Mutation Analysis

Statistics Results

Generated on February 20, 2020 - 11:56

Result Directory
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\results\mutation_analysis

Genes File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\genes.tsv

Mutations File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\mutations.tsv

Genes Of Interest File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\genes_of_interest.tsv

Substitution Matrix File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\substitution_matrices\PAM10.tsv

Protein Domains File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\domains.tsv

Regulation File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\regulation.tsv

TFBS File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\TFBS.tsv

TF Motif MSA File
D:\OneDrive\Projects\MutaNET\MutaNET 2.0\example_data\mutation_analysis\artificial\TF_MSA.fasta

Gene Table

gene type	genes	TFBSs	TFs
MDR efflux pump	2 (9.1%)	2 (40.0%)	0 (0.0%)
MDR efflux regulator	1 (4.5%)	1 (20.0%)	1 (14.3%)
antibiotic	2 (9.1%)	1 (20.0%)	0 (0.0%)
total genes of interest	5 (22.7%)	3 (60.0%)	1 (14.3%)
other genes	17 (77.3%)	3 (60.0%)	6 (85.7%)
total	22	5	7

Description

Number of genes in the database, number of transcription factors and number of transcription factor binding sites for different types of genes.'Total genes of interest' is the sum of MDR efflux pump, MDR efflux regulator, antibiotic. 'Total' is the sum of 'total genes of interest' and 'other genes'.

Transcription factor binding sites can belong to an entire operon instead of a single gene, and an operon can contain both 'genes of interest' and 'other genes'. Thus, the total number of transcription factor binding sites can be lower than the sum of the TFBS column.

Mutation Table

mutations in	in coding	in promoter	in TFBSs	total
MDR efflux pump	5 (23.8%)	2 (11.1%)	0 (0.0%)	7 (17.9%)
MDR efflux regulator	3 (14.3%)	1 (5.6%)	1 (50.0%)	4 (10.3%)
antibiotic	4 (19.0%)	1 (5.6%)	1 (50.0%)	5 (12.8%)
total genes of interest	12 (57.1%)	3 (16.7%)	1 (50.0%)	15 (38.5%)
other genes	9 (42.9%)	16 (88.9%)	2 (100.0%)	25 (64.1%)
total	21	18	2	39

Description

Number of mutations in different types of genes (rows) and gene regions (columns). 'Total genes of interest' is the sum of MDR efflux pump, MDR efflux regulator, antibiotic. 'Total' is the sum of 'total genes of interest' and 'other genes'. TFBS stands for transcription factor binding site.

Gene regions can overlap, including with regions of other genes. As a result, some mutations are located in both 'genes of interest' and in 'other genes'. In such a case, only mutation information associated with 'genes of interest' is counted for the 'genes of interest' rows, same for mutation information associated with 'other genes' genes. The total numbers can be a bit smaller than the sum for that reason.

Genes Of Interest

MDR Efflux Pump

- lt16 (gn16)
- lt17 (gn16)

MDR Efflux Regulator

- lt6 (gn6)

Antibiotic

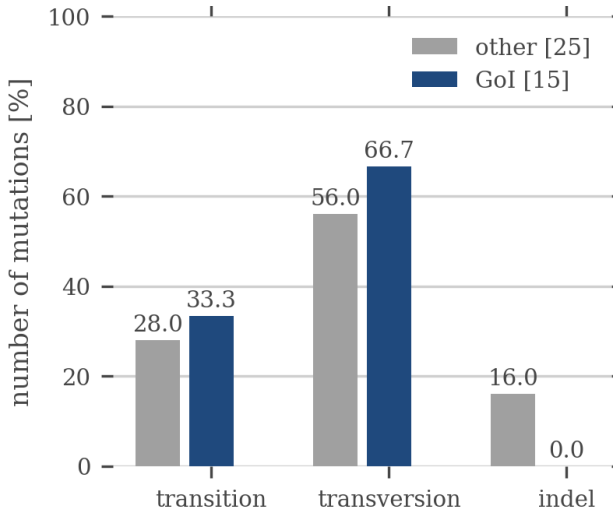
- lt26 (gn26)
- lt8

Mutation Distribution - Type

Method

Mutations were divided based on their type. Then, the number of mutations (in percent) in 'genes of interest' and 'other genes' was compared.

Plot



Description

The plot show the number of mutations (in percentage) in 'genes of interest' and 'other genes'. Brackets denote the number of mutations in the respective category of genes. '*' indicates a p-values < 0.05 and '**' a p-value < 0.01.

P-Values

The Wilcoxon rank-sum test was used to statistically compare the 'genes of interest' and 'other genes' mutation distributions. Note that the test requires a sample size > 20 to be reliable.

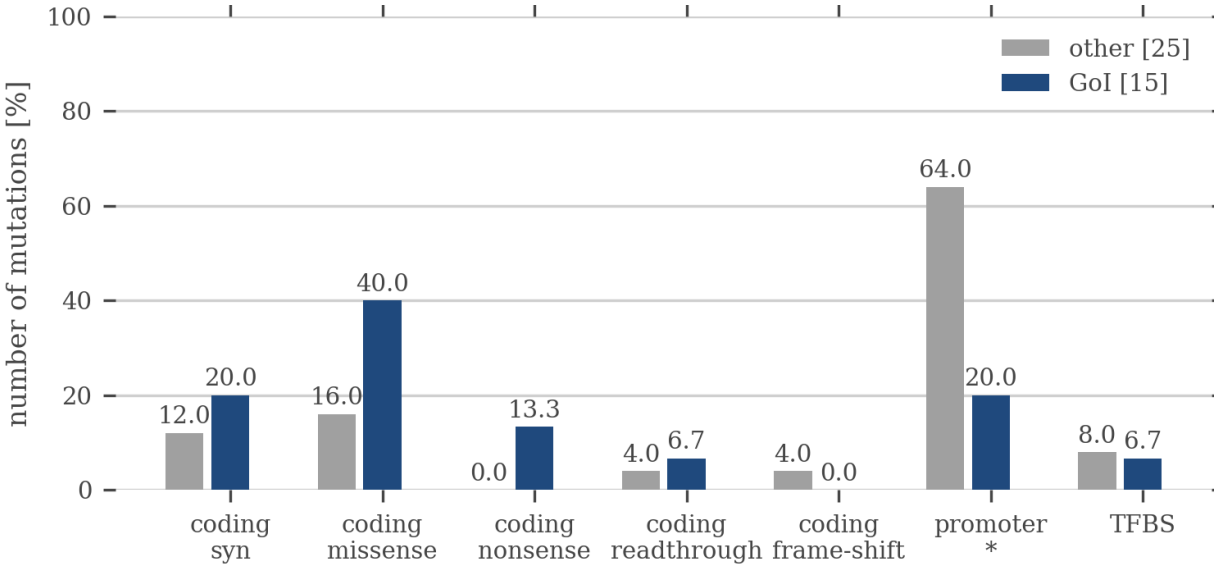
	p-value
transition	0.77995927
transversion	0.57633628
indel	0.40196536

Mutation Distribution - Gene Regions

Method

Mutations were divided based on their location in different gene regions. Mutations in coding regions were further divided into four groups depending on their effect on the amino acid sequence of the protein. 'Missense' means a single amino acid substitution, 'nonsense' the replacement of an amino acid by a translation stop, 'readthrough' a translation stop replaced by an amino acid and 'frame-shift' a reading frame shift caused by an insertion or deletion. Then, the number of mutations (in percent) in all 'genes of interest' and in 'other genes' was compared.

Plot



Description

The plot shows the number of mutations (in percentage) in 'genes of interest' and 'other genes'. Brackets denote the number of mutations in the respective category of genes. The percentages can add up to more than 100% due to an overlap of gene regions. '*' indicates a p-values < 0.05 and '**' a p-value < 0.01.

P-Values

The Wilcoxon rank-sum test was used to statistically compare the 'genes of interest' and 'other genes' mutation distributions. Note that the test requires a sample size > 20 to be reliable.

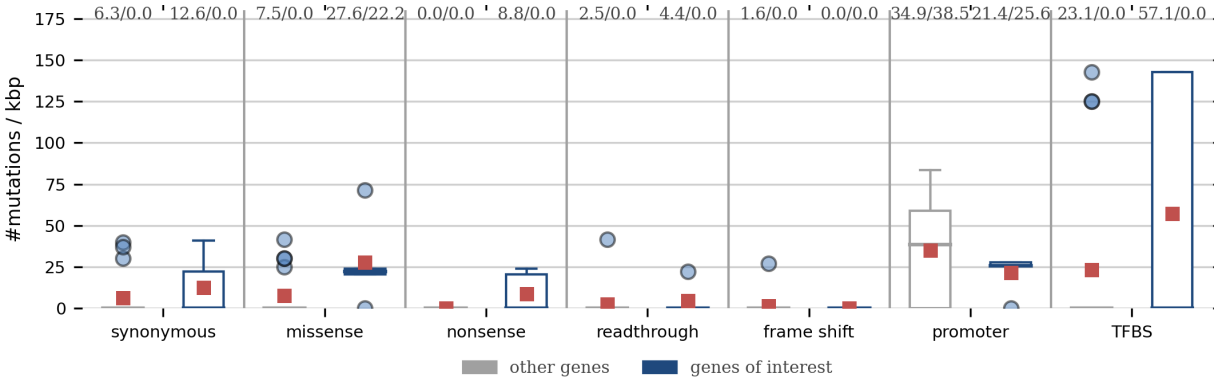
	p-value
coding: synonymous	0.67517361
coding: missense	0.20869044
coding: nonsense	0.48490814
coding: readthrough	0.88890804
coding: frame-shift	0.83403523
promoter	0.02117666
TFBS	0.94431860

Mutation Density

Method

For each gene, the mutation density (#mutations / kbp) for different gene regions was computed. Mutations in coding regions were divided into four groups depending on their effect on the amino acid sequence of the protein. 'Missense' means a single amino acid substitution, 'nonsense' the replacement of an amino acid by a translation stop, 'readthrough' a translation stop replaced by an amino acid and 'frame-shift' a reading frame shift caused by an insertion or deletion.

Plot



Description

The plot compares the mutation density between 'genes of interest' and 'other genes'. At the top are the mean/median values. The blue circles represent outliers, the red squares the mean, and the middle line the median. '*' indicates a p-values < 0.05 and '**' a p-value < 0.01.

P-Values

The Wilcoxon rank-sum test was used to statistically compare the mutation density distributions between 'genes of interest' and 'other genes'. Note that the test requires a sample size > 20 to be reliable.

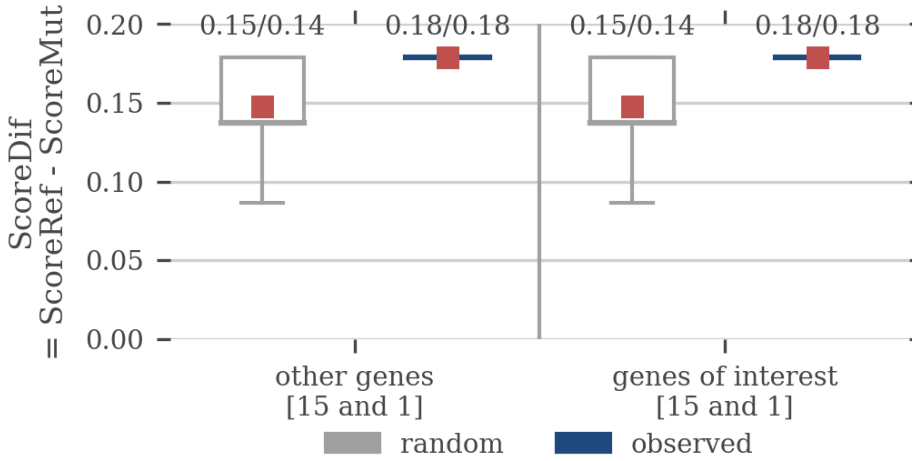
	p-value
coding: synonymous	0.45670244
coding: missense	0.11713304
coding: nonsense	0.18289855
coding: readthrough	0.66653860
coding: frame-shift	0.84471569
promoter	0.25594900
TFBS	0.36759903

Transcription Factor Binding Site Analysis

Method

Mutations in transcription factor binding sites were scored with the position weight matrix of the corresponding transcription factor to ascertain if the mutation likely increases or decreases the ability of the transcription factor to bind to the mutated binding site. Additionally, for each mutation, 15 mutations were randomly generated in the binding site, such that the type of the mutation was conserved. I.e., if the observed mutation is a transition, then the randomly generated mutations are also transitions. These random mutations were also scored with the position weight matrix.

Plot



Description

The scores of observed and random mutations were then normalised with the minimum and maximum score possible for the transcription factor binding site, resulting in a score between 0 and 1. A normalised score of 0 means that the sequence differs as much as possible from the transcription factor motif, whereas a normalised score of 1 means that the sequence is as similar as possible to the motif.

P-Values

The final score of observed and random mutations was computed as the difference from the score of the reference binding site sequence (score = score_ref - score_mut). Scores > 0 suggest that a mutation decreases the ability of the transcription factor to bind to the mutated binding site, whereas scores < 0 suggest an increase. The more the score deviates from 0, he higher the difference between the reference and mutated binding site sequence and thus the higher he potential impact of the mutation.

	p-value
observed vs random (in '{0}')	0.32897177
observed vs random (in '{0}')	0.32897177
observed vs observed	1.00000000