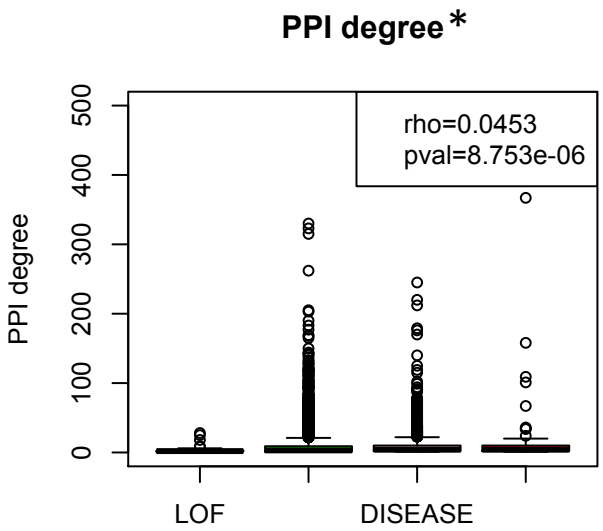
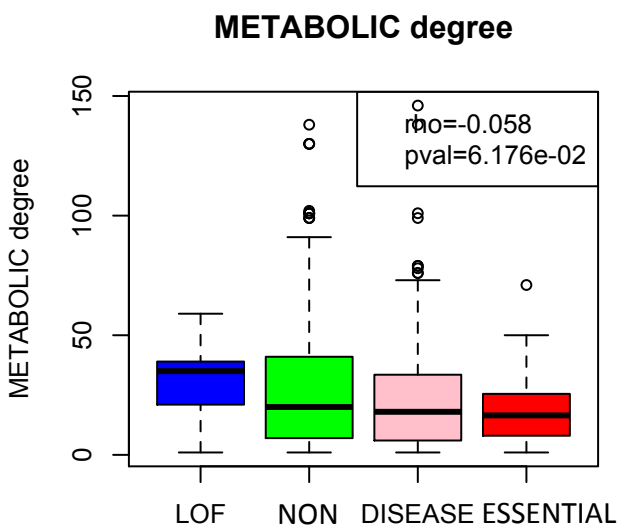
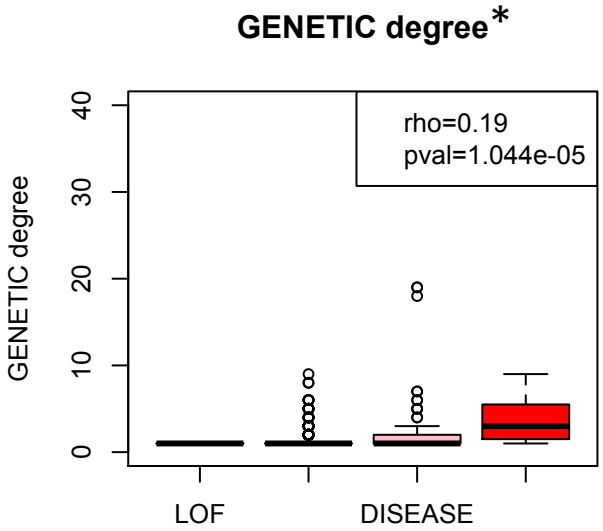
Netsnp

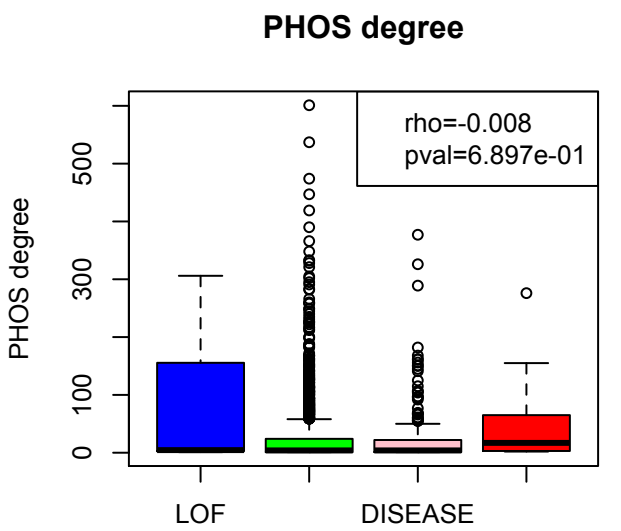
EK

NETWORK CONSTRAINTS ON GENE MUTATION

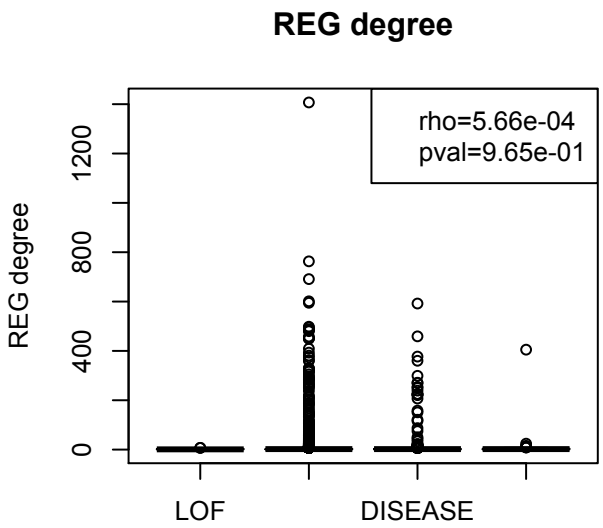


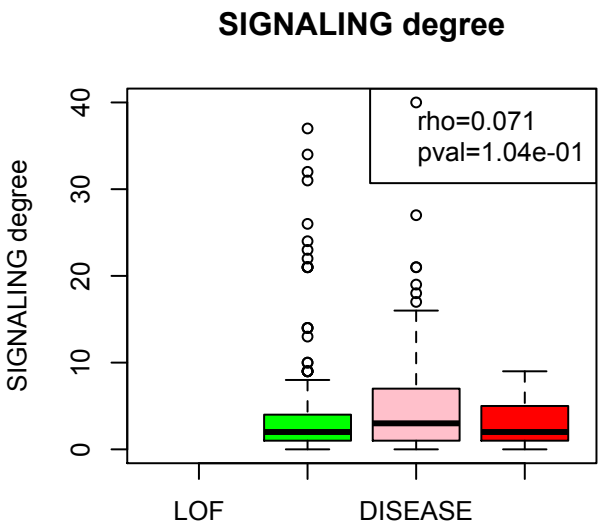


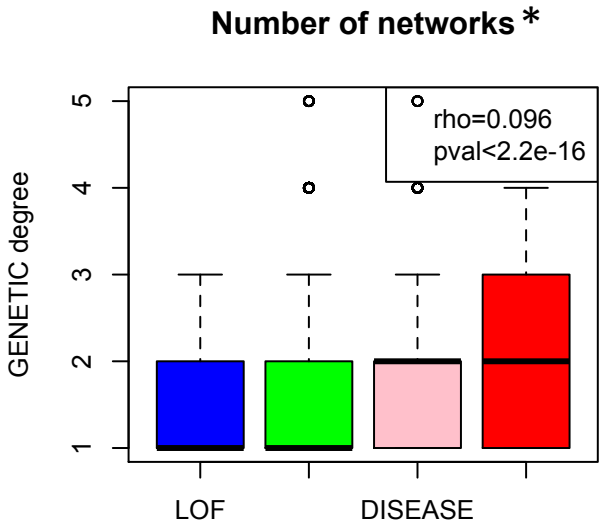


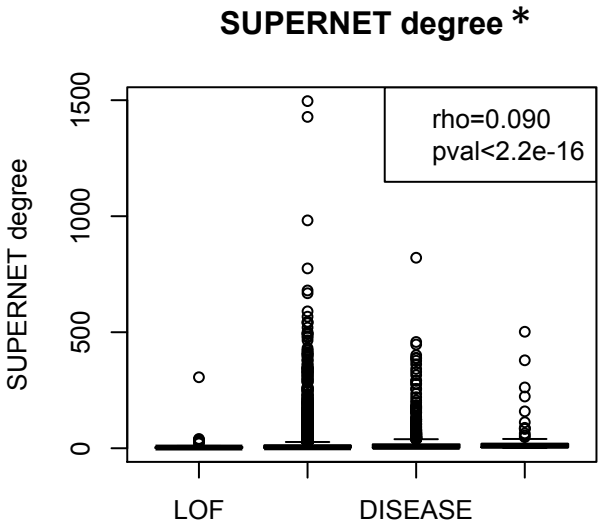


Spearman correlation with susceptibility score
LOF tolerant=-1
Non=0
Disease=1
Essential=2

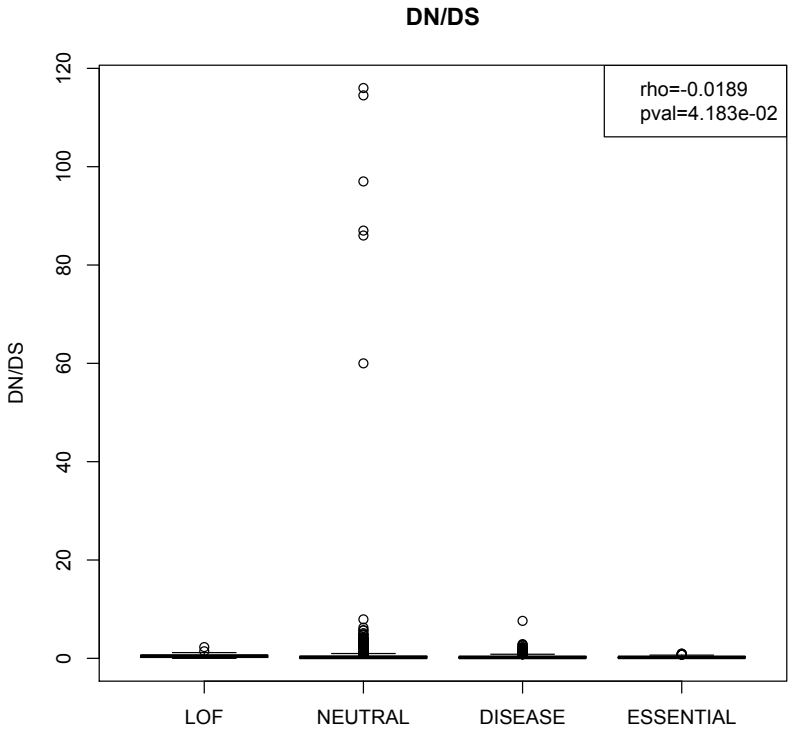








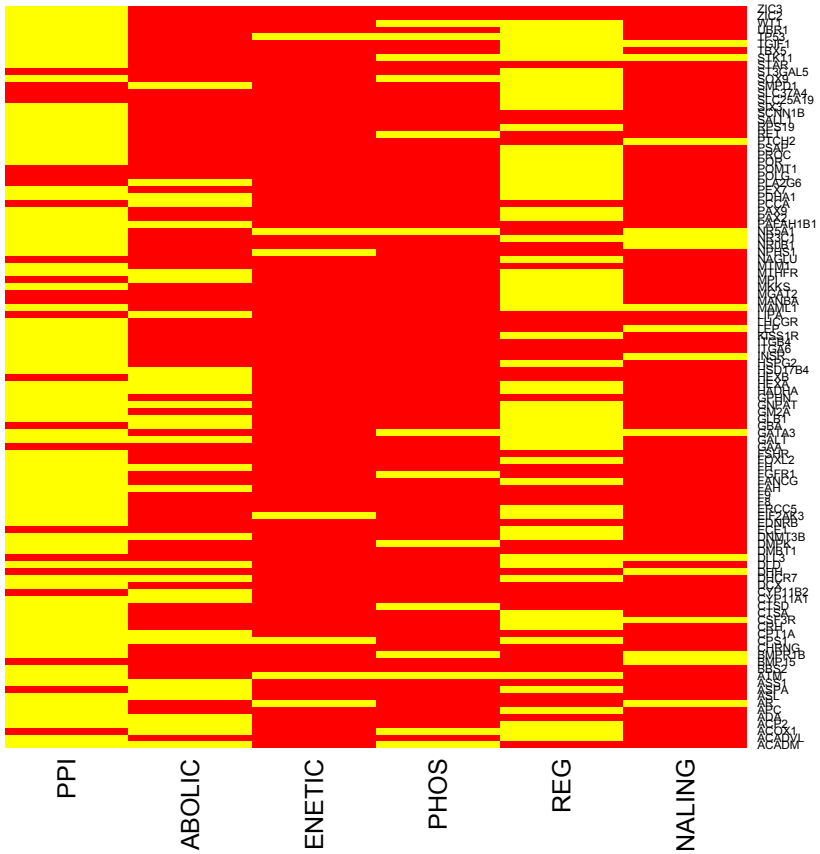
If a gene pair interacts in multiple networks, it is counted once to compute 'Supernet' degree


DN/DS
rho=-0.0189
pval=4.183e-02
DN/DS
0
20
40
60
80
100
120
LOF
NEUTRAL
DISEASE
ESSENTIAL

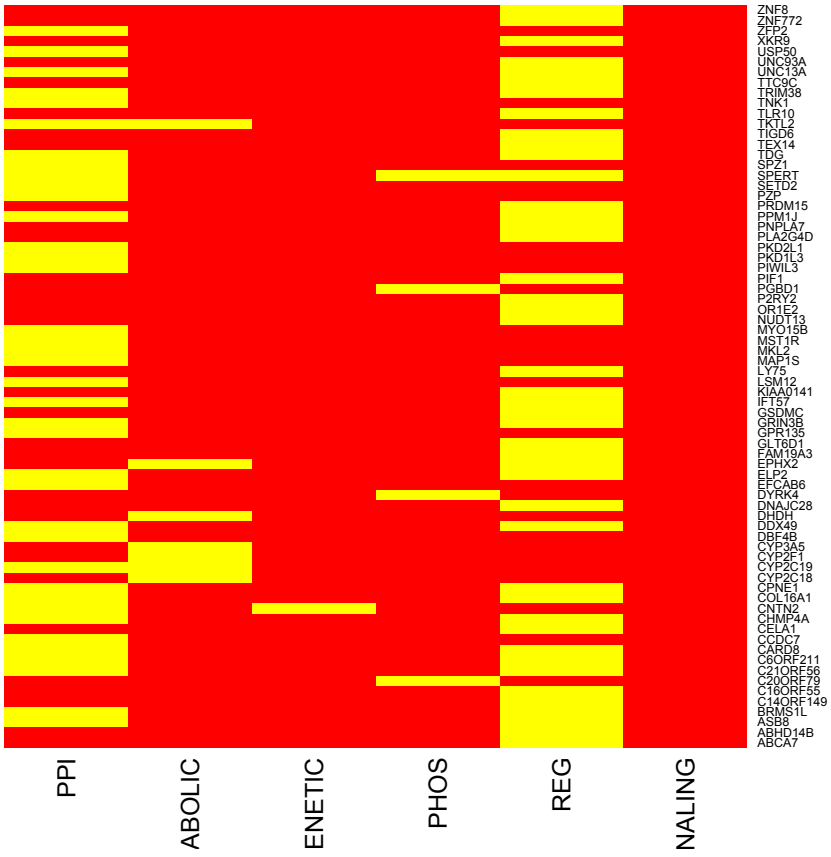
Feature	LOF tolerant	Non-LOF/ Non-disease	Disease	Essential	Pvalue (LOF-Essential)
PPI degree	4.18	8.87	10.08	15.96	**7.68e-04**
Metabolic	30.71	27.57	23.88	19.63	9.35e-02
Genetic	1	1.55	3.14	3.86	1.86e-01
Phosphorylation	79	23.81	27.78	50.64	3.35e-01
Regulatory	2.16	5.85	7.56	10.02	**1.45e-02**
Signaling	NA	3.53	4.90	2.94	NA
Total #of networks	1.31	1.57	1.78	2	**7.85e-09**
dN/dS	0.454	0.375	0.279	0.249	**2.53e-05**
# of interfaces	1.33	1.78	2.17	2.19	**1.34e-01**

Page's trend test, PVAL < 0.001

Network presence of Essential genes



Network presence of LOF genes



HYPOTHESIS: This is actually a result of high network connectivity of Essential genes. Since they are connected to many genes they have a higher chance to be present in networks with limited number of genes

Yellow:present
Red:absent

Networks and population genetics

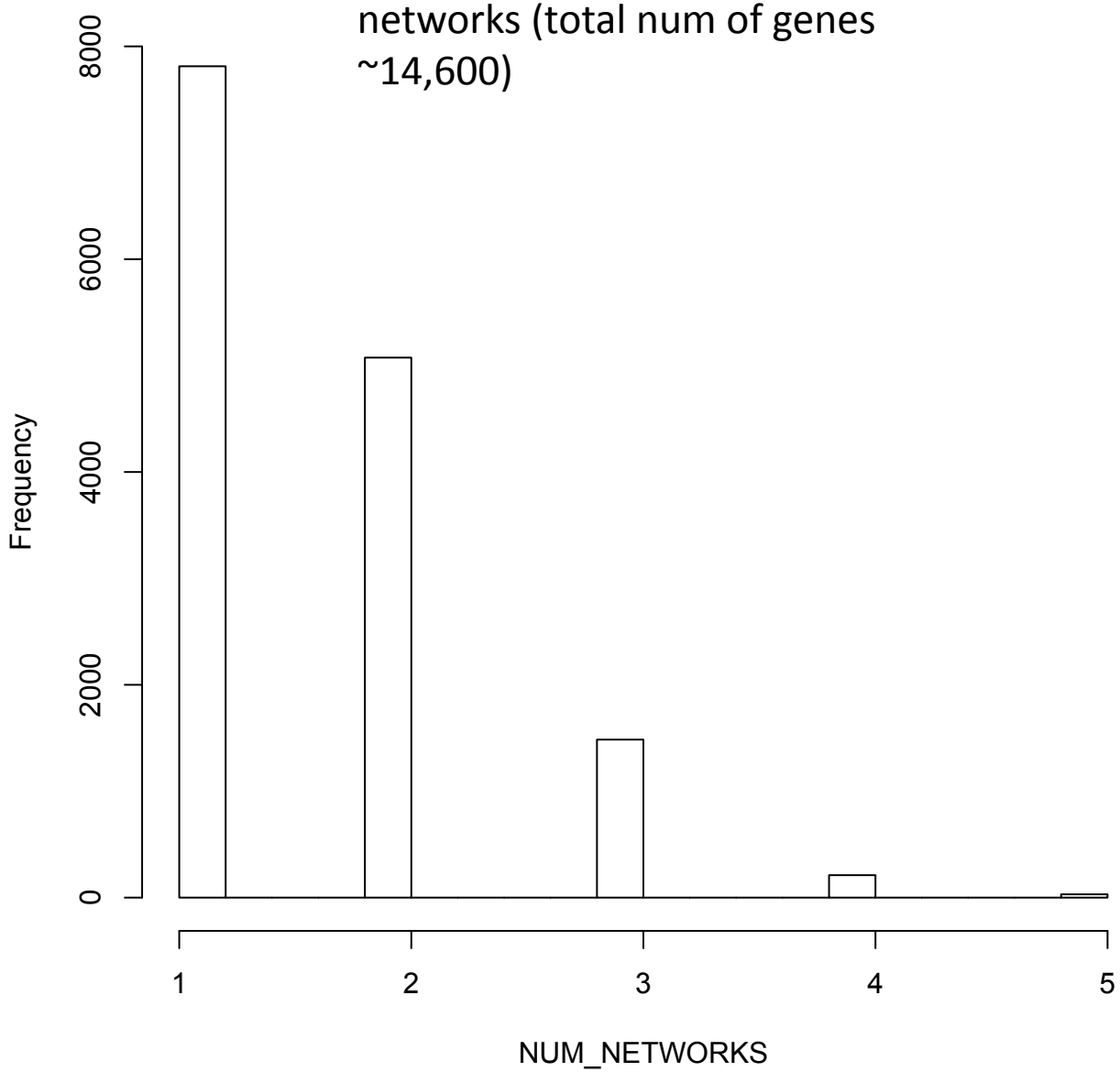
	CEU ALF	CHBJPT ALF	YRI ALF	(CEU+CHBJPT+YRI) ALF
PPI Degree	P: 0.003912; Cor: -0.01789	P: 0.02280; Cor: -0.01412	P: 4.431e-10; Cor: -0.03867	P: 0.000115; Cor: -0.023915
Metabolic Degree	P: 0.8824; Cor: -0.00289	P: 0.7764; Cor: -0.00555	P: 0.09256; Cor: -0.03289	P: 0.1221; P: -0.03023
Genetic Degree	P: 0.6608; Cor: -0.01481	P: 0.9211; Cor: 0.00334	P: 0.5534; Cor: -0.0200	P: 0.2912; Cor: -0.03562
Phosphorylation Degree	P: 0.9295; Cor: -0.00135	P: 0.55; Cor: 0.009107	P: 0.03955; Cor: -0.03136	P: 0.1389; Cor: -0.02255
Encode Nets Degree	P: 0.4235; Cor: -0.00529	P: 0.4631; Cor: -0.004850	P: 0.00572; Cor: -0.01826	P: 0.05381; Cor: -0.01275
Signaling Degree	P: 0.7772; Cor: 0.00834	P: 0.693; Cor: 0.01164	P: 0.7239; Cor: 0.01042	P: 0.9658; Cor: 0.001264
Num_networks	P: 3.434e-6; Cor: -0.0181	P: 0.005133; Cor: -0.01462	P: 3.62e-13; Cor: -0.03796	P: 2.275e-8; Cor: -0.02920
Supernet_degree	P: 0.0005322; Cor: -0.01810	P: 0.0040; Cor: -0.01503	P: <2.2e-16; Cor: -0.04447	P: 4.198e-11; Cor: -0.034466
dN/dS	P: < 2.2e-16; Cor:0.07276	P: 7.69e-15; Cor: 0.04179	P: < 2.2e-16; Cor: 0.1157	P: <2.2e-16; Cor: 0.0944
Num_interfaces	P: 0.8543; Cor : - 0.0035	P: 0.8519; Cor: -0.00356	P: 0.9778; Cor: 0.00053	P: 0.573; Cor: -0.01075

Results obtained from susceptibility scores are consistent with those obtained from allele frequencies of variants in healthy humans

From Yao

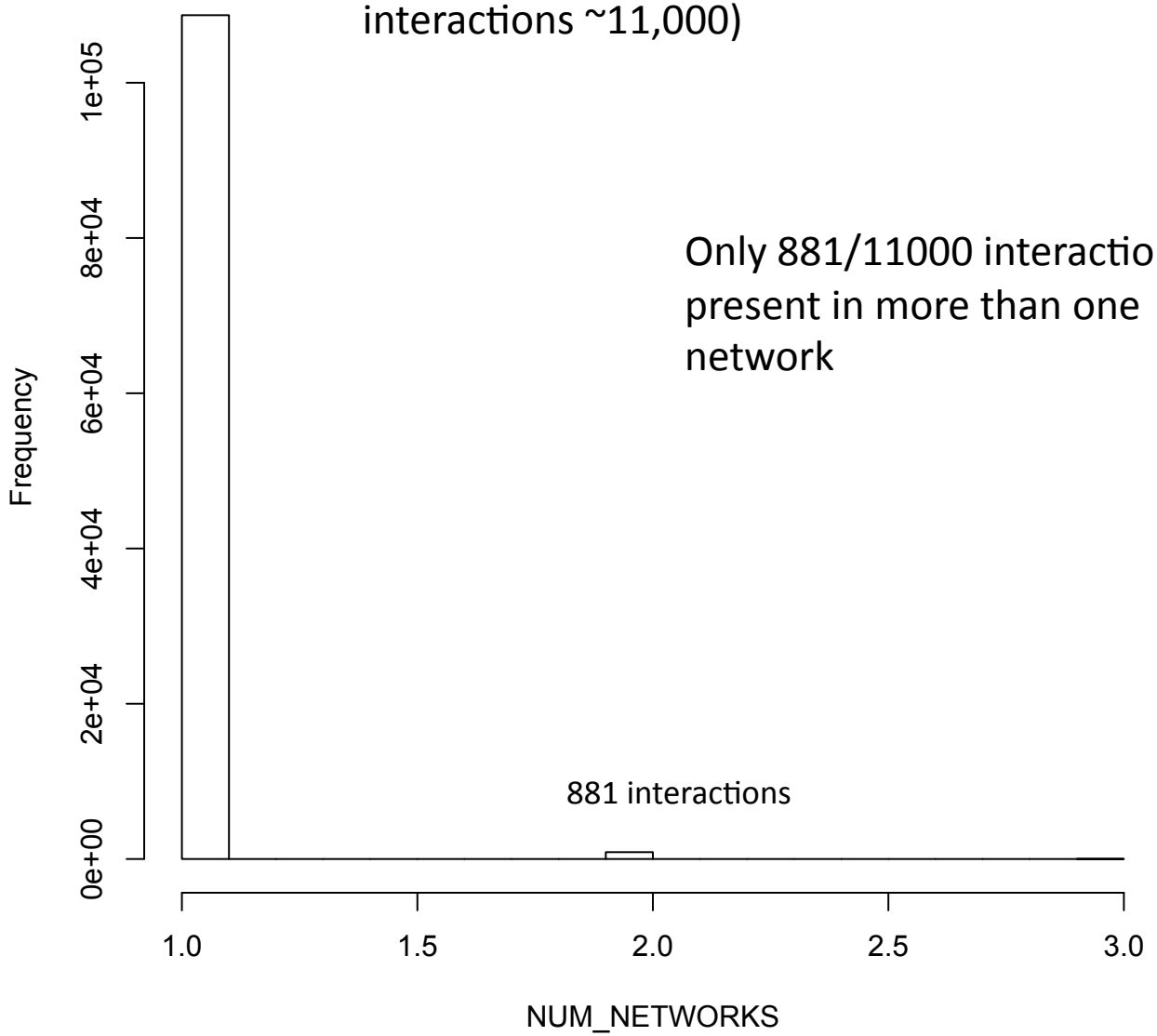
UNDERSTANDING UNDERLYING NETWORK DATA: DO WE HAVE MANY INTERACTIONS COMMON IN DIFFERENT NETWORKS

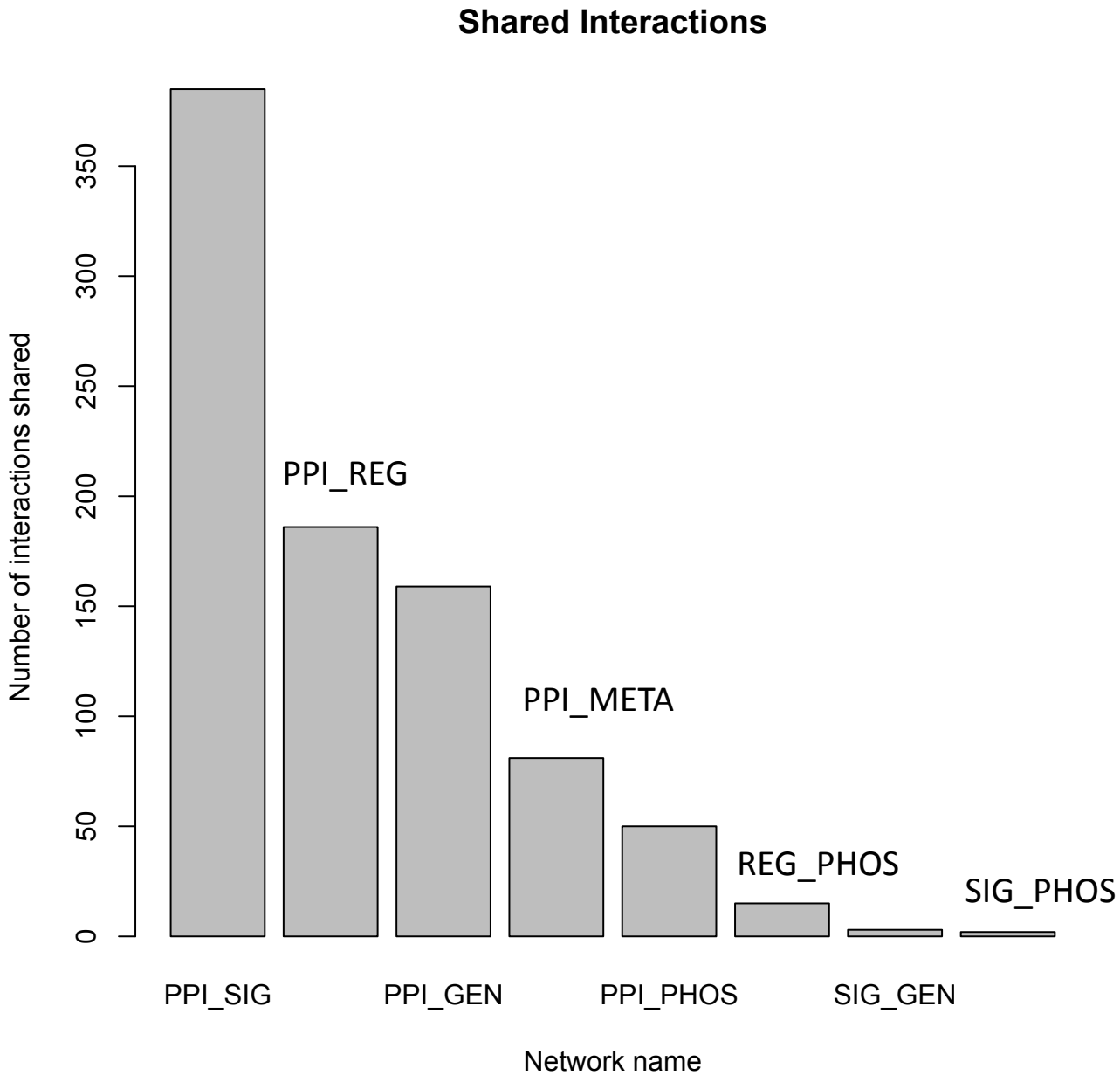
Presence of genes in different networks (total num of genes ~14,600)



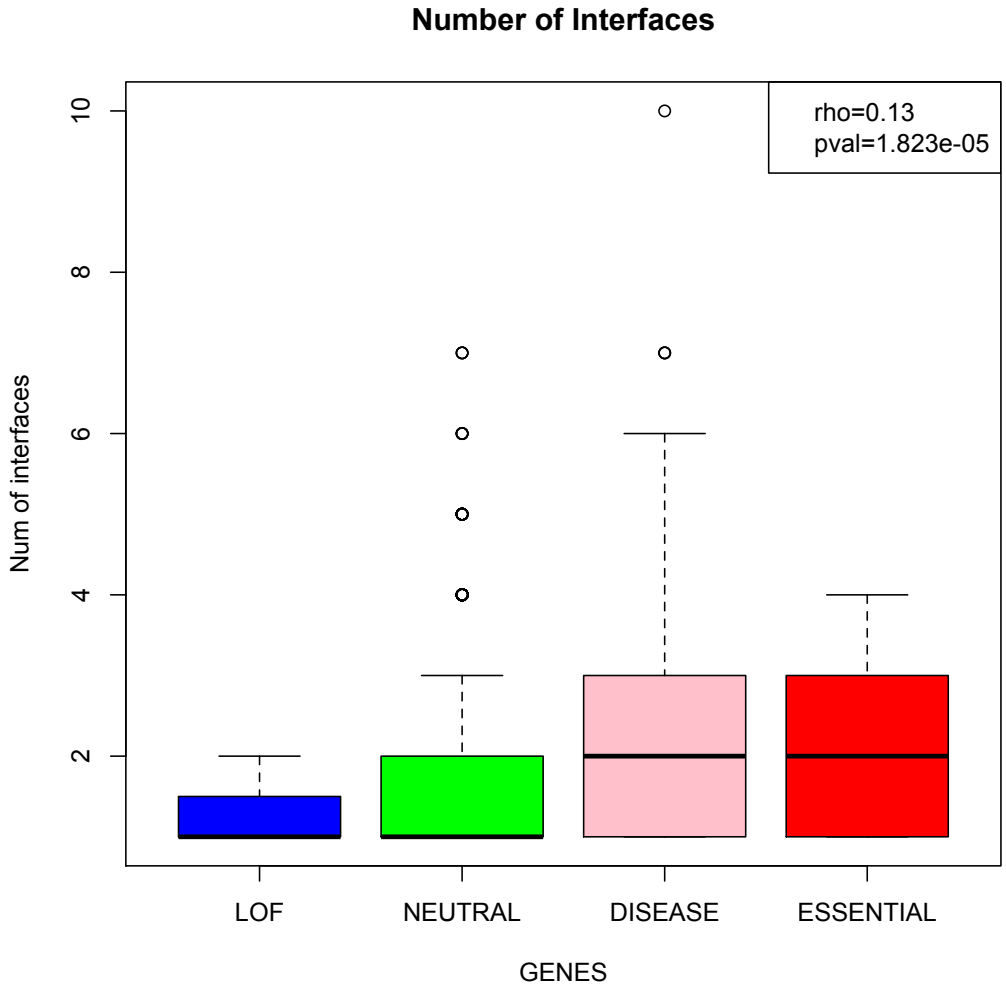
Presence of same interactions in different networks (total num of interactions ~11,000)

Only 881/11000 interactions present in more than one network

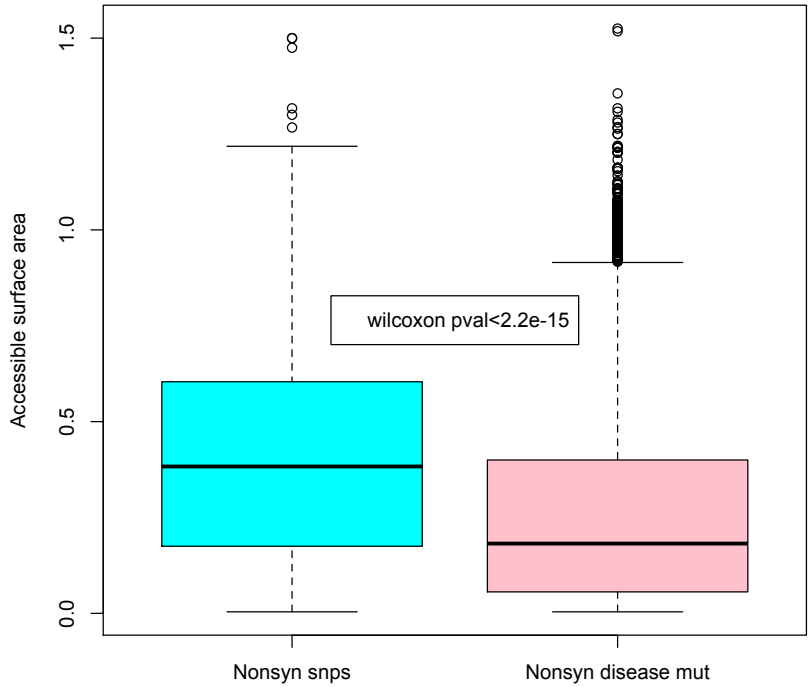



Shared Interactions
Number of interactions shared
0
50
100
150
200
250
300
350
PPI_REG
PPI_META
REG_PHOS
SIG_PHOS
PPI_SIG
PPI_GEN
PPI_PHOS
SIG_GEN
Network name

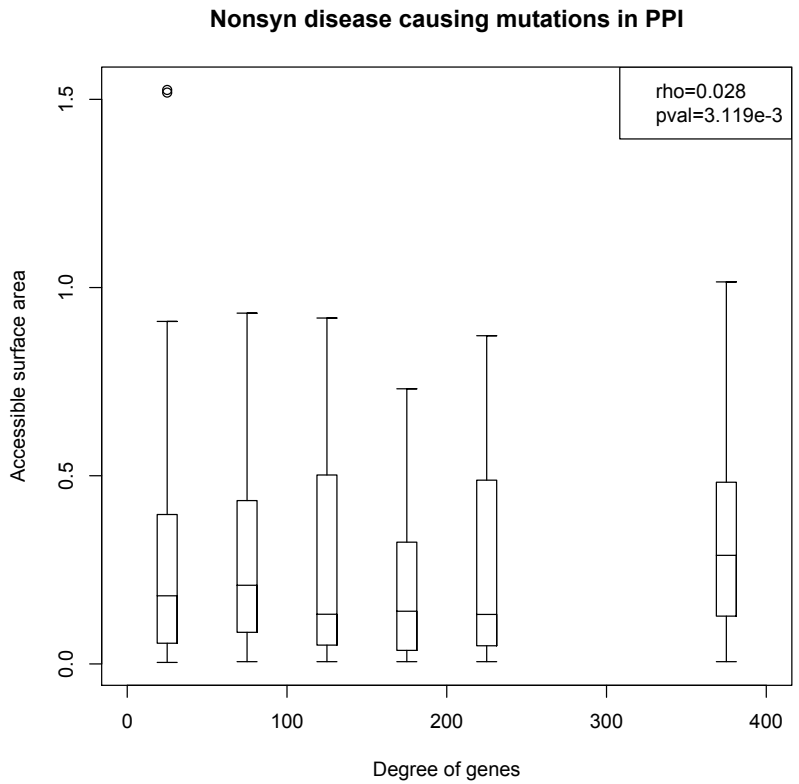
INTEGRATING STRUCTURE WITH PPI NETWORK



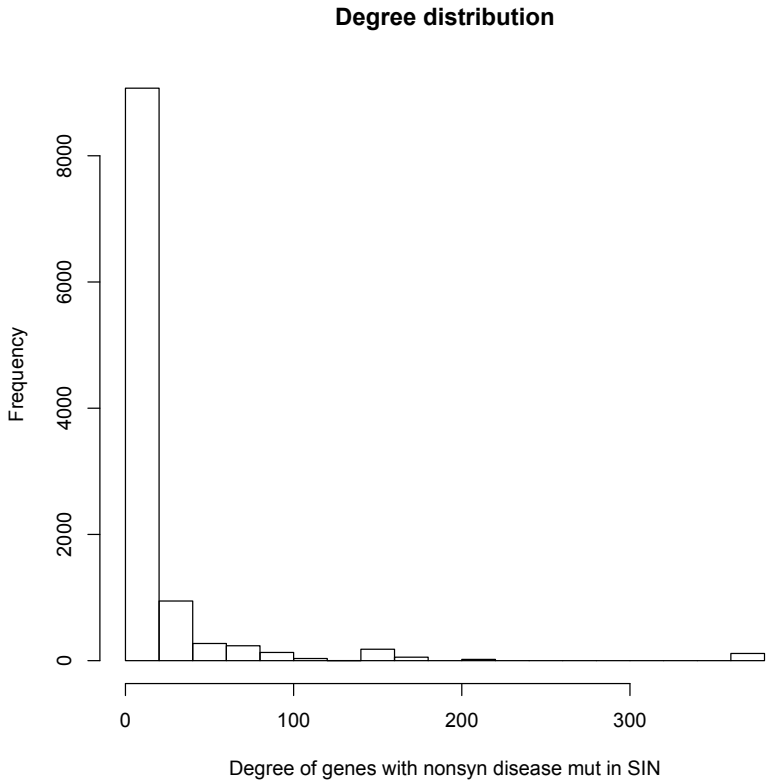
Number of interfaces in PPI network is correlated with susceptibility score of gene

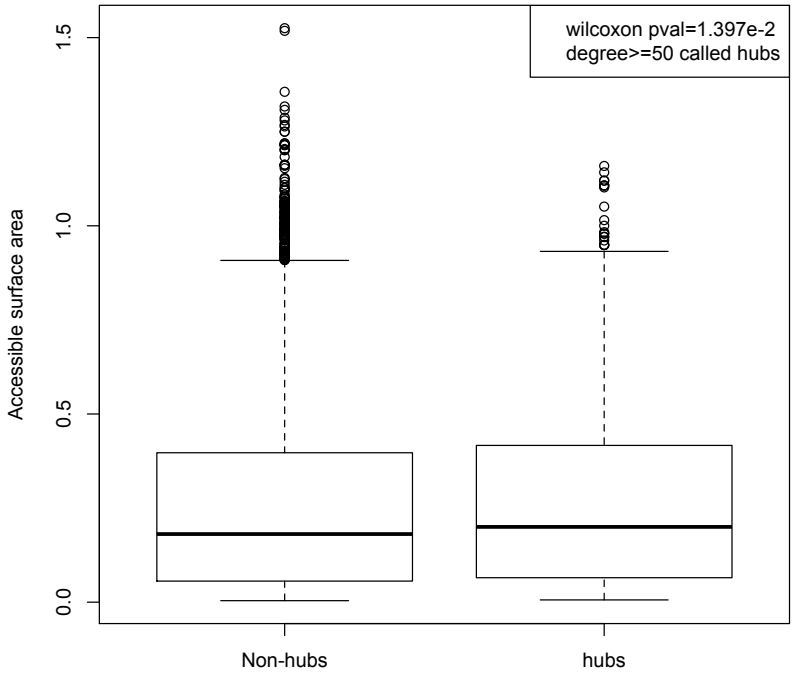


Disease-causing mutations tend to be more buried in general



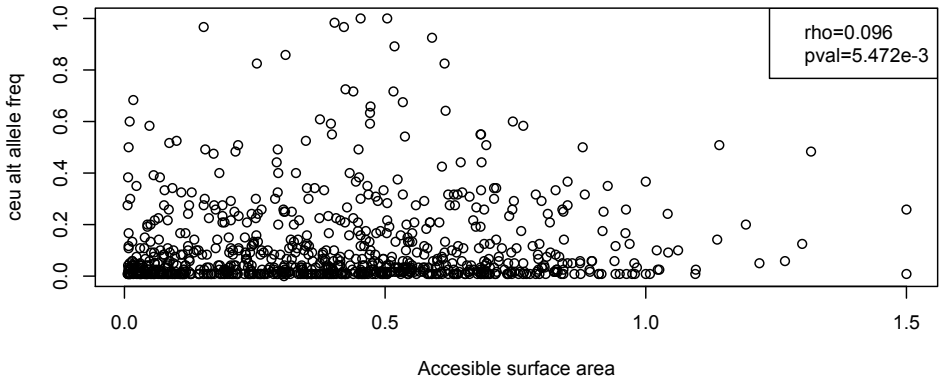
Hubs tend to have more disease mutations on surface than non-hubs

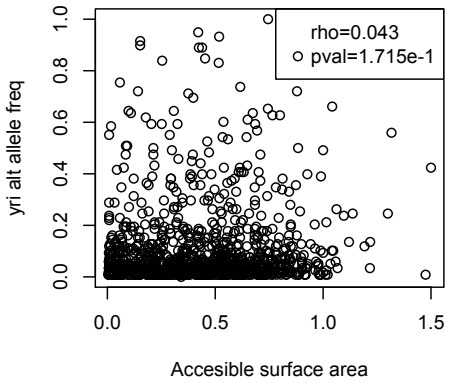


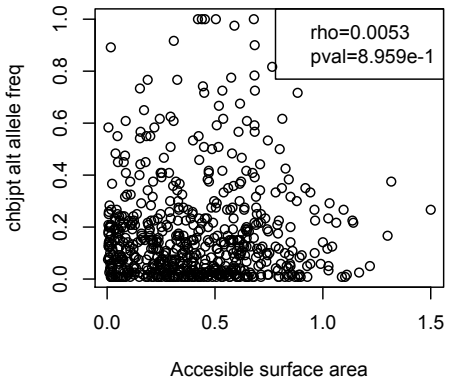


Hubs tend to have more disease mutations on surface than non-hubs

(better way to define hubs and non-hubs?)







Significant positive correlation between area and allele freq of nonsyn snps in CEU

Outline

1) Divide genes into four categories based on kinds of mutations they exhibit: lof-tolerant, neutral, essential, neutral. Their network centralities and presence in networks shows a trend.
2) When perform logistic regression on lof and essential genes: num of networks does almost as good as dn/ds and better than degrees in individual networks
3) Compare evolutionary sel. pressure for these categories betn. human-chimp (dn/ds) and within humans (allele freqs.) and that shows a trend.
4) Integrate structure with PPI network and observe that there are more constraints on surface of central proteins which is one reason leading to reduced tolerance to mutations
5) Data resource useful for functional interpretation of clinical variants

If combine with prosnp

1) Divide genes into four categories based on kinds of mutations they exhibit: lof-tolerant, neutral, essential, neutral. Their network centralities and presence in networks shows a trend.
2) Compare evolutionary sel. pressure for these categories betn. human-chimp (dn/ds) and within humans (allele freqs.) and that shows a trend.
3) Then zoom in to find which sites in genes show more mutations. dN/dS over vertebrates, structural conservation, seq. conservation for indels.
4) Integrate structure with PPI network and observe that there are more constraints on surface of central proteins which is one reason leading to reduced tolerance to mutations

- If keep prosnp separate put some examples of exceptions as reviewer suggested. Eg dN/dS is low but high allele freq in humans etc