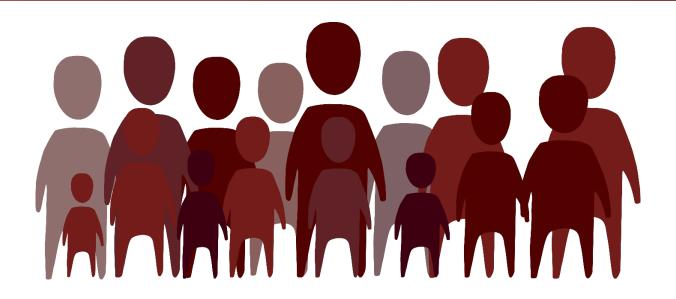# Practical

MutationTaster & RegulationSpotter

Daniela Hombach & Jana Marie Schwarz

Exzellenzcluster NeuroCure
Charité Universitätsmedizin Berlin

AG Translational Genomics

08.11.2017

Rare diseases
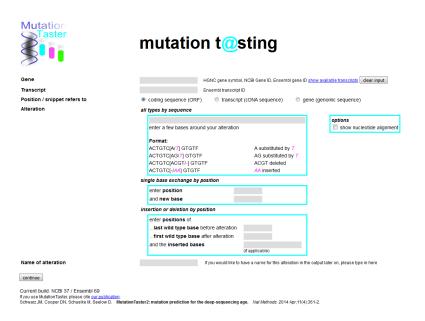

7000 – 8000 rare diseases are known.

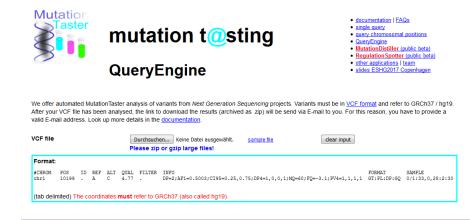>> World-wide more than **300 Millions** patients

Rare diseases

	Mendelian	**Complex**
# loci	one	multiple (polygenic)
effect of environment	small	high
frequency in population	rare	common
effect	strong	weak
example	cystic fibrosis, hemophilia, PKU	allergy, cancer, diabetes

MutationTaster

Single variant query

VCF query

highly-parallel fashion (500,000 variants / hour)

InDels, SNVs, introns, exons, coding and non-coding variants

Go to: www.mutationtaster.org

8 ▾ Google

Mutation Taster

mutation t@sting

- *NEWS*
- documentation | FAQs
- single query
- query chromosomal positions
- QueryEngine
- other applications | team

Gene   [SOD1]   HGNC gene symbol, NCBI Gene ID, Ensembl gene ID show available transcripts   [clear input]

Transcript   [ENST00000270142]   Ensembl transcript ID
Choose the transcript:
- ○ ENST00000470944 (*processed_transcript*, 1746 base
- ⦿ ENST00000270142 (*protein_coding*, 966 bases)   NM_000454
- ○ ENST00000389995 (*protein_coding*, 865 bases)
- ○ ENST00000476106 (*processed_transcript*, 586 bases)

SOD1

Position / snippet refers to   ⦿ coding sequence (ORF)   ○ transcript (cDNA sequence)   ○ gene (genomic sequence)

Alteration   *all types by sequence*

C[A/G]TGTTCATGAGTTTGGAGATAATACAGCAGGCTGT

enter a few bases around your alteration

Format:
ACTGTC[A/*T*] GTGTF	A substituted by *T*
ACTGTC[AG/*T*] GTGTF	AG substituted by *T*
ACTGTC[ACGT/-] GTGTF	ACGT deleted
ACTGTC[-/*AA*] GTGTF	*AA* inserted

options
☐ show nucleotide alignment

single base exchange by position

enter **position**   [   ]
and **new base**   [   ]

insertion or deletion by position

enter **positions** of
...**last wild type base** before alteration   [   ]
...**first wild type base** after alteration   [   ]
and the **inserted bases**   [   ]
(if applicable)

Name of alteration   [SOD1_ALS]   if you would like to have a name for this alteration in the output later on, please type in here

[continue]

Mutation Taster

mutation t@sting

Alteration SOD1_ALS

Prediction	disease causing	Model: *simple_aae*, prob: 0.999999999908447 (classification due to ClinVar, real probability is shown anyway) (explain)

Summary

- amino acid sequence changed
- known disease mutation at this position (HGMD CM930680)
- known disease mutation: rs121912443 (pathogenic)
- protein features (might be) affected
- splice site changes

analysed issue	analysis result
name of alteration	SOD1_ALS
alteration (phys. location)	chr21:33036170A>G
HGNC symbol	SOD1
Ensembl transcript ID	ENST00000270142
UniProt peptide	P00441
alteration type	single base exchange
alteration region	CDS
DNA changes	c.140A>G cDNA.288A>G g.4236A>G
AA changes	H47R Score: 0.79 explain score(s)
position(s) of altered AA if AA alteration in CDS	47
frameshift	no
dbSNP / TGP / HGMD(public) / ClinVar	**known disease mutation: rs121912443 (pathogenic for *Amyotrophic lateral sclerosis type 1*)** dbSNP NCBI variation viewer known disease mutation at this position, please check HGMD for details (HGMD ID CM930680)
regulatory features	H3K9me1, Histone, Histone 3 Lysine 9 mono-methylation H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation H3K79me2, Histone, Histone 3 Lysine 79 di-methylation H4K20me1, Histone, Histone 4 Lysine 20 mono-methylation
phyloP / phastCons	5.162, 1 (flanking) 4.283, 0.996

Computer based classification

- „either-or" decision

- automated classification by **Bayes classifier**

- **simple handling**
- **CPU/RAM friendly**
- **Fast & robust results**

Bayes classifier: Spam E-Mails

new message | important

sign up and win! Every click wins!

The best exclusive exciting HD videos online.

Shop now – cheap and easy.

Payments...

Our "Low Price Pharmacy Store" design sports a professional array of pharmaceuticals.

msg me with a valid email for an account

Bayes classifier: Spam E-Mails

<u>Subject:</u> new message | important

sign up and **win**! Every click **wins!**

The **best exclusive** exciting HD videos online.

Shop now – cheap and easy.

Payments...

Our "Low Price **Pharmacy** Store" design sports a professional array of pharmaceuticals.

msg me with a valid email for an account

Recognition by Bayes classifier

term / phrase	spam	ham
shop now	++	-
pharmacy	++	o
win	++	+
account	o	+
MutationTaster	-	++
abstract	-	++

Bayes classifier in MutationTaster

attribute	disease mutation	polymorphism
loss of splice site	49.5 %	0.06 %
amino acid exchange transmembrane domain	7.3 %	4.5 %
loss of disulfide bridge	2.9 %	0.1 %
amino acid exchange Arg>His	2.3 %	2.4 %

Bayes classifier in MutationTaster

- training with

  a) > 100,000 known disease mutations
  Human Gene Mutation Database (HGMD)

  b) > 6,000,000 harmless polymorphisms
  1000 Genomes Project (1000G)

- classification model:
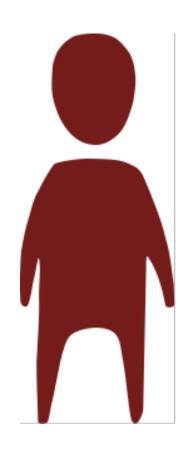 stores frequencies of of different values / test results

VCF upload

mutation t@sting

QueryEngine

We offer automated MutationTaster analysis of variants from *Next Generation Sequencing* projects. Variants must be in VCF format and refer to GRCh37 / hg19. After your VCF file has been analysed, the link to download the results (archived as .zip) will be send via E-mail to you. For this reason, you have to provide a valid E-mail address. Look up more details in the documentation.

VCF file	Durchsuchen... Keine Datei ausgewählt. sample file	clear input
	Please zip or gzip large files!	

Format:

```
#CHROM  POS    ID  REF  ALT  QUAL  FILTER  INFO                                                    FORMAT     SAMPLE
chr1    10199  .   A    C    4.77  .       DP=2;AF1=0.5003;CI95=0.25,0.75;DP4=1,0,0,1;MQ=60;FQ=-3.1;PV4=1,1,1,1  GT:PL:DP:GQ  0/1:33,0,28:2:30
```

(tab delimited) The coordinates **must** refer to GRCh37 (also called hg19).

Project name   Practical_Prague

queue status   2017-11-01 17:57 CEST   (refresh)
low load - our bored server is looking forward to analysing your data
0 jobs running, 0 queued, 1515.6 millions alterations analysed.

DB queries	free slots (0 running, 16 of 16 available - at 0% capacity)
large jobs	free slots (0 running, 20 of 20 available - at 0% capacity)
small jobs	free slots (0 running, 30 of 30 available - at 0% capacity)
medium jobs	free slots (0 running, 25 of 25 available - at 0% capacity)

Unfortunately, E-Mail notification does currently not work. We are sorry for the inconvenience and try to solve the problem as soon as possible

E-mail address

generate HTML files (slower) ☐

Analysis settings

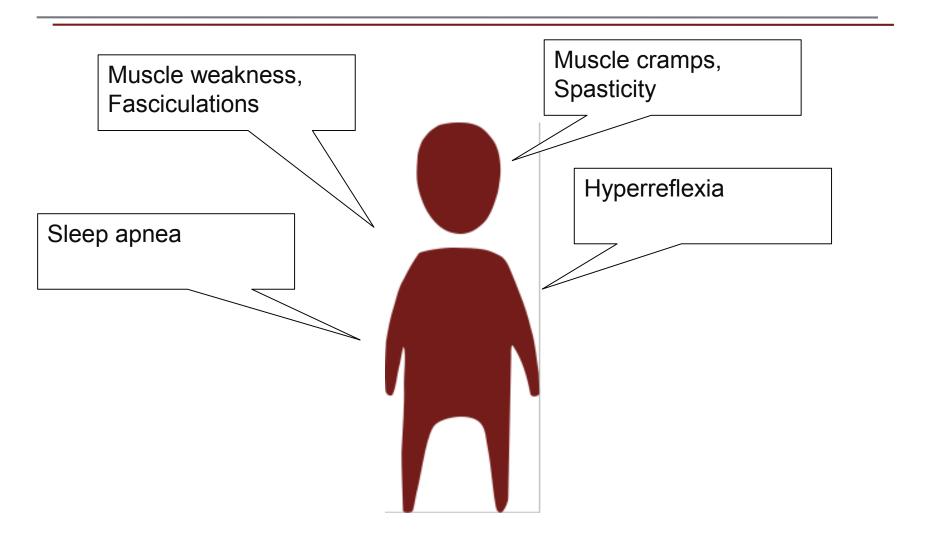| search for homozygous variants	☐ yes	⦿ analyse complete VCF
		☐ ...but only exons with 10 bases intron flanking
combine neighbouring variants	☐ yes	◯ analyse variants on chr
		☐ ...but only exons with 10 bases intron flanking
filter against TGP	homozygous in 4 or more TGP samples ☑	◯ analyse custom regions (select to enter)
		◯ exclude custom regions (select to enter)
	heterozygous in 20 or more TGP samples ☐	
minimum coverage	4	

Submit

Example case: Whole Exome Sequencing

http://bit.ly/mt_exome

Example case: Phenotype

Muscle weakness, Fasciculations

Muscle cramps, Spasticity

Hyperreflexia

Sleep apnea

http://bit.ly/mt_exome

Example case: Phenotype

Muscle weakness, Fasciculations

Muscle cramps, Spasticity

Hyperreflexia

Sleep apnea

Amyotrophic lateral sclerosis (ALS)
OMIM #105400
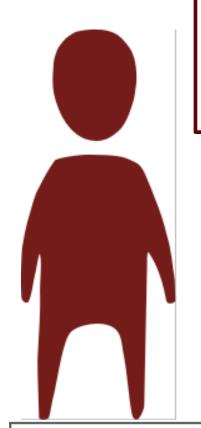
http://bit.ly/mt_exome

Example case: WholeExome Sequencing

Exome from 1000G (HG00377)

with known disease mutation from ClinVar

http://bit.ly/mt_exome

Amyotrophic lateral sclerosis (ALS)
OMIM #105400

VCF upload

Analysis Settings

Results

Results: Overview

Mutation Taster

MT results Practical_Prague

due to the recursion, limit/offset do not have any effect !
304 subgroups

MutationTaster predictions were generated using data from Ensembl69

chromosome	position	genesymbol	prediction	model	probability	type	AAE	dbSNP	ref	alt	ClinVar	features	coverage	gt/freq(var)	filename
21	33036170	SOD1	disease_causing_automatic	simple_aae	1	single base exchange	H47R	121912443	A	G	known disease mutation	splicing affected; protein features affected	170	hom	914
21	33036170	SOD1	disease_causing_automatic	simple_aae	1	single base exchange	H28R	121912443	A	G	known disease mutation	splicing affected; protein features affected	170	hom	915
10	21435341	C10orf113	disease_causing_automatic	complex_aae	1	deletion and insertion	S32Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom	489
10	21435341	C10orf113	disease_causing_automatic	complex_aae	0.999985	deletion and insertion	S32Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom	490
10	21435341	C10orf113	disease_causing_automatic	complex_aae	1	deletion and insertion	S22Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom	491
12	11420334	PRB3	disease_causing_automatic	complex_aae	1	deletion and insertion	P241Sfs*14		GGG	GG		frameshift; splicing affected; protein features affected; NMD	196	hom	580
12	11420334	PRB3	disease_causing_automatic	complex_aae	1	deletion and insertion	P241Sfs*14		GGG	GG		frameshift; splicing affected; protein features affected; NMD	196	hom	581
12	11421094	PRB3	polymorphism	without_aae	1	deletion and insertion			GG	GGGGG		splicing affected; protein features affected	43	hom	582
12	11421094	PRB3	polymorphism	without_aae	1	deletion and insertion			GG	GGGGG		splicing affected; protein features affected	43	hom	583
19	52004792	SIGLEC12	disease_causing_automatic	complex_aae	1	deletion and insertion	A66Gfs*50		CCC	CCCC		frameshift; splicing affected; protein features affected; NMD	44	hom	885
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom	208
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom	209
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom	210
3	64527465	ADAMTS9	polymorphism_automatic	without_aae	1	single base exchange		1036917	C	T		splicing affected; protein features affected	59	hom	206

Results: Overview

Results: Overview

MutationTaster MT results Practical_Prague

due to the recursion, limit/offset do not have any effect !
304 subgroups

back to overview

MutationTaster predicti... s were generat... using data from Ensembl69

chromosome	position	genesymbol	prediction	model	probability	type	AAE	dbSNP	ref	alt	ClinVar	features	coverage	gt/freq(var)	filename
21	330361 0	SOD1	disease_causing_automatic	simple_aae	1	single base exchange	H47R	121912443	A	G	known disease mutation	splicing affected; protein features affected	170	hom	914
21	33036170	SOD1	disease_causing_automatic	simple_aae	1	single base exchange	H28R	121912443	A	G	known disease mutation	splicing affected; protein features affected	170	hom	915
10	21435341	C10orf113	disease_causing_automatic	complex_aae	1	deletion and insertion	S32Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom	489
10	21435341	C10orf113	disease_causing_automatic	complex_aae	0.999985	deletion and insertion	S32Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom	490
10	21435341	C10orf113	disease_causing_automatic	complex_aae	1	deletion and insertion	S22Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom	491
12	11420334	PRB3	disease_causing_automatic	complex_aae	1	deletion and insertion	P241Sfs*14		GGG	GG		frameshift; splicing affected; protein features affected; NMD	196	hom	580
12	11420334	PRB3	disease_causing_automatic	complex_aae	1	deletion and insertion	P241Sfs*14		GGG	GG		frameshift; splicing affected; protein features affected; NMD	196	hom	581
12	11421094	PRB3	polymorphism	without_aae	1	deletion and insertion			GG	GGGGG		splicing affected; protein features affected	43	hom	582
12	11421094	PRB3	polymorphism	without_aae	1	deletion and insertion			GG	GGGGG		splicing affected; protein features affected	43	hom	583
19	52004792	SIGLEC12	disease_causing_automatic	complex_aae	1	deletion and insertion	A66Gfs*50		CCC	CCCC		frameshift; splicing affected; protein features affected; NMD	44	hom	885
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom	208
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom	209
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom	210
3	64527465	ADAMTS9	polymorphism_automatic	without_aae	1	single base exchange		1036917	C	T		splicing affected; protein features affected	59	hom	206

Results: GeneDistiller

Candidate genes. *1 gene(s)*

genesymbol	type	description	chr.	startpos	endpos	synonyms
SOD1	protein-coding	**superoxide dismutase 1, soluble**	21	33031935	33041244	ALS1, IPOA, SOD, homodimer, ALS, hSod1, HEL-S-44
	links	NCBI ENSEMBL SwissProt GeneCards STRING UniHI PubMed create primers for all transcripts				
	KEGG pathways	Peroxisome, Amyotrophic lateral sclerosis (ALS), Huntington's disease, Prion diseases				
	Reactome pathways	Platelet activation, signaling and aggregation, Response to elevated platelet cytosolic Ca2+, Hemostasis, Platelet degranulation				
	PFAM	sodcu;				
	InterPro domains	Superoxide dismutase, copper/zinc binding domain				
	paralogs	SOD3 (24%), CCS (26%)				

HPO
- Muscle cramps
- Pseudobulbar paralysis
- Degeneration of anterior horn cells
- Amyotrophic lateral sclerosis
- Sleep apnea
- Fasciculations
- Autosomal dominant inheritance
- Autosomal recessive inheritance
- Spasticity
- Degeneration of the lateral corticospinal tracts
- Hyperreflexia
- Skeletal muscle atrophy
- Muscle weakness
- Heterogeneous

OMIM
*147450 SUPEROXIDE DISMUTASE 1; SOD1
SUPEROXIDE DISMUTASE, CYTOSOLIC
SUPEROXIDE DISMUTASE, SOLUBLE
SOD, SOLUBLE
SUPEROXIDE DISMUTASE, COPPER-ZINC
INDOPHENOL OXIDASE A; IPOA
text:
DESCRIPTION: The SOD1 gene encodes superoxide dismutase-1 (EC 1.15.1.1), a major cytoplasmic antioxidant enzyme that metabolizes superoxide radicals to molecular oxygen and hydrogen peroxide, thus providing a defense against oxygen toxicity (Niwa et al., 2007). Soluble cytoplasmic SOD1 is a copper- and zinc-containing enzyme; the SOD1 gene maps to chromosome 21q22 (Sherman et al., 1983). SOD2 (147460) is a distinct mitochondrial enzyme that contains manganese; the SOD2 gene maps to 6q25. SOD1 is a homodimer and SOD2 a tetramer (Beckman et al., 1973). [...]
See report at OMIM's website.

OrphaNet   Amyotrophic lateral sclerosis   germline (assessed)

generifs
- That hSOD1 overexpression improves sensorimotor coordination in young organisms may suggest possible therapeutic strategies for early-onset ALS in humans.
- astrocyte-derived exosomes efficiently transfer mutant SOD1 to spinal neurons and induce selective motor neuron death
- ATPase-deficient Hsp104 mutants did not restore mobility, suggesting that, rather than preventing aggregation, Hsp104 disaggregates mutant SOD1 after it has aggregated
- The results of thi study shows for the first time that the functional consequence of G93A hSOD1 overexpression in intact motor neurons is indeed a disturbance of the ER mitochondria calcium cycle
- The A4V mutation as well as the absence of GSH, reduced hSod1 activity.
- In conclusion, increased catecholamine production in PHEO is accompanied by decreased levels of vitamin C and increased levels of MDA which may indicate the activation of oxidative stress in PHEO.
- This study demonistrated that in mouse in mouse models of sod1 mutation amyotrophic lateral sclerosis Enhancing mitochondrial calcium buffering capacity reduces aggregation of misfolded SOD1 and motor neuron cell death without extending survival

http://www.genedistiller.org

Results: Overview

MutationTaster

MT results Practical_Prague

due to the recursion, limit/offset do not have any effect !
304 subgroups

back to overview

MutationTaster predictions were generated using data from Ensembl69

chromosome	position	genesymbol	prediction	model	probability	type	AAE	dbSNP	ref	alt	ClinVar	features	coverage	gt/fre		ihename
21	33036170	SOD1	disease_causing_automatic	simple_aae	1	single base exchange	H47R	121912443	A	G	known disease mutation	splicing affected; protein features affected	170	hom		914
21	33036170	SOD1	disease_causing_automatic	simple_aae	1	single base exchange	H28R	121912443	A	G	known disease mutation	splicing affected; protein features affected	170	hom		915
10	21435341	C10orf113	disease_causing_automatic	complex_aae	1	deletion and insertion	S32Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom		489
10	21435341	C10orf113	disease_causing_automatic	complex_aae	0.999985	deletion and insertion	S32Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom		490
10	21435341	C10orf113	disease_causing_automatic	complex_aae	1	deletion and insertion	S22Ffs*5		CACTC	C		frameshift; splicing affected; NMD	92	hom		491
12	11420334	PRB3	disease_causing_automatic	complex_aae	1	deletion and insertion	P241Sfs*14		GGG	GG		frameshift; splicing affected; protein features affected; NMD	196	hom		580
12	11420334	PRB3	disease_causing_automatic	complex_aae	1	deletion and insertion	P241Sfs*14		GGG	GG		frameshift; splicing affected; protein features affected; NMD	196	hom		581
12	11421094	PRB3	polymorphism	without_aae	1	deletion and insertion			GG	GGGGG		splicing affected; protein features affected	43	hom		582
12	11421094	PRB3	polymorphism	without_aae	1	deletion and insertion			GG	GGGGG		splicing affected; protein features affected	43	hom		583
19	52004792	SIGLEC12	disease_causing_automatic	complex_aae	1	deletion and insertion	A66Gfs*50		CCC	CCCC		frameshift; splicing affected; protein features affected; NMD	44	hom		885
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom		208
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom		209
3	64640207	ADAMTS9	disease_causing	without_aae	0.999883	deletion and insertion			AAAA	AAA		splicing affected; protein features affected	55	hom		210
3	64527465	ADAMTS9	polymorphism_automatic	without_aae	1	single base exchange		1036917	C	T		splicing affected; protein features affected	59	hom		206

Results: Details

Prediction **disease causing** Model: *simple_aae*, prob: 0.999999999908447 (classification due to ClinVar, real probability is shown anyway) (explain)

Summary

hyperlink

- **amino acid sequence changed**
- **known disease mutation at this position (HGMD CM930680)**
- **known disease mutation: rs121912443 (pathogenic)**
- **protein features (might be) affected**
- **splice site changes**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr21:33036170A>G show variant in all transcripts IGV
HGNC symbol	SOD1
Ensembl transcript ID	ENST00000270142
Genbank transcript ID	NM_000454
UniProt peptide	P00441
alteration type	single base exchange
alteration region	CDS
DNA changes	c.140A>G cDNA.288A>G g.4236A>G
AA changes	H47R Score: 29 explain score(s)
position(s) of altered AA if AA alteration in CDS	47
frameshift	no
known variant	Reference ID: rs121912443 Allele 'G' was neither found in ExAC nor 1000G. **known disease mutation: rs121912443 (pathogenic for *Amyotrophic_lateral_sclerosis_type_1*) dbSNP NCBI variation viewer** known disease mutation at this position, please check HGMD for details (HGMD ID CM930680)
regulatory features	H3K9me1, Histone, Histone 3 Lysine 9 mono-methylation H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation H3K79me2, Histone, Histone 3 Lysine 79 di-methylation H4K20me1, Histone, Histone 4 Lysine 20 mono-methylation
phyloP / phastCons	

	PhyloP	PhastCons
(flanking)	5.162	1
	4.283	0.996
(flanking)	-2.023	0.059

explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

Results: Details

Kozak consensus sequence altered?	N/A

conservation				
protein level for non-synonymous changes	species	match	gene	aa alignment
	Human			47 I K G L T E G L H G F H V H E F G D N T A G C T
	mutated	not conserved		47 I K G L T E G L H G F R V H E F G D N T A G C
	Ptroglodytes	all identical	ENSPTRG00000013847	47 I K G L T E G L H G F H V H E F G D N T A G C
	Mmulatta	all identical	ENSMMUG00000001711	47 I T G L T E G L H G F H V H Q F G D N T Q G C
	Fcatus	no homologue		
	Mmusculus	all identical	ENSMUSG00000022982	47 I T G L T E G Q H G F H V H Q Y G D N T Q G C
	Ggallus	all identical	ENSGALG00000015844	48 I T G L S D G D H G F H V H E F G D N T N G C
	Trubripes	all identical	ENSTRUG00000008179	69 I K G L T P G E H G F H V H A F G D N T N G C
	Drerio	all identical	ENSDARG00000043848	47 I T G L T P G K H G F H V H A F G D N T N G C
	Dmelanogaster	all identical	FBgn0003462	47 V C G L A K G L H G F H V H E F G D N T N G C
	Celegans	no homologue		
	Xtropicalis	all identical	ENSXETG00000007350	48 I Y G L T D G K H G F H I H E F G D N T N G C

protein features	start (aa)	end (aa)	feature	details	
	41	50	STRAND		lost
	47	47	METAL	Copper; catalytic.	lost
	49	49	METAL	Copper; catalytic.	might get lost (downstream of altered splice site)
	51	52	MUTAGEN	FG->EE: Abolishes dimerization; when associated with Q-134.	might get lost (downstream of altered splice site)
	54	56	HELIX		might get lost (downstream of altered splice site)
	57	61	HELIX		might get lost (downstream of altered splice site)
	58	58	MUTAGEN	C->S: Enhances formation of fibrillar aggregates in the absence of bound zinc; when associated with S-7; S-112 and S- 147.	might get lost (downstream of altered splice site)
	58	58	DISULFID		might get lost (downstream of altered splice site)
	63	65	STRAND		might get lost (downstream of altered splice site)
	64	64	METAL	Copper; catalytic.	might get lost (downstream of altered splice site)
	64	64	METAL	Zinc; via pros nitrogen.	might get lost (downstream of altered splice site)
	67	69	STRAND		might get lost (downstream of altered splice site)
	71	71	MOD_RES	N6-acetyllysine (By similarity).	might get lost (downstream of altered splice site)
	72	72	METAL	Zinc; via pros nitrogen.	might get lost (downstream of altered splice site)
	74	76	STRAND		might get lost (downstream of altered splice site)
	77	79	STRAND		might get lost (downstream of altered splice site)
	81	81	METAL	Zinc; via pros nitrogen.	might get lost (downstream of altered splice site)
	81	81	MUTAGEN	H->A: Loss of zinc binding and enhanced tendency to form aggregates; when associated with A-84.	might get lost (downstream of altered splice site)
	81	81	MUTAGEN	H->S: Destabilization of dimer and loss of zinc binding; when associated with S- 84.	might get lost (downstream of altered splice site)
	84	84	METAL	Zinc; structural.	might get lost (downstream of altered splice site)
	84	84	MUTAGEN	D->S: Destabilization of dimer and loss of zinc binding; when associated with S- 81.	might get lost (downstream of altered splice site)
	84	84	MUTAGEN	D->A: Loss of zinc binding and enhanced tendency to form aggregates; when associated with A-81.	might get lost (downstream of altered splice site)
	84	90	STRAND		might get lost (downstream of altered splice site)
	92	94	HELIX		might get lost (downstream of altered splice site)
	96	108	STRAND		might get lost (downstream of altered splice site)
	99	99	CONFLICT	S -> V (in Ref. 3; no nucleotide entry).	might get lost (downstream of altered splice site)
	99	99	MOD_RES	Phosphoserine.	might get lost (downstream of altered splice site)
	109	111	HELIX		might get lost (downstream of altered splice site)
	112	112	MUTAGEN	C->S: Enhances formation of fibrillar aggregates in the absence of bound zinc; when associated with S-7; S-58 and S-147.	might get lost (downstream of altered splice site)
	112	115	STRAND		might get lost (downstream of altered splice site)
	116	123	STRAND		might get lost (downstream of altered splice site)
	121	121	METAL	Copper; catalytic.	might get lost (downstream of altered splice site)
	123	123	MOD_RES	N6-acetyllysine	might get lost (downstream of altered splice site)

Results: Details

121	121	METAL	Copper; catalytic.			might get lost (downstream of altered splice site)
123	123	MOD_RES N6-acetyllysine.				might get lost (downstream of altered splice site)
127	129	STRAND				might get lost (downstream of altered splice site)
130	132	STRAND				might get lost (downstream of altered splice site)
133	136	HELIX				might get lost (downstream of altered splice site)
134	134	MUTAGEN E->Q: Abolishes dimerization; when associated with E-50 and E-51.				might get lost (downstream of altered splice site)
138	140	TURN				might get lost (downstream of altered splice site)
143	149	STRAND				might get lost (downstream of altered splice site)
147	147	MUTAGEN C->S: Enhances formation of fibrillar aggregates in the absence of bound zinc; when associated with S-7; S-58 and S-112.				might get lost (downstream of altered splice site)
147	147	DISULFID				might get lost (downstream of altered splice site)
151	153	STRAND				might get lost (downstream of altered splice site)

length of protein	normal
AA sequence altered	yes
position of stopcodon in wt / mu CDS	465 / 465
position (AA) of stopcodon in wt / mu AA sequence	155 / 155
position of stopcodon in wt / mu cDNA	613 / 613
poly(A) signal	N/A
conservation nucleotide level for all changes - no scoring up to now	N/A
position of start ATG in wt / mu cDNA	149 / 149
chromosome	21
strand	1
last intron/exon boundary	506
theoretical NMD boundary in CDS	307
length of CDS	465
coding sequence (CDS) position	140
cDNA position (for ins/del: last normal base / first normal base)	288
gDNA position (for ins/del: last normal base / first normal base)	4236
chromosomal position (for ins/del: last normal base / first normal base)	33036170
original gDNA sequence snippet	TGAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATA
altered gDNA sequence snippet	TGAAGGCCTGCATGGATTCCGTGTTCATGAGTTTGGAGATA
original cDNA sequence snippet	TGAAGGCCTGCATGGATTCCATGTTCATGAGTTTGGAGATA
altered cDNA sequence snippet	TGAAGGCCTGCATGGATTCCGTGTTCATGAGTTTGGAGATA
wildtype AA sequence	MATKAVCVLK GDGPVQGIIN FEQKESNGPV KVWGSIKGLT EGLHGFHVHE FGDNTAGCTS AGPHFNPLSR KHGGPKDEER HVGDLGNVTA DKDGVADVSI EDSVISLSGD HCIIGRTLVV HEKADDLGKG GNEESTKTGN AGSRLACGVI GIAQ*
mutated AA sequence	MATKAVCVLK GDGPVQGIIN FEQKESNGPV KVWGSIKGLT EGLHGFRVHE FGDNTAGCTS AGPHFNPLSR KHGGPKDEER HVGDLGNVTA DKDGVADVSI EDSVISLSGD HCIIGRTLVV HEKADDLGKG GNEESTKTGN AGSRLACGVI GIAQ*
speed	0.08 s
	Report bugs / help us to improve MutationTaster!

All positions are in basepairs (bp) if not explicitely stated differently.

RegulationSpotter

- *Extratranscriptic* variants: located outside of known transcripts

- free, web-based tool

- InDels, SNVs

- hyperlinks to MutationTaster for intragenic variants

RegulationSpotter

RegulationSpotter – submit VCF file

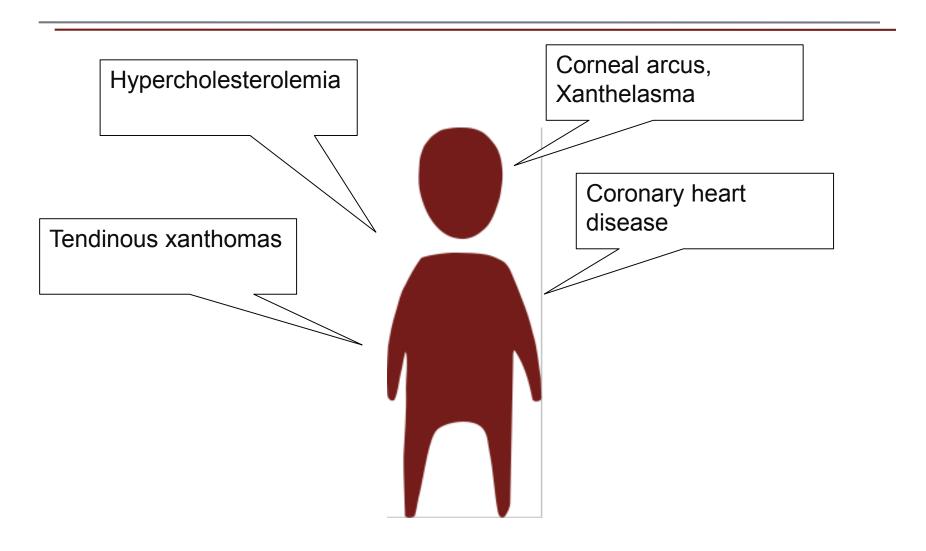**Regul@tionSpotter - annotation of extratranscriptic variants**

We offer automated analysis of variants from Whole Genome Sequencing projects in either user-defined genomic regions or in candidate genes and the genomic elements interacting with these. Variants must be in VCF format and refer to GRCh37. Look up more details in the documentation. You can also follow the provided tutorial to become familiar with the software and the results.

Project name GenomTutorialLDLR   **VCF file** [Durchsuchen...] Keine Datei ausgewählt.   **Please zip or gzip large files!**   sample file

Format:

```
#CHROM   POS    ID   REF  ALT  QUAL   FILTER   INFO                                                            FORMAT       SAMPLE
chr1     10199  .    A    C    4.77   .        DP=2;AF1=0.5003;CI95=0.25,0.75;DP4=1,0,0,1;MQ=60;FQ=-3.1;PV4=1,1,1,1  GT:PL:DP:GQ  0/1:33,0,28:2:30
```

(tab delimited) The coordinates **must** refer to GRCh37.

Analysis settings

recessive trait   ☐ homozygous variants only

                                                             1000GExAC
filter polymorphisms   homozygous in >=   4   10
                                             present (het or hom) in >=   [ ] [ ]
                                             number of individuals in 1000G or ExAC, set values to 0 to stop filtering

minimum coverage   4
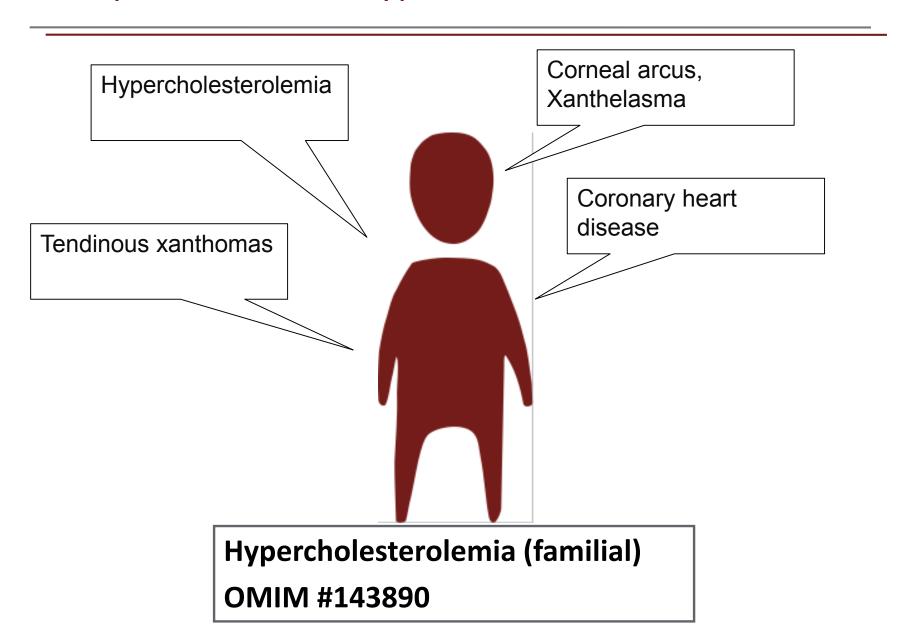
                                             ⦿ analyse complete VCF file
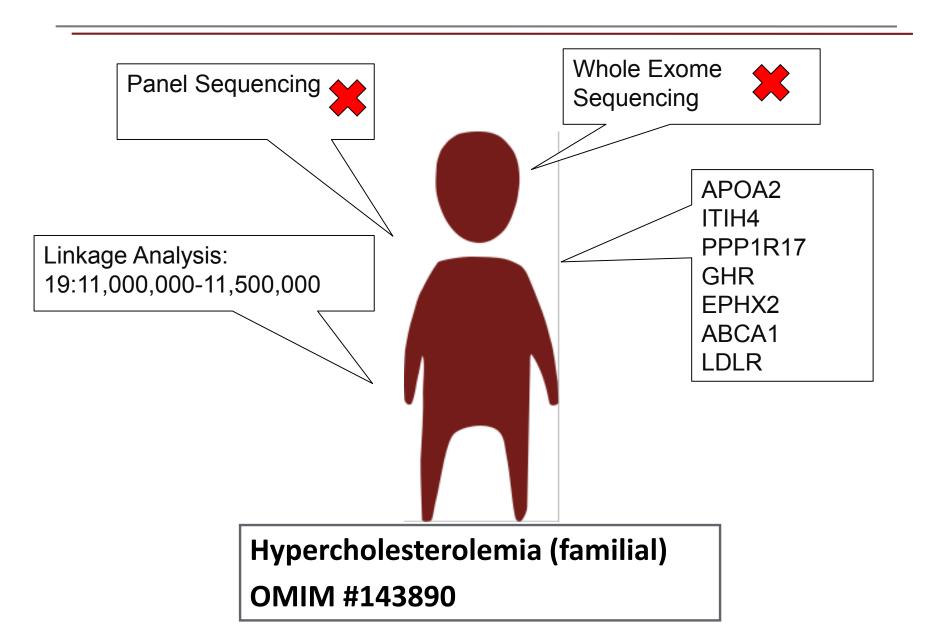   ◯ analyse custom regions (select to enter)
analyse the following regions or genes…   ◯ analyse custom genes (select to enter)
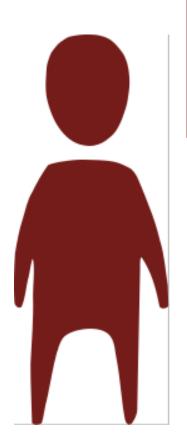
[Submit]

Example case: Phenotype

Example case: Phenotype

Example case: Previous examinations

Example case: Whole Genome Sequencing
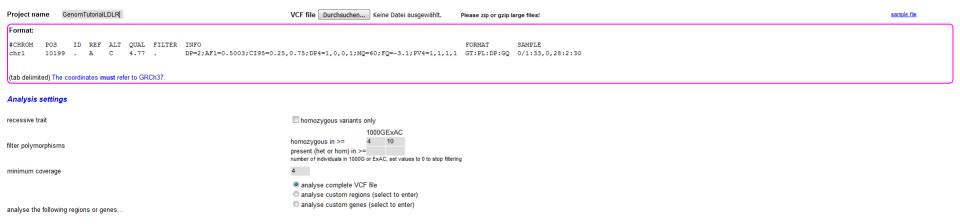
1000G Genome
(HG00101)

with known disease
mutation from ClinVar

RegulationSpotter – submit VCF file

Regul@tionSpotter - annotation of extratranscriptic variants

We offer automated analysis of variants from Whole Genome Sequencing projects in either user-defined genomic regions or in candidate genes and the genomic elements interacting with these. Variants must be in VCF format and refer to GRCh37. Look up more details in the documentation. You can also follow the provided tutorial to become familiar with the software and the results.

Project name   GenomTutorialLDLR| **VCF file** [Durchsuchen...] Keine Datei ausgewählt.   **Please zip or gzip large files!**   sample file

Format:

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
chr1	10199	.	A	C	4.77	.	DP=2;AF1=0.5003;CI95=0.25,0.75;DP4=1,0,0,1;MQ=60;FQ=-3.1;PV4=1,1,1,1	GT:PL:DP:GQ	0/1:33,0,28:2:30

(tab delimited) The coordinates **must** refer to GRCh37.

Analysis settings

recessive trait   ☐ homozygous variants only

                                                          1000G ExAC
filter polymorphisms   homozygous in >=   4   10
                                        present (het or hom) in >= ▢   ▢
                                        number of individuals in 1000G or ExAC, set values to 0 to stop filtering

minimum coverage   4

                                        ◉ analyse complete VCF file
                                        ○ analyse custom regions (select to enter)
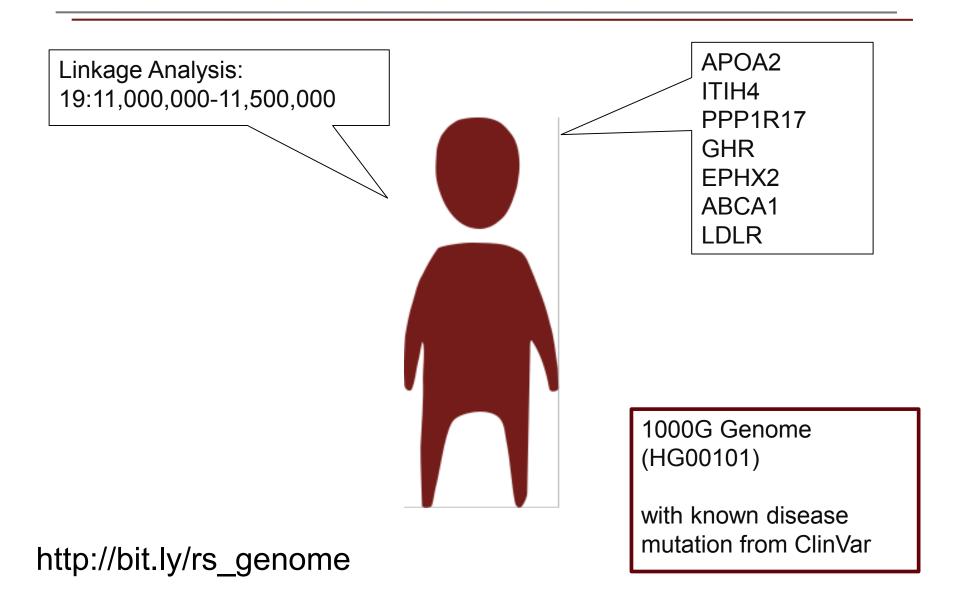analyse the following regions or genes… ○ analyse custom genes (select to enter)

[Submit]

Analysis settings:
Discard variants which appear more than 4 times in 1000G or 10 times in ExAC
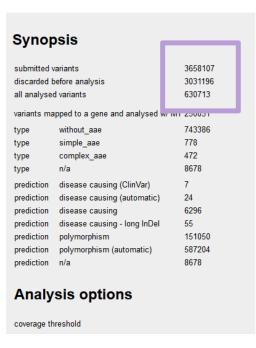
Example case: Whole Genome Sequencing

Linkage Analysis:
19:11,000,000-11,500,000

APOA2
ITIH4
PPP1R17
GHR
EPHX2
ABCA1
LDLR

1000G Genome
(HG00101)

with known disease
mutation from ClinVar

http://bit.ly/rs_genome

Filter: none

RegulationSpotter documentation

[RegulationSpotter] GenomTutorialLDLR - synopsis & display settings

project ID 738_807596

Synopsis

submitted variants		3658107
discarded before analysis		3031196
all analysed variants		630713

variants mapped to a gene and analysed w/ MT 230031

type	without_aae	743386
type	simple_aae	778
type	complex_aae	472
type	n/a	8678
prediction	disease causing (ClinVar)	7
prediction	disease causing (automatic)	24
prediction	disease causing	6296
prediction	disease causing - long InDel	55
prediction	polymorphism	151050
prediction	polymorphism (automatic)	587204
prediction	n/a	8678

Analysis options

coverage threshold

Display settings

display / filter options

- ○ show all variants
- ○ show only intragenic vars
- ⦿ show only extratranscriptic vars
- ☑ hide extratranscriptic variants without annotation

- ☐ hide known polymorphisms
- ☐ hide predicted polymorphisms
- ☐ hide silent alterations (intragenic variants)
- ☐ hide prediction problems

restrict to these genes

TP53, LBR, DMD

restrict to these regions

1:0-1000000

sort & group results

- ○ sort by chromosome, position, effect
- ⦿ sort by effect, chromosome, position
- ○ sort by gene, effect, position
- ○ sort by effect, gene, position
- ○ sort by gene, position

get the data

[display]

Filter: needed for Genome

[RegulationSpotter ... is & display settings

project ID 738_807596

Synopsis

submitted variants	3658107
discarded before analysis	3031196
all analysed variants	630713

variants mapped to a gene and analysed w/ MT 250031

type	without_aae	743386
type	simple_aae	778
type	complex_aae	472
type	n/a	8678
prediction	disease causing (ClinVar)	7
prediction	disease causing (automatic)	24
prediction	disease causing	6296
prediction	disease causing - long InDel	55
prediction	polymorphism	151050
prediction	polymorphism (automatic)	587204
prediction	n/a	8678

Analysis options

coverage threshold

- ☐ hide known polymorphisms
- ☐ hide predicted polymorphisms
- ☐ hide silent alterations (intragenic variants)
- ☐ hide prediction problems

restrict to these regions

```
1:0-1000000
```

get the data

display

Filter: needed for Genome

Regulation Spotter

RegulationSpotter results GenomTutorialLDLR

Too many variants

Please restict your search to fewer candidate genes or less / smaller regions.

Filter: regions

Display settings

display / filter options		
⦿ show all variants		☐ hide known polymorphisms
○ show only intragenic vars		☐ hide predicted polymorphisms
○ show only extratranscriptic vars		☐ hide silent alterations (intragenic variants)
☑ hide extratranscriptic variants without annotation		☐ hide prediction problems

restrict to these genes

```
TP53, LBR, DMD
```

restrict to these regions

```
19:11000000-11500000
```

sort & group results	get the data
○ sort by chromosome, position, effect	
⦿ sort by effect, chromosome, position	**display**
○ sort by gene, effect, position	
○ sort by effect, gene, position	
○ sort by gene, position	

Filter: candidate genes

Display settings

display / filter options

- ○ show all variants
- ○ show only intragenic vars
- ● show only extratranscriptic vars
- ☑ hide extratranscriptic variants without annotation

restrict to these genes

```
APOA2
ITIH4
GHR
PPP1R17
EPHX2
ABCA1
ldlr
```

- ☐ hide known polymorphisms
- ☐ hide predicted polymorphisms
- ☐ hide silent alterations (intragenic variants)
- ☐ hide prediction problems

restrict to these regions

```
1:0-1000000
```

sort & group results

- ○ sort by chromosome, position, effect
- ● sort by effect, chromosome, position
- ○ sort by gene, effect, position
- ○ sort by effect, gene, position
- ○ sort by gene, position

get the data

[display]

Example case: Whole Genome Sequencing

Linkage Analysis:
19:11,000,000-11,500,000

APOA2
ITIH4
PPIR17
GHR
EPHX2
ABCA1
LDLR

http://bit.ly/rs_genome

Results

Results - Filter: regions

RegulationSpotter results GenomTutorialLDLR

RegulationSpotter documentation

I am now applying your regions filter 19:11000000-11500000.

show / hide details

Results - Filter: regions

RegulationSpotter results GenomTutorialLDLR

RegulationSpotter documentation

I am now applying your regions filter 19:11000000-11500000.

show / hide details

Results – Details 19:11200032_112000034delCTC

Results – Details 19:11200032_112000034delCTC

Regulation Spotter

documentation

regul@tion spotting

Likely effect	**functional (high confidence)**

Model: *extratranscriptic*, score: 164.955853658537

Summary	• within TFBS • within active promoter

hyperlink

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr19:11200032_11200034delCTC IGV
alteration type	insertion / deletion
alteration region	extratranscriptic
known variant	Variant was neither found in ExAC nor 1000G.

promoters	*Ensembl Promoter* ENSR00000143671
	active A549, Aorta, CD14+CD16- monocyte (CB), CD14+CD16- monocyte (VB), CD4+ ab T cell (VB), CM CD4+ ab T cell (VB), DND-41, EPC (VB), erythroblast (CB), Fetal Adrenal Gland, Fetal Intestine Large, Fetal Intestine Small, Fetal Muscle Leg, Fetal Muscle Trunk, Fetal Stomach, Gastric, GM12878, H1-mesenchymal, H1-trophoblast, H9, HeLa-S3, HepG2, HMEC, HSMM, HSMMtube, HUVEC, HUVEC prol (CB), IMR90, iPS-20b, iPS DF 19.11, iPS DF 6.9, K562, Left Ventricle, Lung, M0 macrophage (CB), M0 macrophage (VB), M1 macrophage (CB), M1 macrophage (VB), M2 macrophage (CB), M2 macrophage (VB), Monocytes-CD14+, MSC (VB), Natural Killer cells (PB), neutrophil (CB), NH-A, NHDF-AD, NHEK, NHLF, Osteobl, Ovary, Pancreas, Placenta, Psoas Muscle, Right Atrium, Small Intestine, Spleen, T cells (PB) Roadmap, Thymus
	inactive CD8+ ab T cell (CB), neutrophil myelocyte (BM)
	poised B cells (PB) Roadmap, eosinophil (VB), Fetal Thymus, H1ESC, H1-neuronal progenitor, Monocytes-CD14+ (PB) Roadmap, naive B cell (VB), neutrophil (VB)

	RegulationSpotter				
	19: 11199639-11200189	DNase1	H3K4me3	LDLR	ENST00000557958, ENST00000545707, ENST00000535915, ENST00000455727, ENST00000560502
	19: 11199638-11200188	DNase1	H3K4me3	LDLR	ENST00000557933
	19: 11199538-11200088	DNase1	H3K4me3	LDLR	ENST00000558518
	19: 11199654-11200204	DNase1	H3K4me3	LDLR	ENST00000558013

enhancers	none found
epigenetic marks (RegulationSpotter)	**DNase1** DNase1 hypersensitive site in promoter 19:11199106-11201769 **H3K4me3** H3K4me3 in 3 cell lines, overlap with promoter 19:11199071-11204139

histone modifications	H2AZ	GM12878		show features in Ensembl explain
	H3K27ac	NHEK, DND-41, H1ESC, Osteobl, IMR90, NHLF, HUVEC, GM12878, K562, NH-A, HSMM, HSMMtube, HepG2, NHDF-AD, A549, HMEC, HeLa-S3, Monocytes-CD14+		
	H3K4me1	DND-41, HUVEC		

Results – Details 19:11200032_112000034delCTC

Results – Details 19:11200032_112000034delCTC

polymerase	PolII	H1ESC, HUVEC, GM12878, K562, HepG2, HeLa-S3	show features in Ensembl explain
open chromatin	DNase1	NHEK, H1ESC, IMR90, NHLF, GM12878, K562, NH-A, HepG2, HSMM, HSMMtube, NHDF-AD, Monocytes-CD14+, A549, HMEC, HeLa-S3	show features in Ensembl explain

TF binding sites			
	CTCF	H1ESC	show features in Ensembl explain
	Cmyc	K562	
	E2F4	HeLa-S3	
	Egr1	K562	
	FOXA1	HepG2	
	HEY1	K562	
	Ini1	HeLa-S3	
	Max	HUVEC, K562, HeLa-S3	
	SP1	H1ESC	
	TAF1	H1ESC, K562, HepG2, HeLa-S3	
	TAF7	H1ESC	

test for effects on TFBSs with ePOSSUM

genomic interactions

HiC interactions

affected gene(s)		element	cell line	transcripts
gene symbol	Ensembl ID		HMEC	
LDLR	ENSG00000130164	promoter	+	*protein-coding:* ENST00000455727, ENST00000535915, ENST00000545707, ENST00000557933, ENST00000558013, ENST00000558518 *retained_intron:* ENST00000557958, ENST00000560502
LDLR	ENSG00000130164	overlaps with promoter	+	*protein-coding:* ENST00000252444
SMARCA4	ENSG00000127616	modifier, interacts with promoter	+	*protein-coding:* ENST00000592158 *processed_transcript:* ENST00000538456 *retained_intron:* ENST00000586892

show interactions as plot

phyloP / phastCons		PhyloP PhastCons		
	(flanking)	1.259	0.854	
		3.039	0.996	
		1.298	1	
		3.039	1	
	(flanking)	3.039	1	

explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

chromosome	19
strand	1
chromosomal position (for ins/del: last normal base / first	11200031 / 11200035

Results – Details Interaction

RegulationSpotter interactions

► SMARCA4 (ENSG00000127616)

► LDLR (ENSG00000130164)

HMEC

19:11200031N>NCTC

show legend

	location of the variant
	protein-coding gene: orientation genesymbol (ENSEMBL stable ID)
	non-protein-coding gene: orientation ENSEMBL stable ID (biotype)
	►: gene on forward (+) strand ◄: gene on reverse (-) strand
	interaction between two elements reported in these cell lines

hide legend

affected gene(s)		element	cell line	transcripts
gene symbol	Ensembl ID		HMEC	
LDLR	ENSG00000130164	promoter	+	protein-coding: ENST00000455727, ENST00000535915, ENST00000545707, ENST00000557933, ENST00000558013, ENST00000558518 retained_intron: ENST00000557958, ENST00000560502
LDLR	ENSG00000130164	overlaps with promoter	+	protein-coding: ENST00000252444
SMARCA4	ENSG00000127616	modifier, interacts with promoter	+	protein-coding: ENST00000592158 processed_transcript: ENST00000538456 retained_intron: ENST00000586892

show details

Results – Score

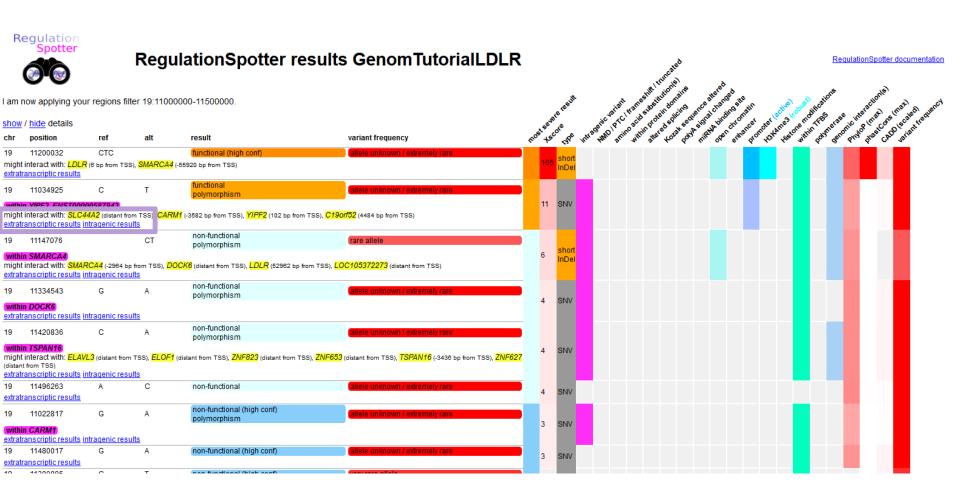All positions are in basepairs (bp) if not explicitly stated differently.
Abbreviations: aa: amino acid; CDS: coding sequence; mu: mutated; NMD: nonsense-mediated mRNA decay; nt: nucleotide; wt: wildtype; TGP: 1000 Genomes Project

show score composition

Scoring

feature	value	weight
EnsemblCellBasedFeatures*Histone|H2AZ	present	3
EnsemblCellBasedFeatures*Histone|H3K27ac	present	1
EnsemblCellBasedFeatures*Histone|H3K4me1	present	0.6
EnsemblCellBasedFeatures*Histone|H3K4me2	present	2
EnsemblCellBasedFeatures*Histone|H3K79me2	present	2
EnsemblCellBasedFeatures*Histone|H3K9ac	present	3
EnsemblCellBasedFeatures*Polymerase|PolII	present	3
EnsemblCellBasedFeatures*Transcription Factor|CTCF	present	1
EnsemblCellBasedFeatures*Transcription Factor|Cmyc	present	10
EnsemblCellBasedFeatures*Transcription Factor|E2F4	present	20
EnsemblCellBasedFeatures*Transcription Factor|Egr1	present	10
EnsemblCellBasedFeatures*Transcription Factor|FOXA1	present	5
EnsemblCellBasedFeatures*Transcription Factor|HEY1	present	10
EnsemblCellBasedFeatures*Transcription Factor|Ini1	present	10
EnsemblCellBasedFeatures*Transcription Factor|Max	present	5
EnsemblCellBasedFeatures*Transcription Factor|SP1	present	10
EnsemblCellBasedFeatures*Transcription Factor|TAF1	present	10
EnsemblCellBasedFeatures*Transcription Factor|TAF7	present	20
HiC*interaction	present	0.2
RS*DNase1	present	10
RS*H3K4me3	present	10
RS*activepromoter	present	10
phastcons	1	5
phylop	0.831170731707317	5

Results – Details intragenic variant

Results – Details intragenic variant

5 result(s)

1 some transcripts had annotation problems and are not shown

genesymbol	prediction	prob	model	problems	splicing	known mutations	AA changes	region	dbSNP	features	link
-	functional	n/a	extratranscriptic					extratranscriptic	538930743		show file
YIPF2	polymorphism	1.06326226900766e-06	without_aae					intron	538930743		show file
YIPF2	polymorphism	1.06326226900766e-06	without_aae					intron	538930743		show file
YIPF2	polymorphism	1.06326226900766e-06	without_aae					intron	538930743		show file

Results – view MutationTaster

Alteration 19:11034925C>T_0_ENST00000586748

Prediction   **polymorphism**   Model: *without_aae*, prob: 0.999998936737731   (explain)

name of alteration	19:11034925C>T_0_ENST00000586748
alteration (phys. location)	chr19:11034925C>T show variant in all transcripts IGV
HGNC symbol	YIPF2
Exac loss-of-function metrics	LOF: 0.00, misssense: -0.38, synonymous: 0.27
Ensembl transcript ID	ENST00000586748
Genbank transcript ID	none found
UniProt peptide	N/A
alteration type	single base exchange
alteration region	intron
DNA changes	c.368-57G>A g.4433G>A
AA changes	N/A
position(s) of altered AA if AA alteration in CDS	N/A
frameshift	no
length of protein	N/A
known variant	Reference ID: rs538930743

known variant	database	homozygous (T/T)	heterozygous	allele carriers
	1000G	0	1	1
	ExAC	-	-	-

regulatory features	TF binding site, Regulatory Feature, Transcription factor binding site H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation PolII, Polymerase, RNA Polymerase II
multicell regulatory features	N/A
external features from FANTOM5 and/or VISTA	N/A
miRNA binding site(s)	no miRNA binding sites annotated

phyloP / phastCons

	PhyloP	PhastCons
(flanking)	0.023	0
	-1.176	0
(flanking)	-0.356	0.001

explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

Thank you!

Contact:

Charité Translational Genomics
https://translationalgenomics.charite.de/

jana-marie.schwarz@charite.de

mutationtaster.org
regulationspotter.org

```perl
my %predictions = (                                              #hash to encode
                      '1' => 'disease causing (ClinVar)',
                      '2' => 'disease causing (automatic)',
                      '3' => 'disease causing',
                      '4' => 'disease causing - long InDel',
                      '21' => 'functional (high conf)',
                      '22' => 'functional',
                      '29' => 'non-functional (high conf)',
                      '28' => 'non-functional',
                      '95' => 'polymorphism',
                      '99' => 'polymorphism (automatic)'
                      );
$self->{build}=37;
$self->{schema}->{qe}='regulationspotter.';
$self->{schema}->{ensembl}="ensembl37_84.";
$self->{schema}->{regulation}="ensembl_regulation37_84."; ## ACHTUNG: Hier noch
$self->{schema}->{encode}='encode'. $self->{build}.'.'        ;
$self->{schema}->{build}='build'.$self->{build}.'.';
$self->{links}->{ensembl_URL}='http://grch37.ensembl.org/';
if ($self->{id}) {
    ($self->{project_number},$self->{accesscode})=split /_/,$self->{id};
}
my %predictions_reverse = ();
foreach my $key (keys %predictions) {
    $predictions_reverse{$predictions{$key}}=$key;
}
my %prediction_problems = (
                      '0' => 'error',
                      '1' => 'wrong input format',
                      '2' => 'DB error',
                      '3' => 'network problem',
                      '4' => 'data problem',
                      '5' => 'query cannot be processed',
                      '6' => 'software error',
                      '7' => 'annotation problem'
);
```