human.
 - Positive selected genes are defined as the ones with $dn/ds > 1$
- Paper: **Patterns of positive selection in six mammalian genomes** This paper identify 544 genes (gene name from Vega, UCSC known gene, Refseq
- Ensembl match. Download from Ensembl website. 411 genes have Ensembl IDs. Manually match 22, total=433.



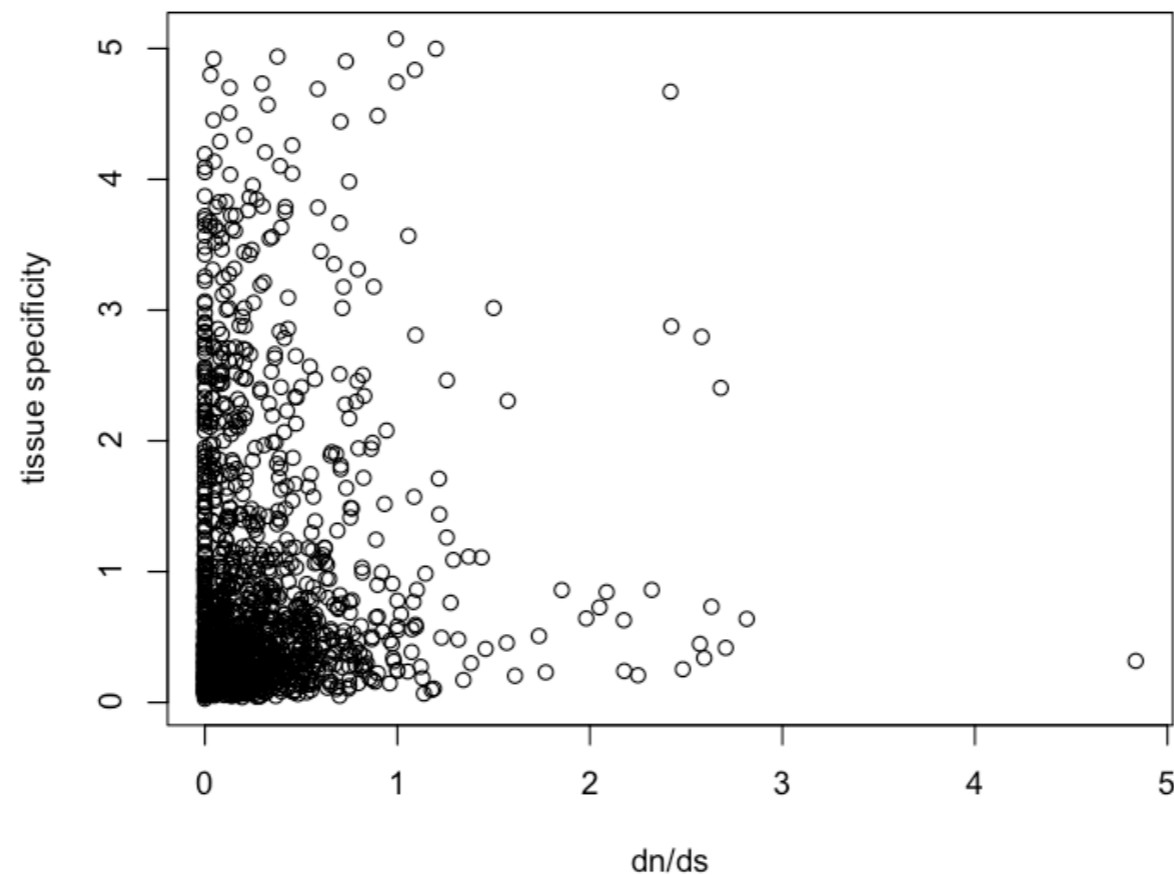
Positive selection & tissue specificity

Data: positive selection(Ensembl) & tissue specificity (1984 TFs in 34 tissues).

Correlate dn/ds and tissue specificity score--spearman correlation

Result: p-value = 0.0004633; rho=0.09198712

Higher dn/ds score genes have higher tissue specificity scores



Positive selection (intra species Vs. inter-species)

McDonald and Kreitman's test: $\frac{p_n}{p_s} > \frac{d_n}{d_s}$

- Download Human and Chimp coding sequences.
- Find orthologs
- Sequence Alignment (MUSCLE)
- Find fixed synonymous & replacement (remove SNPs 1000 genomes)
- Human polymorphism data
- Calculate α & Fisher-test $\alpha = 1 - \frac{D_s P_n}{D_n P_s}$

Result

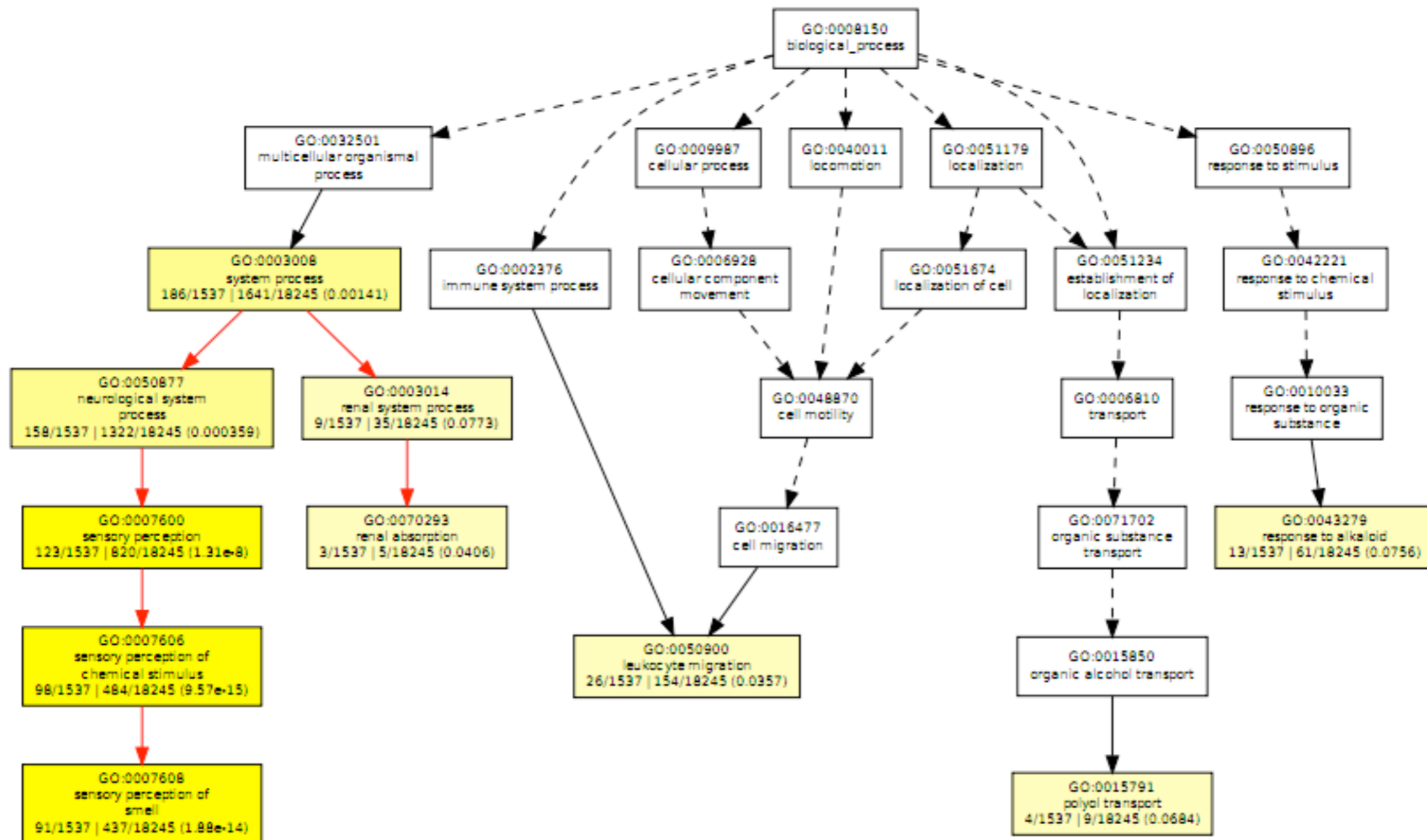
- 30,859 transcripts analyzed
- Fisher test ($p < 0.05$): 1,109 transcripts, involved in 671 genes.
- Human-chimp $dn/ds > 1$: 1497 genes.
- 2136 genes total.

GOEAST: GO enrichment

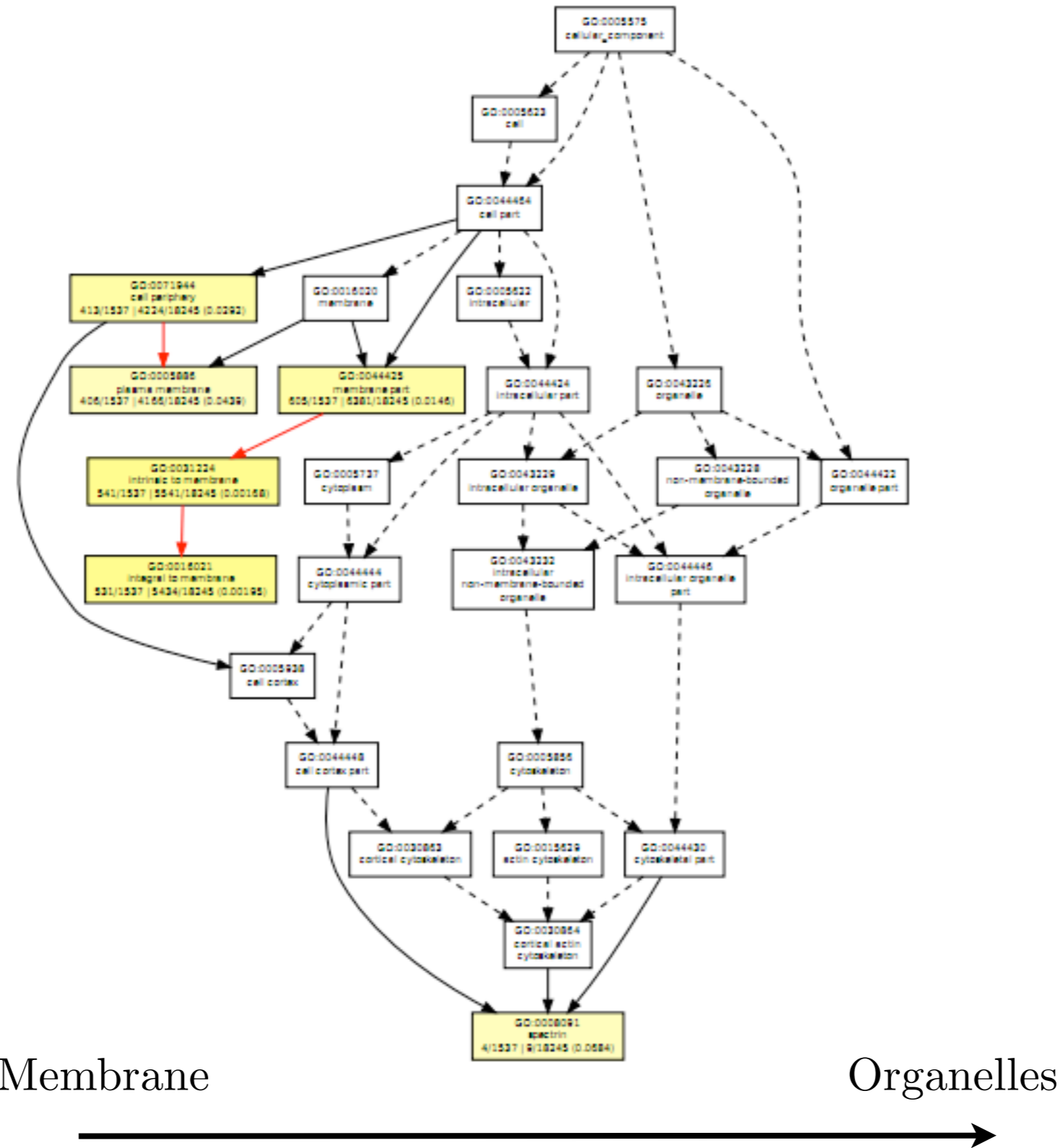
GOID	Ontology	Term	p
GO:0071944	cellular_component	cell periphery	0.0291902844048018
GO:0005886	cellular_component	plasma membrane	0.0439158145846039
GO:0003008	biological_process	system process	0.00141470500289491
GO:0050877	biological_process	neurological system process	0.000359499733367988
GO:0016021	cellular_component	integral to membrane	0.0019538102993518
GO:0031224	cellular_component	intrinsic to membrane	0.00168052213793801
GO:0044425	cellular_component	membrane part	0.0145547468776665
GO:0003014	biological_process	renal system process	0.077266729166669
GO:0004871	molecular_function	signal transducer activity	5.0035913601629e-10
GO:0004872	molecular_function	receptor activity	1.66789730768034e-15
GO:0060089	molecular_function	molecular transducer activity	5.0035913601629e-10
GO:0004888	molecular_function	transmembrane receptor activity	1.06900427031169e-10
GO:0004930	molecular_function	G-protein coupled receptor activity	3.55699640980892e-09
GO:0004984	molecular_function	olfactory receptor activity	5.37696486841723e-14
GO:0007600	biological_process	sensory perception	1.31177694406180e-08
GO:0007606	biological_process	sensory perception of chemical stimulus	9.5685426641483e-15
GO:0007608	biological_process	sensory perception of smell	1.88282319133241e-14
GO:0050900	biological_process	leukocyte migration	0.0356568514268142
GO:0043279	biological_process	response to alkaloid	0.0756289090076829
GO:0005035	molecular_function	death receptor activity	0.098314898060769
GO:0008091	cellular_component	spectrin	0.0684218796155176
GO:0005218	molecular_function	intracellular ligand-gated calcium channel activity	0.040625858420242
GO:0005219	molecular_function	ryanodine-sensitive calcium-release channel activity	0.040625858420242
GO:0015101	molecular_function	organic cation transmembrane transporter activity	0.0185066242696302
GO:0008519	molecular_function	ammonium transmembrane transporter activity	0.040625858420242
GO:0015791	biological_process	polyol transport	0.0684218796155176
GO:0008109	molecular_function	N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl transferase activity	0.000765895629521642
GO:0005087	molecular_function	Ran guanyl-nucleotide exchange factor activity	0.040625858420242
GO:0005132	molecular_function	interferon-alpha/beta receptor binding	0.00154134695161400
GO:0070293	biological_process	renal absorption	0.040625858420242

~

Biological Process



Cellular Component



Acknowledgement

- Special thanks to: Ekta, Koon-Kiu, Nitin, Chong
- Thanks to everyone in the lab.