Funseq and NCVARG

Lou Shaoke

Department of Molecular Biophysics and Biochemistry

loushaoke@gmail.com

August 11, 2016

Yale

Mutation list for K562

Step1: get a subset of whole genome funseq score

K562 ChromHMM regions: remove that extremely large region ($>$ 1Kb) and about 70K regions left. After filtering , about 10% genome position left.

Step2: Remove non-annotated positions

Remove the genomic position:
Not have any annotation
GERP scores$>$ 0, and
funseqScore$>$ 0.01

Example:

chr1 11416 11417 G A—C—T chr1-0 .;No;.;.;.;.;.;.;.;.;.;.;.;REPEAT(TAR1—chr1:10468-11447);.;0;.;.
chr1 11417 11418 C A—G—T chr1-0 .;No;.;.;.;.;.;.;.;.;.;.;.;REPEAT(TAR1—chr1:10468-11447);.;0;.;.
chr1 11486 11487 T A—C—G chr1-0 .;No;.;.;.;.;.;.;.;.;.;.;.;.;0;.;.
chr1 11487 11488 A C—G—T chr1-0 .;No;.;.;.;.;.;.;.;.;.;.;.;.;0;.;.
chr1 11488 11489 T A—C—G chr1-0 .;No;.;.;.;.;.;.;.;.;.;.;.;.;0;.;.

Step3: get distribution of overlapping with other type markers

K562 ChromHMM vs (DHS + GRO−Seq + P300) region overlapping;

#Overlapped Markers	number
0	42686
1	29204
2	619
3	145
4	26

Step4: randomly select enhancer region

Random selection of about 2000 enahancer regions, same distribution as the overlapped marker count

Step5: randomly select SNV according to low, mid, high funseq score quantile

Sort by funseq Score, get top 2, medium 2 and bottom 2 SNV randomly according to funseq Score (1911 enhancer regions with 11466 SNVs)