

Table S1. Top-10 transfer RNA (tRNA) clusters.

cluster_id	category	chr	start	end	n_features	n_bystanders
C3	tRNA	chr6	26,286,526	28,989,083	165	NA
C11	tRNA	chr1	161,399,700	161,621,748	38	NA
C14	tRNA	chr1	143,584,701	146,488,936	37	NA
C8	tRNA	chr7	149,310,190	149,707,740	23	NA
C6	tRNA	chr17	8,119,155	8,227,064	19	NA
C7	tRNA	chr16	2,927,661	3,196,225	19	NA
C4	tRNA	chr5	181,097,070	181,222,467	18	NA
C9	tRNA	chr14	20,609,336	20,684,087	16	NA
C13	tRNA	chr1	148,011,059	148,529,330	16	NA
C1	tRNA	chr1	16,520,585	16,889,750	14	NA

tRNA clusters analysis

Human (genome assembly: GRCh38/hg38) tRNA gene predictions were downloaded in bed format (<http://gtrnadb.ucsc.edu/genomes/eukaryota/Hsapi38/hg38-tRNAs.tar.gz>) from the genomic tRNA database GtRNAdb (<http://gtrnadb.ucsc.edu/>). The file named hg38-tRNAs.bed was passed to ClusterScan as FEATURES file without any modification. The ANNOTATION file consists of a two-columns, tab-delimited table containing the unique tRNA id in the first column and a custom made category (in this case: “tRNA”) in the second column and was produced through the free software environment for statistical computing and graphics R. ClusterScan was ran with clustermean using the following parameters: -k 3; -e 2; -n2; -w 500000; -s 250000 resulting in 17 total clusters found starting from 619 unique features.

Table S2. Top-10 full length non-intact L1s (FLnI-L1) clusters.

cluster_id	category	chr	start	end	n_features	n_bystanders
C204	FLnI-L1	chrX	75,554,661	76,747,370	51	NA
C189	FLnI-L1	chrX	56,805,326	58,024,091	46	NA
C212	FLnI-L1	chrX	66,268,161	67,495,006	41	NA
C207	FLnI-L1	chrX	72,977,213	73,777,090	32	NA
C188	FLnI-L1	chrX	56,046,767	56,747,914	30	NA
C213	FLnI-L1	chrX	64,602,482	65,393,281	28	NA
C202	FLnI-L1	chrX	76,878,710	77,396,845	25	NA
C272	FLnI-L1	chr11	48,185,251	48,808,099	20	NA
C201	FLnI-L1	chrX	78,151,947	78,649,427	19	NA
C14	FLnI-L1	chr6	85,756,815	86,084,705	17	NA

Full length non-intact L1s (FLnI-L1s) clusters analysis

Human (genome assembly: NCBI38) full length non-intact L1s (FLnI-L1s) annotations were downloaded in bed format (http://l1base.charite.de/BED/hsflil1_8438.bed) from the L1Base (v2) database of putatively active LINE-1 (L1) insertions (<http://l1base.charite.de/>). The file named hsflil1_8438.bed was passed to ClusterScan as FEATURES file without any modification. The ANNOTATION file consists of a two-columns, tab-delimited table containing the unique FLnI-L1 id in the first column and a custom made category (in this case: “FLnI-L1”) in the second column and was produced through the free software environment for statistical computing and graphics R. ClusterScan was ran with clustermean using the following parameters: -k 3; -e 2; -n 2; -w 500000; -s 250000 resulting in 435 total clusters found starting from 13,418 unique features.

Table S3. Comparison of commonly used tools to identify genomic clusters.

Tool name	Availability	Distribution	Example dataset	Input	Used feature	Used category	Search strategy	Principal output	Clusters coordinates in output	Bystanders	Singletons	Substrate prediction	Used species	Pubmed id
antiSMASH	https://antismash.secondarymetabolites.org	web server; standalone	yes	EMBL or GenBank file of DNA sequences.	gene	none	sequence similarity; co-localization	Clusters annotation in non standard format.	yes	no	no	yes	bacteria; fungi	21672958
C-Hunter	http://fglab.org	standalone	yes	Ordered list of genes; GO annotation.	gene	Gene Ontology (GO)	co-localization	List of clustered genes.	no	no	no	no	any	19531248
CASSIS/SMIPS	https://sbi.hki-jena.de/cassis/	web server; standalone	yes	Fasta file of DNA sequences; gene coordinates.	gene	none	co-localization; co-regulation	List of clusters (predicted starting from a selected anchor gene).	no	no	no	no	fungi	26656005
ClusterFinder	https://github.com/petercim/ClusterFinder	standalone	yes	Gene coordinates; Pfam coordinates.	gene	Pfam domains	co-localization	Clusters annotation in non standard format.	yes	no	no	no	bacteria	25036635
ClusterScan	https://github.com/pyrevo/ClusterScan	standalone	yes	Bed file of features; any categorical annotation.	any mappable feature	(custom made included)	co-localization; shared category	Bed formatted clusters annotation.	yes	yes	yes	no	any	N/A
ClustScan-Professional	http://bioserv.pbf.hr/cms/index.php?page=clustscan	standalone	yes	Fasta file of DNA sequences.	gene	none	sequence similarity; co-localization	Putative domains coordinates of clustered PKS and NRPS genes.	no	no	no	yes	bacteria	18978015
EvoMining	https://github.com/nselem/EvoMining	standalone	yes	Fasta files of DNA and protein sequences.	gene	none	phylogenomics	Phylogenetic tree.	no	no	no	no	bacteria	27289100
FunGeneClusterS	https://fungiminions.shinyapps.io/FunGeneClusterS/	web server; standalone	yes	Gene coordinates; expression data.	gene	expression	co-expression; co-localization	List of clustered genes.	no	no	no	no	fungi	29062935
NP:searcher	http://dna.sherman.lsi.umich.edu/	web server; standalone	yes	Fasta file of DNA sequences.	gene	none	sequence similarity; co-localization	SMILES of the predicted PKS and NRPS natural products.	no	no	no	yes	bacteria	19531248
SMURF	http://www.jcvi.org/smurf	web server	yes	Fasta file of protein sequences; gene coordinates.	gene	Pfam; TIGRFAM	co-localization	List of clustered genes.	no	no	no	no	fungi	26656005