LIMSTILL platform

managing resequencing and TILLING projects in model organisms Victor Guryev, Eugene Berezikov, Ronald H. A. Plasterk and Edwin Cuppen Hubrecht Laboratory, Uppsalalaan 8, 3584CT Utrecht, The Netherlands

LIMSTILL (LIMS for Identification of Mutations by Sequencing and TILLING) is an open-source software designed to streamline the informatics and management parts of the Hubrecht Laboratory facility for high-throughput screening for induced mutations in model organisms. It includes steps for amplicon selection, primer design, sequence analysis, and annotations. The platform is universal and can be used for any resequencing or TILLING (target induced local lesions in genomes) project for any organism of interest.

1. The purpose of LIMSTILL is to maximally automate and efficiently manage all steps of the resequencing and tilling process. The public LIMSTILL web-service is available at http://limstill.niob.knaw.nl



2. Amplicon selection. Based on structural gene information, LIMSTILL proposes a list of possible amplicons to be selected for further analysis. The graphical representation of the gene structure and amplicon ratings reflecting the chance to identify a knockout based on the mutation spectrum and amplicon characteristics, assist the researcher to pertype: select gene regions to be targeted. User can also specify an amplicon with custom coordinates.

Design	Exons	Exon_IDs	Position	Size	Coding	Stop chance	Score	Status
8	1	ENSRNOE00000272097	10011035	35	35	0.65 %	22.75	
×	12	ENSRNOE00000272097 ENSRNOE00000123895	10011248	248	72	6.27 %	451.44	
8	2	ENSRNOE00000123895	12121248	37	37	11.79 %	436.23	ĺ.
8	3	ENSRNOE00000124319	44004577	178	178	4.16 %	740.48	
	34	ENSRNOE00000124319 ENSRNOE00000124734	44004932	533	285	4.22 %	1202.7	designed
	4	ENSRNOE00000124734	48264932	107	107	4.31 %	461.17	
×	5	ENSRNOE00000272081	62976590	294	293	5.5 %	1611.5	
	c1	Custom	1001	1800	1			0

Tilling (4 primers with ends)

Once registered, user can create new project by specifying Ensembl gene ID, Genbank accession or supplying a file with custom sequence in Genbank/EMBL format.

. Provide project identification (max. 8 characters):		
FGFH		
2. Provide project description:		
Rat fibroblast growth factor H		
Specify data source		
 Specify data source Ensembl gene ID (recommended): 	ENSRN0G0000012912	(e.g. ENSG0000000003
Specify data source Ensembl gene ID (recommended): Genbank ID:	ENSRNOG0000012912	(e.g. ENSG0000000003) (e.g. NM_065433)

LIMSTILL fetches information for the gene from Ensembl database or parses provided annotation and generates a graphical representation of the gene that includes intron/exon structure, alternative transcripts and location of annotated protein domains/ features (central figure).

Single amplicon (2 primers Nested amplicon (4 primers

Start design Reset

For resequencing projects, for scanning genes for SNPs/mutations, all exons can be selected. There are three options for mutation detection method (resequencing using either a single or nested PCR or using TILLING) available in this section.



3. Primer design is based on a custom interface to primer3 program to design primers for nested PCR under universal conditions, allowing easy automated processing robotic in The primer design step setups. performs iterative primer search and masking for repetitive sequences.

				Amplico	n 5
AMPLICON T 200-524:					
PCR1 PRODUCT SIZE: 578	ł				
rnFGFH-T-1	110	20	57.71	50.00	ATGACTTCTTGCTCCCA
rnFGFH-T-4	668	20	57.29	50.00	ATCACCCTGGGATACAC
Parameters: -mingc 40	-maxpoly	х 3 - (gcclamp 1		
PCR2 PRODUCT SIZE: 401	-				
rnFGFH-T-2	166	18	57.12	55.56	CAACCAGGGTGAATGGA
rnFGFH-T-3	548	19	57.02	52.63	AAGAGGGAAGGCTGTTG
Parameters: -mingc 40	-maxpoly	x 3 - (gcclamp 1		
rnF	GFH-T-1				
>>>	>>>>>>>>>	>>>>>	>>>		
cccactagacaggggcacaatg	acttettge	tecca	cacgggttcta	acaacaat	aggeceaaaggeegacea
cetectetgeteetttetteea	tagCCGAGC	GGGAA	GAGCAAAGAC'	GCGTATT	CACCGAGATCGTACTGGAG
CGGCATGAGGGCTGGTTCATGG	CTTTCACCC	GGCAG	accaaccme	ICC AGCC	Recedence
		JULAU	age code ce rei	JUCAGGUU	LUUGGAGUUGUUAGAAUUA
TCTACCAAGGCCAGCTGCCTTT	CCCCAACCA	CGCTG	AAAGGCAGAA	CAGTTCG	ATTCGTGGGCTCCGCCCC
TCTACCAAGGCCAGCTGCCTTT	CCCCAACCA	CGCTG	AAAGGCAGAA	CAGTTCG	AATTCGTGGGCTCCGCCCCC rnFGFH-T-3
TCTACCAAGGCCAGCTGCCTT	CCCCAACCA	CGCTG	AAAGGCAGAAG	CAGTTCG	AATTCGTGGGCTCCGCCCCC rnFGFH-T-3 <<<<<<<<<<<<
GCCCCAGTCCCAAACGTAGt Ca	CCCCAACCA	CGCTG	AAAGGCAGAA	GCAGGTCG	AATTCGTGGGGCTCCGCCCCC rnFGFH-T-3 <<<<<<<<<<<
GCCCCAGTCCCAAACGTAGtca	CCCCAACCA	CGCTG2	AAAGGCAGAAG	GCAGGCC GCAGTTCG	ATTCGTGGGCTCCGCCCC rnFGFH-T-3 <<<<<<<<<<<>
GCCCCAGTCCCAAACGTAGtca	CCCCAACCA	CGCTG	AAAGGCAGAAG	JCAGTTCG	AATTCGTGGGCTCCGCCCCC rnFGFH-T-3 <<<<<<<<<<<<<<<
GCCCCAGTCCCAAACGTAGtca	CCCCCAACCA	CGCTG2	AAAGGCAGAAG	GEAGGTEG	ATTCGTGGGCTCCGCCCC rnFGFH-T-3 <<<<<<<<<<>
GCCCCAGTCCCAAACGTAGtca ggctggggtggaaggtgggggg	CCCCCAACCA	CGCTG tagtgi aaaag	AAAGGCAGAAG tgcaaggctgo gaccccgaggo	JCAGTTCG	ATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<<<<<<> cccctcaacagccttccctc
GCCCCAGTCCCAAACGTAGtca ggctggggtggaaggtggggg catttatggaacaggtt 790	CCCCCAACCA gggggggcct ccaaactcc	CGCTG tagtgi	AAAGGCAGAAG tgcaaggctgg gaccccgaggg	JGTAGATTCG	ATTCGTGGGCTCCGCCCC rnFGFH-T-3 <<<<<<<<<<>cccctcaacagccttccctc
GCCCCAGTCCCAAACGTAGtca ggctggggggggaaggtgggggg cattttatggaacaggtt 790	CCCCAACCA	CGCTG tagtgi	AAAGGCAGAAG	JGTAGATTCG	ATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<<<<>cccctcaacagccttccctd
TCTACCAAGGCCAGCTGCCTTT GCCCCAGTCCCAAACGTAGtca ggctgggggggggaaggtggggggg cattttatggaacaggtt 790 Primers:	CCCCAACCA	CGCTG	AAAGGCAGAAG	JGTAGATTCG	ATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<< <c< td=""></c<>
GCCCCAGTCCCAAACGTAGtca ggctggggggggaaggtgggggg catttatggaacaggtt 790 Primers: EGEH-5-1: ATGACTTCTTGCT	CCCCAACCA	CGCTG	AAAGGCAGAAG	JGTAGATTCG	ATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<<<>cccctcaacagccttccctd
TCTACCAAGGCCAGCTGCCTTT GCCCCAGTCCCAAACGTAGtca ggctgggggggggaaggtgggggg catttatggaacaggtt 790 Primers: FGFH-5-1: ATGACTTCTTGCT	CCCCAACCA	CGCTG	AAAGGCAGAAG	JGTAGATTCG	ATTCGTGGGCTCCGCCCC rnFGFH-T-3 <<<<<<<< <c< td=""></c<>
TCTACCAAGGCCAGCTGCCTTT GCCCCAGTCCCAAACGTAGtca ggctgggggggggaaggtgggggg catttatggaacaggtt 790 Primers: FGFH-5-1: ATGACTTCTTGCT FGFH-5-4: ATCACCCTGGGA	CCCCAACCA	CGCTG tagtgi	AAAGGCAGAAG tgcaaggctgc gaccccgaggc	JGTAGATTCG	ATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<< <c< td=""></c<>
TCTACCAAGGCCAGCTGCCTTT GCCCCAGTCCCAAACGTAGtca ggctgggggggggagggggggggggggggggggggg	CCCCAACCA	CGCTG2 tagtg1 aaaag0	GGGTGAATG	GAC	AATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<< <ccctcaacagccttccctd< td=""></ccctcaacagccttccctd<>
TCTACCAAGGCCAGCTGCCTTT GCCCCAGTCCCAAACGTAGtca ggctggggggggaaggtgggggg catttatggaacaggtt 790 Primers: FGFH-5-1: ATGACTTCTTGCT FGFH-5-4: ATCACCCTGGGA FGFH-5-2: TGTAAAACGACG FGFH-5-3: AGGAAACAGCTA	CCCCAACCA gggggggcct ccaaactcc CCCACAC TACACAAG GCCAGT CA TGACCAT A	CGCTG tagtg1 aaaag ACCAG	AAAGGCAGAAG tgcaaggctgg gaccccgaggg GGGTGAATG	GAC TTGAG	ATTCGTGGGCTCCGCCCCG rnFGFH-T-3 <<<<<<<< <cccctcaacagccttccctc< td=""></cccctcaacagccttccctc<>

	Description	Species	Started	Status
Argonaut	C. elegans Argonaute gene	Caenorhabditis_elegans	26/7/2004	Created
DicerMm	Mouse dicer gene	Mus_musculus	26/7/2004	Created
FGF16	Rat fibroblast growth factor 16	Rattus_norvegicus	26/7/2004	Created
FGFH	Fibroblast growth factor H	Rattus_norvegicus	26/7/2004	Created
FGFR4	Mouse fibroblast growth factor receptor 4	Mus_musculus	26/7/2004	Created
ManBeta	Mannosidase, beta A, lysosomal	Danio_rerio	26/7/2004	Created

4. Amplicon view provides the interactive graphical representation of information on intron/exon structure of an amplicon, coding frames, primer positions and mapped mutations.

5. Mapping and annotation of mutations is accessible from amplicon view and can be done in four ways:

- 1) by providing coordinates in the amplicon and mutant allele
- 2) by specifying nucleotide context of a mutation for pattern search
- 3) by supplying preformatted fasta file
- 4) by submitting sequencing chromatograms of samples from a screen for mutation detection

ort.	taacttet	ctagaactet	accocadada	octaccacat	ceactecta	attt	(SDecit
			accccagggg			l	
	50	60	70	80	90	100	by prim
1	G A A	RLL	PNLT	#			51
GGG	AGCCGCCC	GCCTGCTGCC	TAACCTTACC	CTgtaagtgtg	lefdeefdedd	ectgg	Adiu
	3	1	1	1		1	
	150	160	170	180	190	200	Exo
ccg	cgcgcgcd	cccagggaca	ectctcgggc	aggtggggcaa	igacaaaggto	14444	>5-
		1	1	1	1	1	oct
	250	260	270	280	290	300	001
						105257	Cac
	L	C #L O	L L #I L	ссо	то		cto
cc	aaacadg	GCTTGCAGCT	ATTGATCCTC	TGCTGTCAAAC	GCAGgtagg	Cacco	gad
	1	1	1	1	1	1	1
	350	360	370	380	390	400	tcc
	120	100	2.19	100	1220	100	GA
				FGFU. 1	2-4		CG
				rorn-1	_4-4 		AT
							TT
Jay	yyyayyca I	iggactggeet	lecteracea	aycaactytta	iayycccyga;	Jggcc	
							CA

Accepting candidate mutations in amplicon 1_2, position 185

FGFH-1 2-2

FGFH-1 2-1

110

210

teectecattettecatecetgaaaacetgtggeteegagaga

ggagcagaggggcaaatcgccccaagaagtctctccagcgA

igttettttttgteattteeaggaatgttattaeceeaeeeae

ggetgtacettggggtgeeeacaeetggaaetgeettgtet

tgeetageeeccaaceeccaacettteeccaeetcatetegaa

230

220

The process of primer design can be repeated until satisfactory results are obtained. The suggested boundaries of amplicon can be interactively altered fied by square brackets) followed ners redesign.

st your amplicons (Amplicon design mode: tilling)

n(s): 5 Exon(s) ID(s): ENSRNOE00000272081 Coords: 6297 - 6590 Size: 294

tgctgtcttgctaacaaaggtgcccatcccactagacaggggcacaatgacttcttg ccacacgggttctacaacaattaggcccaaaggccgaccaggcaaccagggtgaatg aaatteeceteetetge

tttcttccatadCCGAGCGGGAAGAGCAAAGACTGCGTATTCACCGAGATCGTACT CTACCAAGGCCAGCTGCCTTTCCCCCAACCACGC CGAATTCGTGGGCTCCGCCCCCCCCCGCAGGACCAAGCGCACTCGGAGGCCC AACGTAGtcagggaggccttag

tgtgcaaggctgggtagatgcccctcaacagccttccctcttttctctaactcatccaaa gatgggggctggggtggaaggtggggagccaaactccaaaaggaccccgagggcatgaag tttcccactgggaagagacagtacttgtgtatcccagggtgatcaccattttatggaaca ggtttgaagtcagaggcaga

In the latter case phred/phrap/polyphred package is used for sequences analysis and identification of potential homo- and heterozygous mutations. The analysis is done in the background and the researcher is presented with chromatograms of potential mutation-harboring regions for manual inspection and approval.

Sample (method)	Context	Position	Transcript	Trivname	Description	Splicing
rat10n02 (coordinate)	5'- GTCTCTCCAG (C>A) GATGGGAGCC -3'	GTCTCTCCAG (C>A) ATGGGAGCC -3' 139139(amplicon 1_2) 999999(gene_map) g2C>A(sysname) ENSRNOP00000017723-ENSRNOT00000017723			No	
rat14f16 (pattern)	5'- CAGCTATTGA (T>G) CCTCTGCTGT -3'	369369(amplicon 1_2) 12291229(gene_map) g.+229T>G(sysname)	ENSRNOP00000017723-ENSRNOT00000017723	1185	substitution, nonconservative	No
rat08a01 <mark>(</mark> fasta)	5'- ACCTTACCCT (G>A) TAAGTGTGCT -3'	176176(amplicon 1_2) 10361036(gene_map) g.+36G>A(sysname)	ENSRNOP00000017723-ENSRNOT00000017723			Affected
rat21g12 (chromatogram)	5'- AACAGGTGCT (TG>T) CAGCTATTGA -3'	357358(amplicon 1_2) 12171218(gene_map) g.+217TG>T(sysname)	ENSRNOP00000017723-ENSRNOT00000017723	L14ins17X	out-of-frame translation, truncation	No

Submitted/discovered mutations are then annotated for their impact on gene structure (silent, AA-change, stopcodon, splice-site) and added to the graphical interface. The summary on mutations obtained in the project can be viewed in the mutations view.



LIMSTILL is a LAMP-type (Linux, Apache, MySQL, Perl) open-source project that uses HTML templates and extensively utilizes open-source software like BioPerl, Ensembl, Emboss, Staden, primer3, phred/phrap/ polyphred.

Design primers Reset

The accepted primers are then put in the user basket for ordering and as well as in the list associated with the project. These lists can be exported as tab-delimited files.

Primers to order (primer basket)

Project	Amplicon	Primer	Start	End	Tm	GC	Sequence	Length
FGFH	3_4	FGFH-3_4-3	5002	5021	57.02	50.00	AGGAAACAGCTATGACCAT	20
FGFH	3_4	FGFH-3_4-2	4303	4322	56.25	55.00	TGTAAAACGACGGCCAGT	20
FGFH	3_4	FGFH-3_4-1	4263	4281	57.49	52.63	TCTTCAGCTCGCACCATAC	19
FGFH	3_4	FGFH-3_4-4	5055	5075	57.05	47.62	AGATCTCATCCTGTGAAGCAG	21
DicerMm	2_3	DicerMm-2_3-F	20881	20898	56.83	55.56	CTCTTGTTCCCGTGCTTC	18
DicerMm	2_3	DicerMm-2_3-R	22034	22054	57.62	52.38	GCAGGACTTCAGTAAGCACAC	21
DicerMm	4	DicerMm-4-R	22870	22889	57.63	55.00	TCAGCTAGCTAAGGCTGAGG	20
DicerMm	4	DicerMm-4-F	22623	22639	57.74	58.82	CTCAGGAGCCGATTTGG	17

