Welcome!

Open Geneious (or your software of choice for viewing, annotating and editing sequences)

Open the workshop webpage: **https://www.allencell.org/scge-workshop-2019.html**

If you had any trouble completing the downloads of the NUP153 or mEGFP sequences, we are available to help you now!

Internet access:

"Institute Guest" pw: sounders

Stem cells and Gene editing workshop

3/27/2019

Ru Gunawardane
Director, Stem cells and Gene editing

ALLEN INSTITUTE *for*
CELL SCIENCE

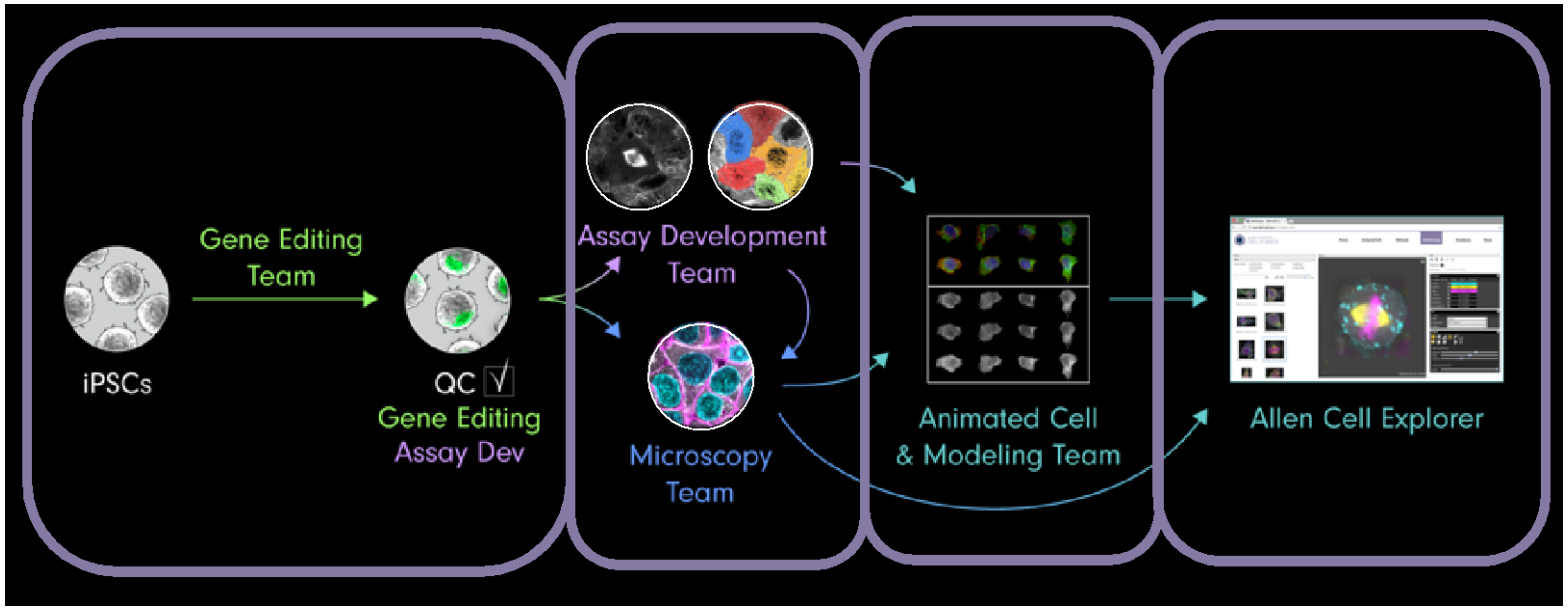
Allen Institute for Cell Science

Institute Mission

- Understand and predict cellular behaviors in normal and pathologic contexts
- Develop dynamic, visual data on cellular organization and activities
- Share our data and tools with the research community

Institute values

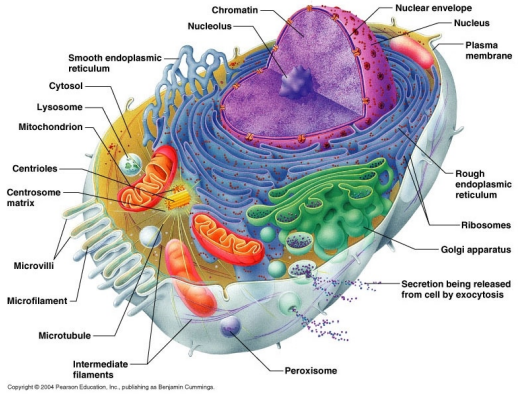
- Team science
- Open science
- Community empowerment



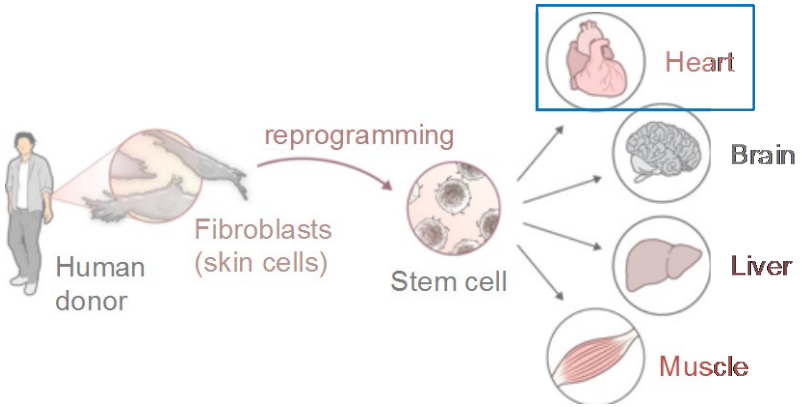
- Research program launched in January, 2016
- Highly multi-disciplinary team of ~65

Creating a stem cell state map with gene edited hiPSCs and live imaging

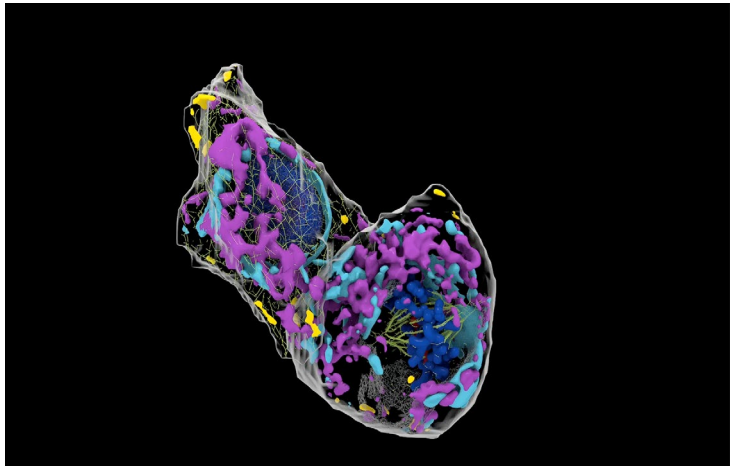
- *Physical organization of the cell – "google" map of the cell*
- *Study how cells change across states (division, differentiation, disease)*



iPSCs as a cell model



- Human
- Diploid
- Non-transformed
- Proliferative
- Pluripotent

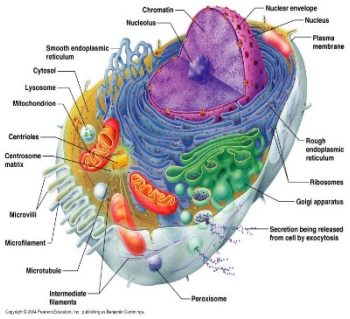


Data driven, integrated model of a cell

Critical points and rationale:

- iPSCs from a healthy donor - to understand baseline and variance
- Live imaging - capture dynamics and avoid artifacts of fixation and staining
- Endogenous tagging - Normal regulation of tagged structure
- High level of QC - data used for creating baseline models of cell behavior

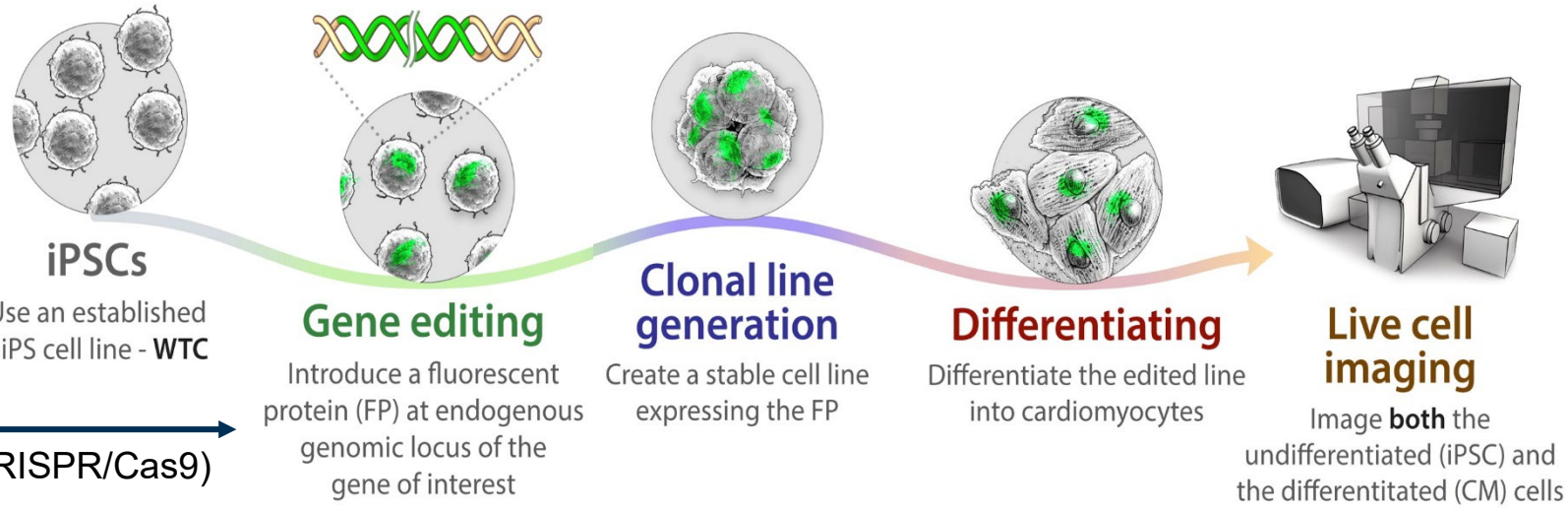
Creating a collection of FP tagged hiPSC lines representing major structures of the cells



Working with experts to select key proteins to tag and how to tag them

	Structure	Gene/Protein
1	Adhesions	Paxilin
2	Microtubules	Alpha tubulin
3	Nucleus	LaminB1
4	Mitochondria	Tom20
5	Cell-cell junctions	Desmoplakin
6	Actin	Beta actin
7	ER	Sec61B
8	Nucleolus	Fibrillarin
9	Golgi	ST6Gal1
10	Centrosome	Centrin2

(CRISPR/Cas9)

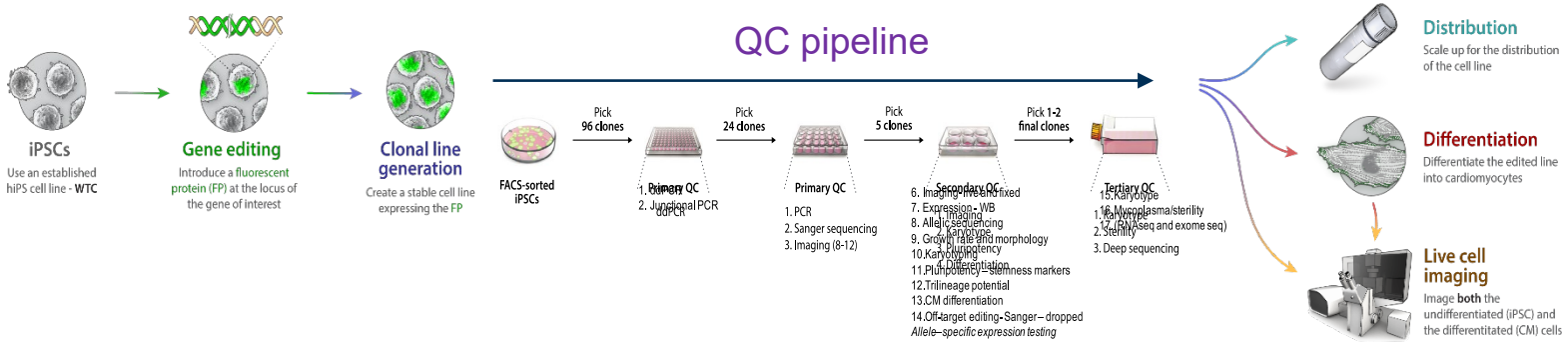


- Generate a collection of high quality clonal FP-tagged iPSC lines
- Develop editing strategy, workflows and QC methods
- Share cell lines, plasmids, images, and methods with the community

The Allen Cell and Plasmid Collections

	Structure	Gene/Protein	FP	Status
1	Adhesions	Paxilin	mEGFP	Available
2	Microtubules	Alpha tubulin	mEGFP	Available
3	Nuclar lamin	LaminB1	mEGFP	Available
4	Mitochondria	Tom20	mEGFP	Available
5	Cell-cell junctions	Desmoplakin	mEGFP	Available
6	Actin	Beta actin	mEGFP	Available
7	ER	Sec61B	mEGFP	Available
8	Nucleolus	Fibrillarin	mEGFP	Available
9	Myosin	Myosin IIB	mEGFP	Available
10	Tight junctions	ZO-1	mEGFP	Available
11,12	Cytoplasm	Safe harbor	mEGFP	Available
13	Centrosome	Centrin2	mTagRFPt	Available
14	Microtubules	Alpha tubulin	mTagRFPt	Available
15,16	Golgi	**ST6Gal1**	mEGFP	Available
17	Lysosome	LAMP1	mEGFP	Available
18	Autophagosomes	LC3	mEGFP	Available
19,20	Endosome	**Rab5**	mEGFP	Available
21	Peroxisomes	PMP34	mEGFP	Available
22	Plasma membrane	Safe harbor-CAAX	mTagRFPt	Available
23	Gap junctions	Connexin43	mEGFP	Available
24	Nucleolar	NPM1	mEGFP	Available
25	Nuclear pore	NUP153	mEGFP	Available
26	Histones	H2B	mEGFP	Available
27	Beta catenin	CTNNB1	mEGFP	Available
28	Phase transition	FUS	mEGFP	Available
29	Nuclear cohesin	SMC1A	mEGFP	Available
30	ER/Nuclear lamin	Sec61B/LMNB1	mEGFP, mTagRFPt	Available
31	CM- Thin filament	ssTnni1	mEGFP	Available
32	CM- Thick filament -	MYL7 (early)	mEGFP	Available
33	CM- M-disc	Titin	mEGFP	Available
34	CM- SR	ATP2A2	mEGFP	Available
35	CM- Thick filament -	MYL2 (late)	mEGFP	Available
36	CM- Z-disc	ACTN2	mEGFP	Available
37	Transcription factor	Sox2	mEGFP	Pending
38	Nucleolar dual	FBL/NPM1	mEGFP	Pending

4-6 month process



- Endogenous FP-tagging with CRISPR/Cas9 (38 lines)
  - Extensive QC for cell biological, genomic, and stem cell criteria
- Sharing cell lines, plasmids, and methods
  - Cell lines and plasmids - through Coriell and Addgene
  - Distribution to stem cell cores at major institutes
  - Methods and tutorials - MBoC and Stem cell Reports (Roberts et al., 2017, 2019)
  - Allencell.org

Bolded- *both mono and bi-allelic lines*
Red – *Cardiomyocyte-specific edits*

The Allen Institute for Cell Science team

The Stem Cell and Gene Editing group

Allen Institute for Cell Science: Gene Editing, Part I: Gene editing design

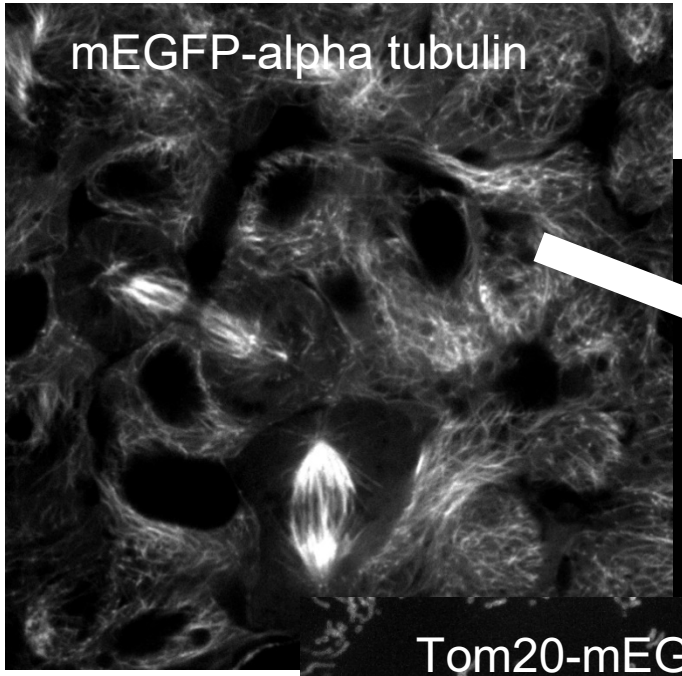
Amanda Haupt & Tanya Grancharova

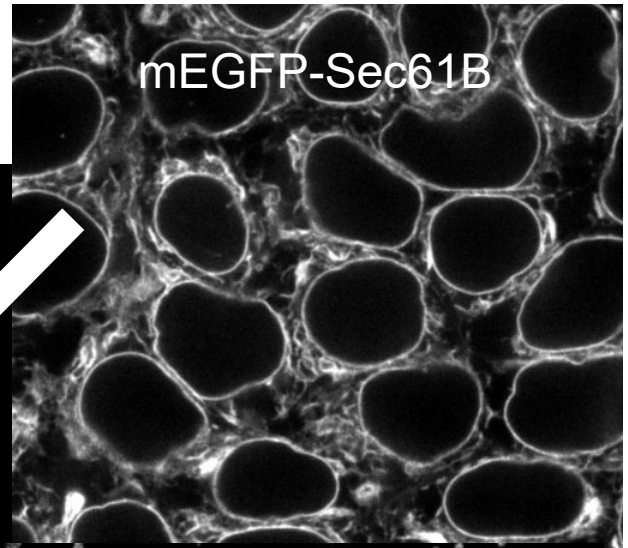
ALLEN INSTITUTE for CELL SCIENCE

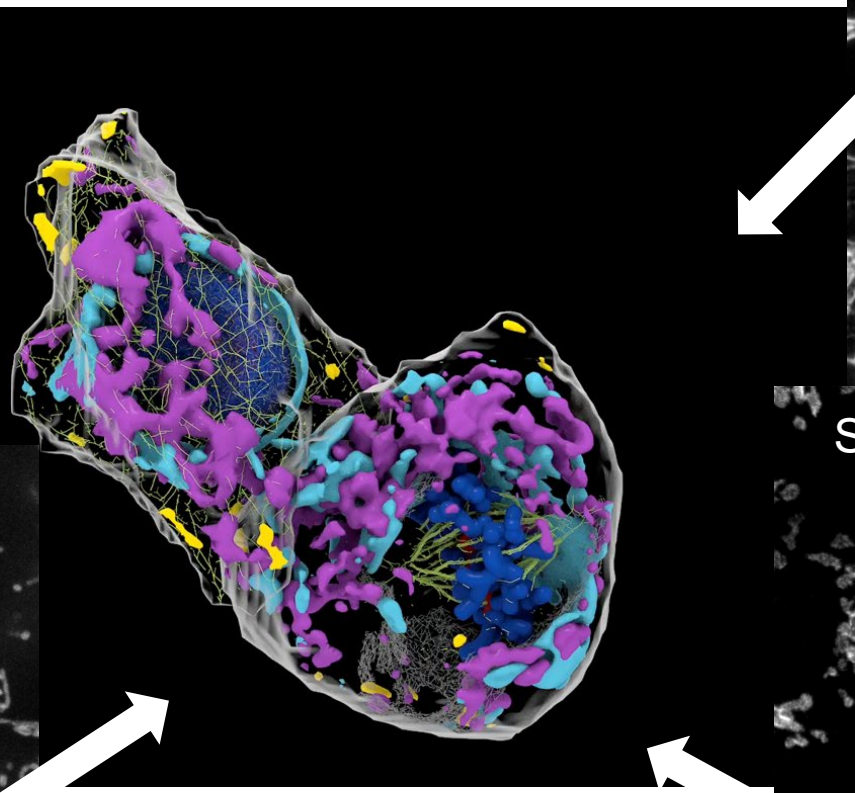
Outline

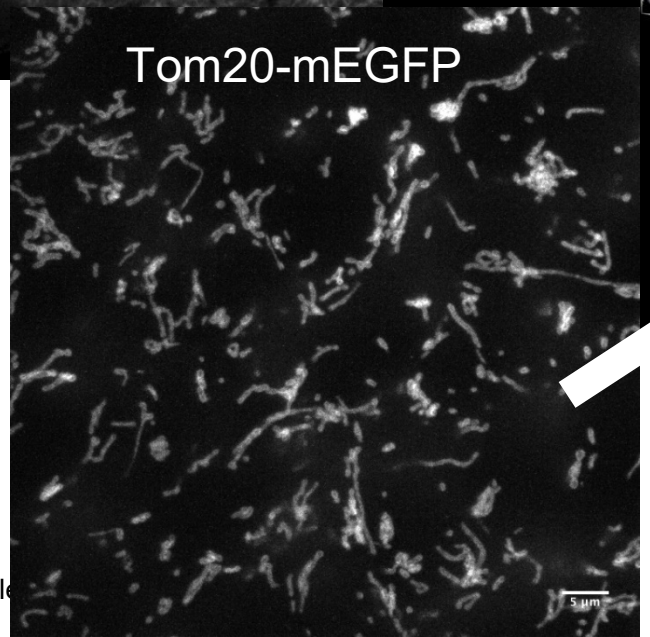
- Intro to the Allen Institute for Cell Science
- **Gene editing design**
  - **Introduction**
  - **crRNA design**
  - **Donor template design**
- Workflows and methods for generating high-quality clonal hiPS cell lines
- In-lab demonstrations
- Roundtable discussion

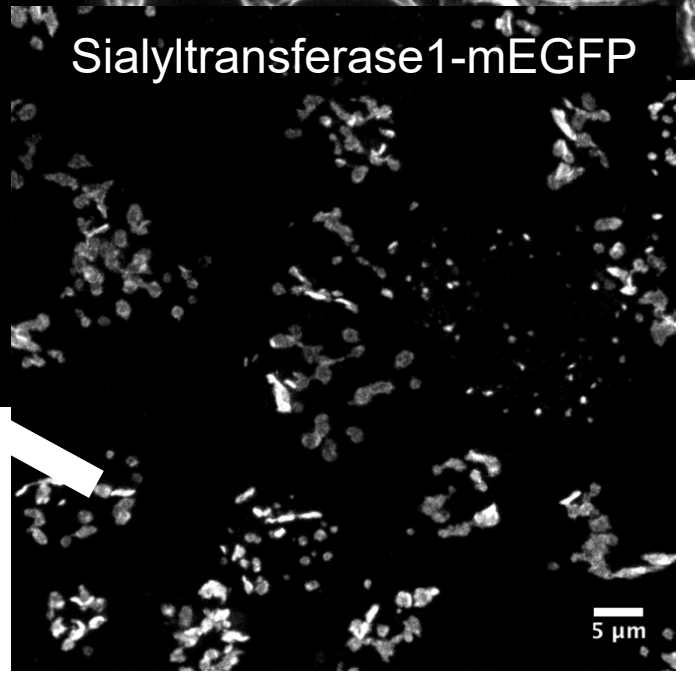
Allen Institute for Cell Science: Creating a collection of FP tagged hiPSC lines representing major structures of the cell











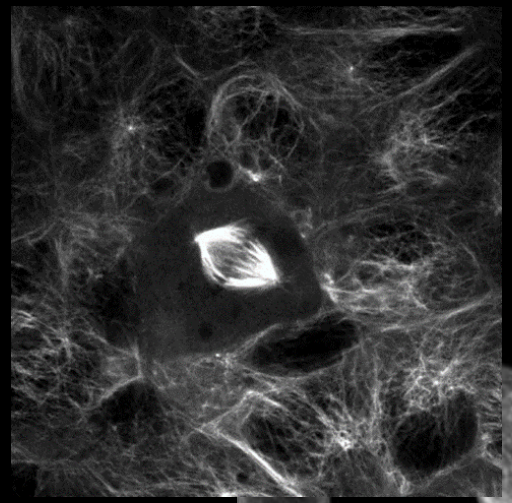
Endogenously tagged clonal cell lines generated

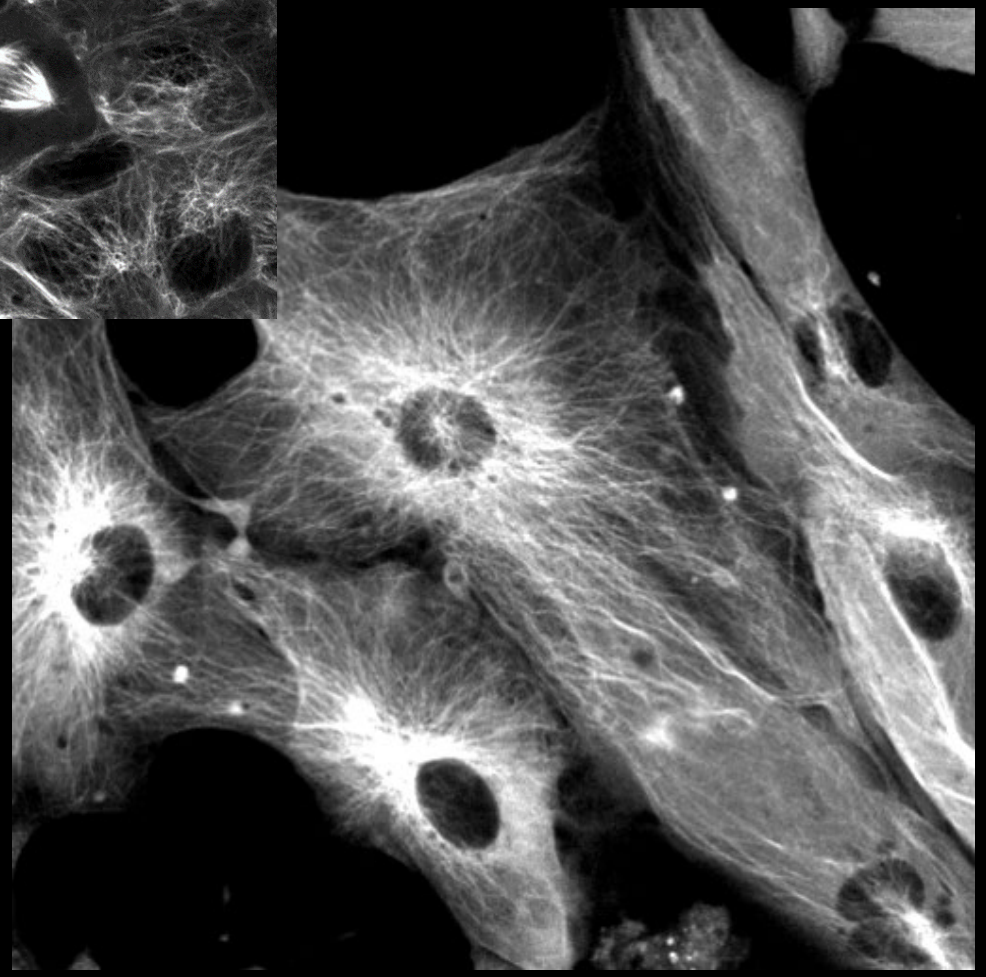
Structures Illuminated

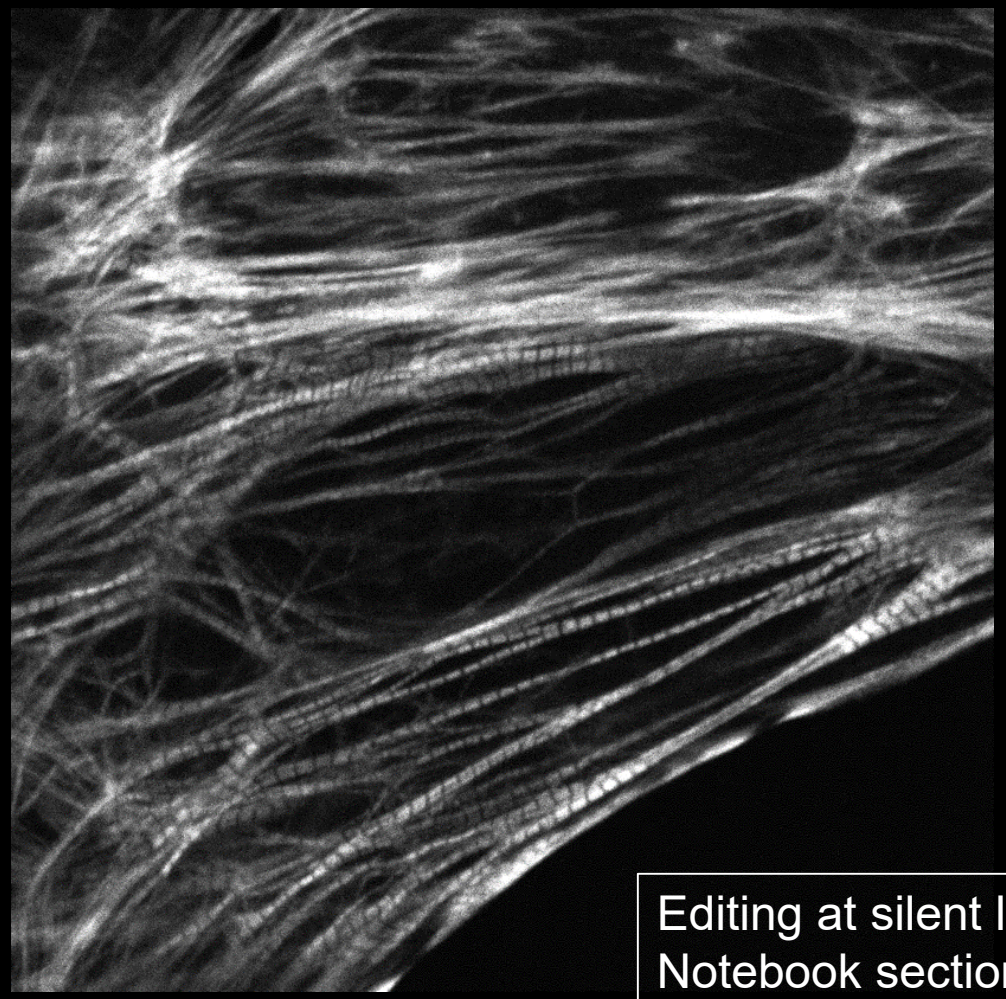
- Nucleus (LMNB1)
- Nucleolus (FBL, NPM1, UBTF)
- Nuclear pores (NUP153)
- Chromatin (SMC1A, HIST1H2BJ)
- Mitochondria (TOMM20)
- Microtubules (TUBA1B)
- Centrosomes (CETN2)
- Endoplasmic Reticulum (SEC61B)
- Golgi (ST6GAL1)
- Lysosomes (LAMP1)
- Peroxisomes (SLC25A17)
- Autophagosomes (MAP1LC3B)
- Endosomes (RAB5A)
- Actins, myosins & filaments (ACTB, ACTN1, ACTN2, MYH10, VIM)
- Cell-cell junctions (TJP1/ZO1, DSP)
- Gap junctions (GJA1)
- Adherens junctions (CTNNB1)
- Matrix Adhesions (PXN)
- Stress Granules and P-Bodies (G3BP1, DCP1A)
- Transcription Factors (SOX2, POU5F1)
- Sub-nuclear structures (CBX1, CTCF, SMC1A)
- Cytoplasm
- Plasma Membrane
- Sarcomere (TTN, MYL7, MYL2, TNNI1, ACTN2)
- Sarcoplasmic reticulum (ATP2A2)

mEGFP-Alpha tubulin as hiPSC

Harnessing the power of hiPSC







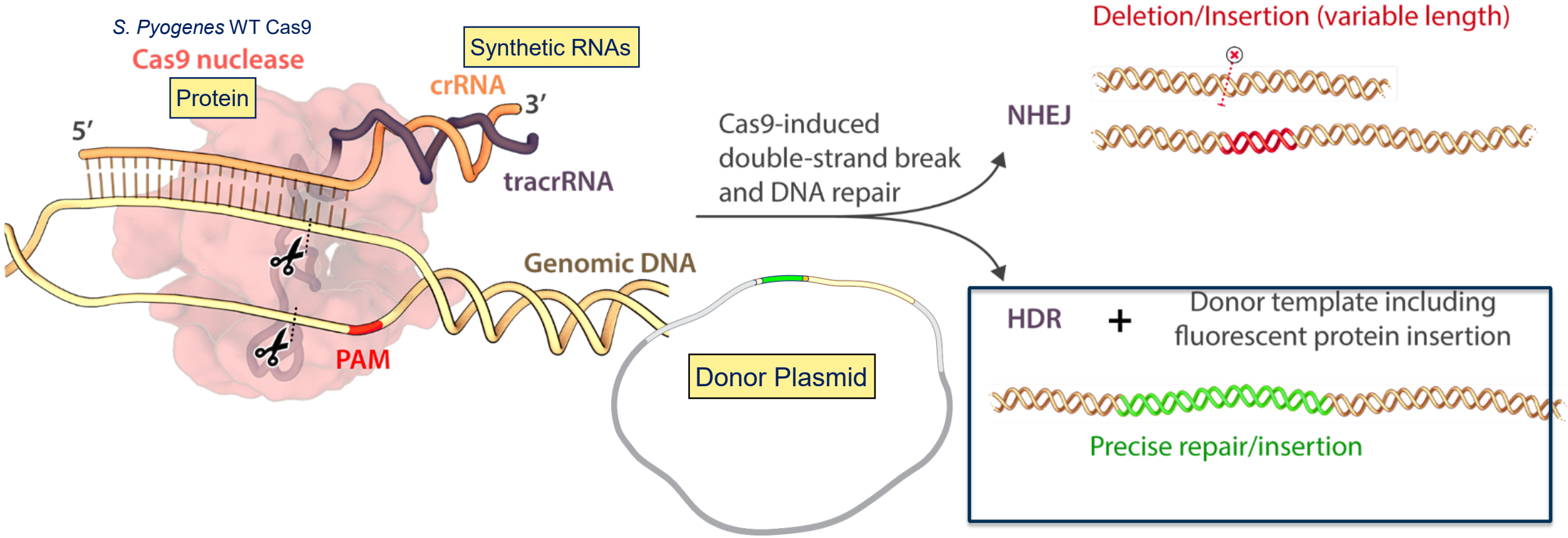
Editing at silent loci:
Notebook section 1C

Cardiomyocytes from mEGFP-Alpha tubulin cell line

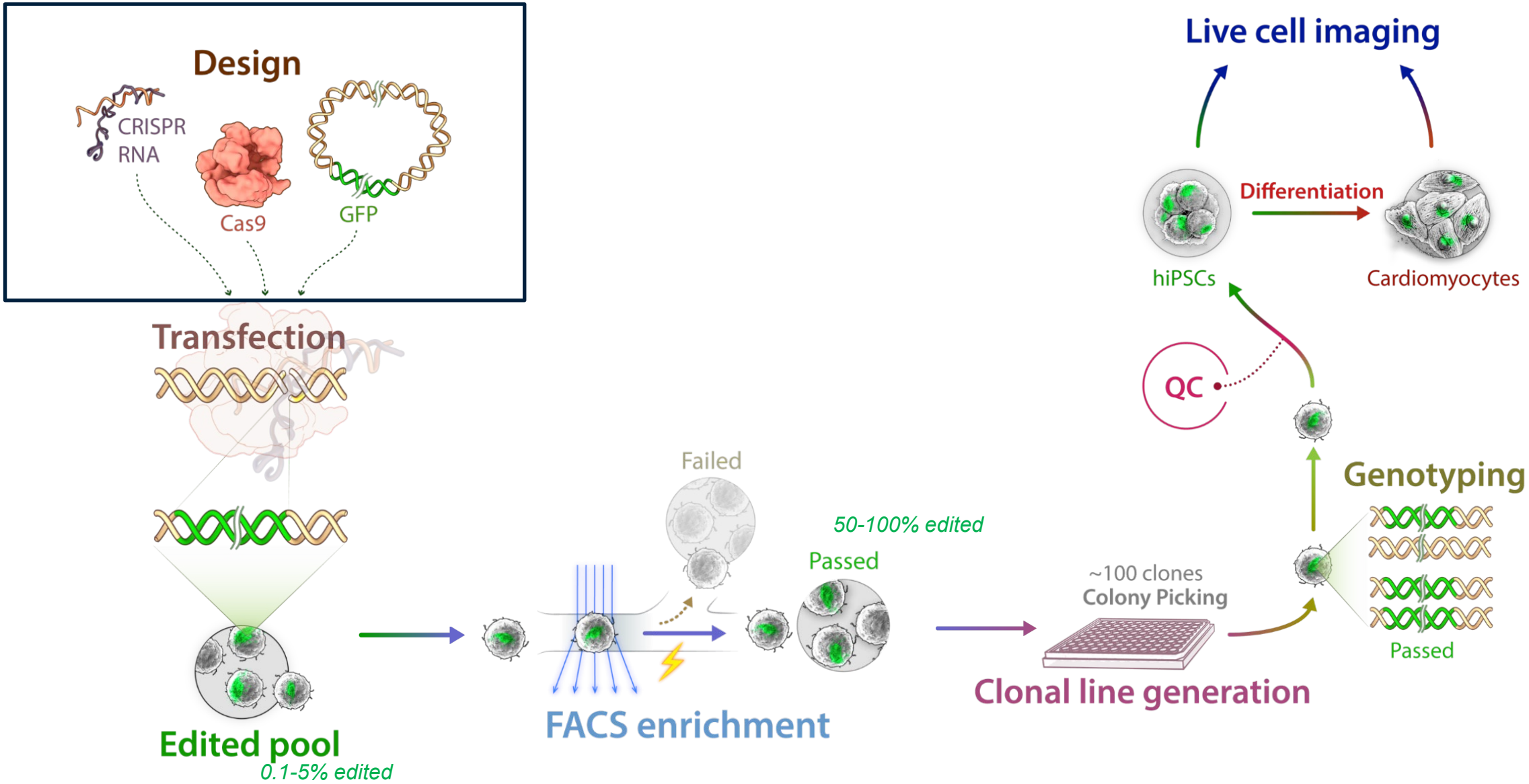
Cardiomyocytes from Troponin 1-mEGFP cell line

Cardiomyocyte movies: 100 ms intervals, 25s total 2D timelapse ~ real time - - Thank you Assay Dev Team!

CRISPR/Cas9 using the RNP Method: A gentle and effective way to introduce specific genomic insertions in stem cells



The Allen Institute for Cell Science Gene Editing Workflow



Outline

- Intro to the Allen Institute for Cell Science
- Gene editing design
  - Introduction
  - **crRNA design**
  - Donor template design
- Workflows and methods for generating high-quality clonal hiPS cell lines
- In-lab demonstrations
- Roundtable discussion

Editing design prerequisites

- **Select target gene, linker, and tag**

Editing design: Selecting a linker

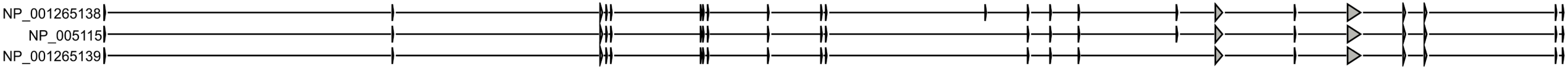
- **Survey literature**
- **Review of linker design: Chen et al. 2013 (Section 1b in workbook)**
- **Short flexible linker: GTSGGS**

Editing design: Selecting a target gene, linker and tag

Gene: NUP153 (Nucleoporin Nup153)

Linker: YSDLE

Tag: mEGFP



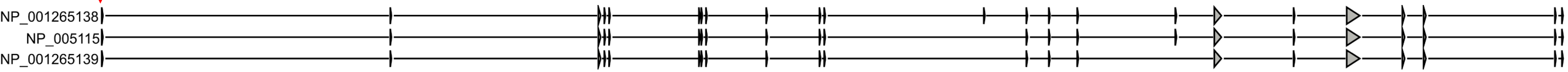
Structure: Nuclear pores

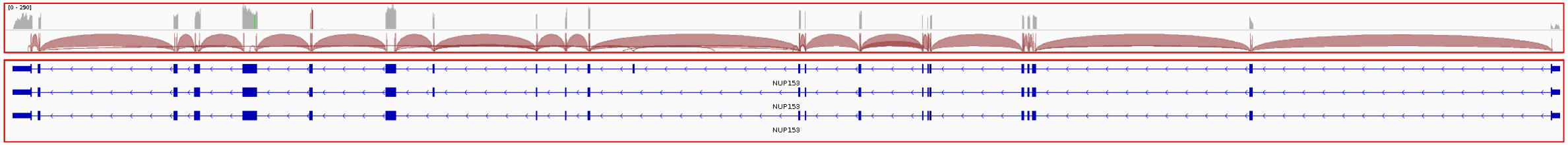
Editing design prerequisites

- Select target gene, linker, and tag
- **Acquire cell line specific data (recommended)**

Editing design: Do you have isoform level information?

Tag here



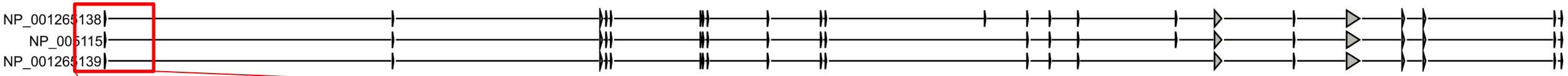


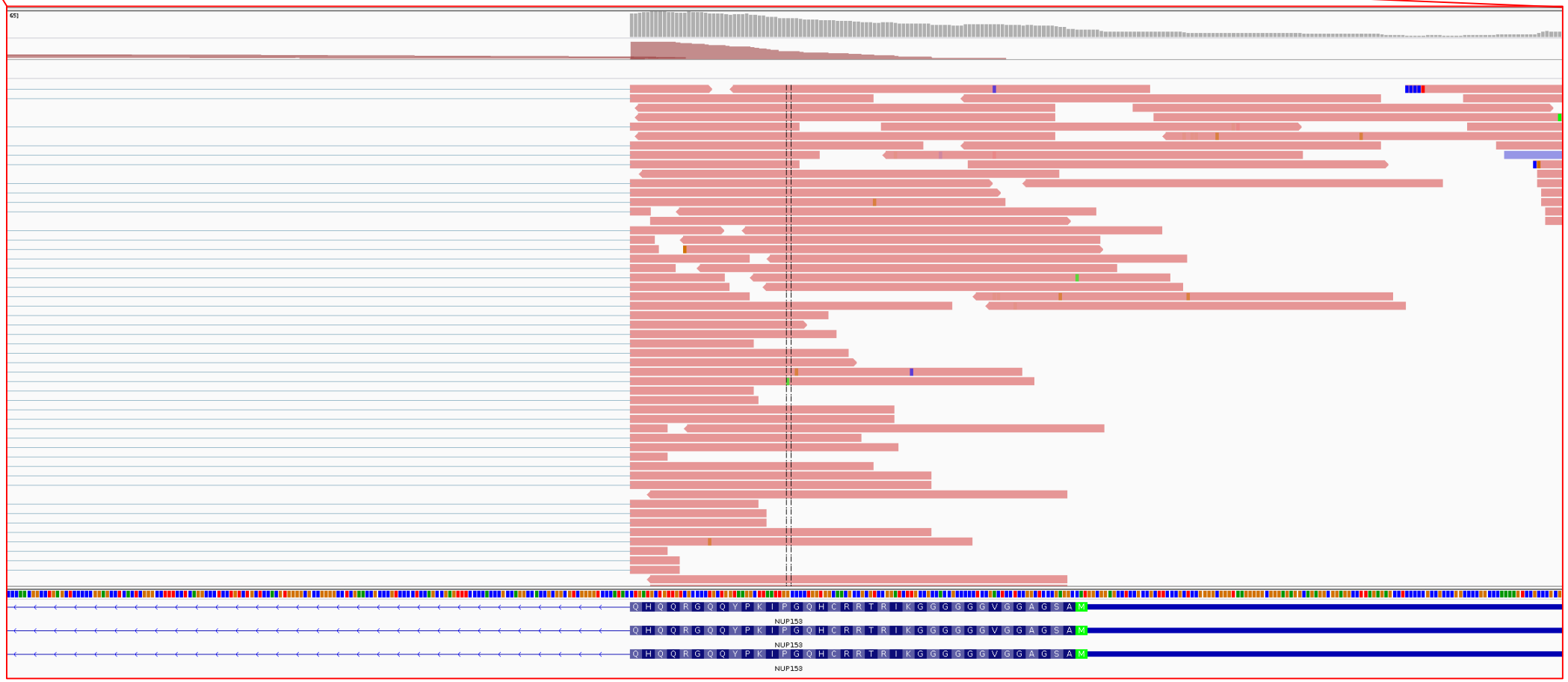
Tag here

Multiple isoforms w/ same N-terminus

Editing design: Do you have isoform level information?

Tag here

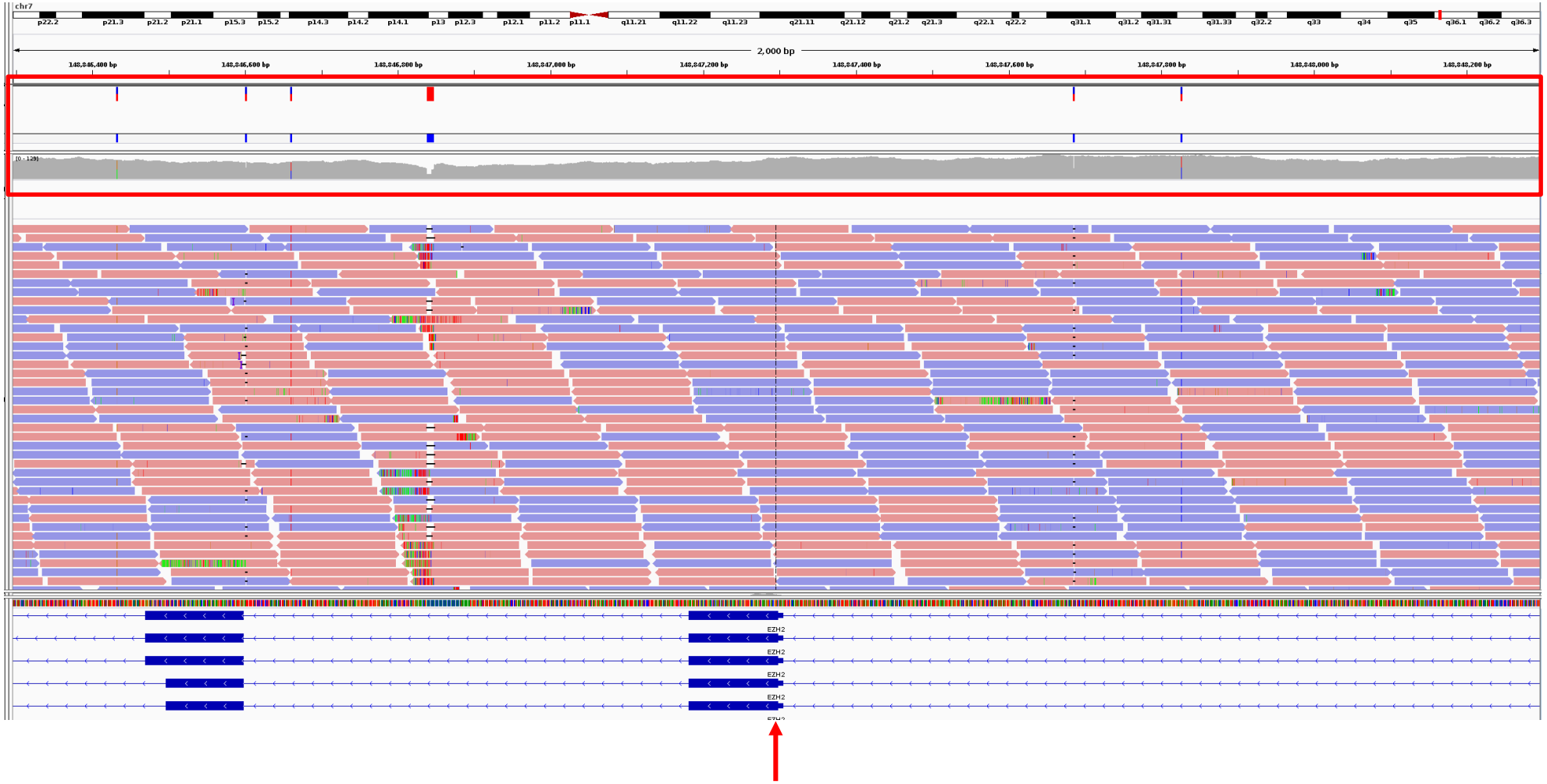




Tag here

Editing design: Do you have cell line variant information?

EZH2



WTC-11 variants in homology arms

Editing design prerequisites

- Select target gene, linker, and tag
- Acquire cell line specific data (recommended)
- **Sequences and tools**
  - **Software to view/annotate/manipulate sequences**
  - Reference sequence, tag sequences, and plasmid backbone
  - crRNA design tool

Editing design prerequisites

- Select target gene, linker, and tag
- Acquire cell line specific data (recommended)
- **Sequences and tools**
  - Software to view/annotate/manipulate sequences
  - **Reference sequence, tag sequences, and plasmid backbone**
  - crRNA design tool

Editing design prerequisites

- Select target gene, linker, and tag
- Acquire cell line specific data (recommended)
- **Sequences and tools**
  - Software to view/annotate/manipulate sequences
  - Reference sequence, tag sequences, and plasmid backbone
  - **crRNA design tool**

Editing design: Download sequences

- Download **NUP153 sequence** from NCBI https://www.ncbi.nlm.nih.gov/gene/ in genbank format

Download NUP153 sequence

Home - Gene - NCBI

https://www.ncbi.nlm.nih.gov/gene/

NCBI Resources How To Sign in to NCBI

Gene | Gene | NUP153 | Search

Help

- Homo sapiens NUP153
- Homo sapiens nuclear pore complex protein Nup153
- Drosophila melanogaster Nup153
- Rattus norvegicus Nup153
- Danio rerio nup153
- Mus musculus Nup153
- Bos taurus NUP153
- Pongo abelii NUP153
- Rattus norvegicus nuclear pore complex protein Nup153
- Danio rerio nuclear pore complex protein Nup153
- Mus musculus nuclear pore complex protein Nup153
- Bos taurus nuclear pore complex protein Nup153
- Pongo abelii nuclear pore complex protein Nup153
- Sus scrofa NUP153
- Pan troglodytes NUP153
- Equus caballus NUP153
- Macaca mulatta NUP153
- Dasypus novemcinctus NUP153
- Oryctolagus cuniculus NUP153
- Sus scrofa nuclear pore complex protein Nup153

lature, Reference
henotype-, and locus-

Using Gene

- Gene Quick Start
- FAQ
- Download/FTP
- RefSeq Mailing List
- Gene News
- Factsheet

Representative queries

Find genes by...	Search text
free text	human muscular dystrophy
chromosome and symbol	(II[chr] OR 2[chr]) AND adh*[sym]

Download NUP153 sequence

Homo sapiens NUP153 - Gene

https://www.ncbi.nlm.nih.gov/gene/?term=Homo+sapiens+NUP153

NCBI Resources How To Sign in to NCBI

Gene | Gene | Homo sapiens NUP153 | Search

Create RSS Create alert Advanced Help

Gene sources
Genomic

Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding

Sequence content
CCDS
Ensembl
RefSeq
RefSeqGene

Status clear
✓ Current

Clear all

Show additional filters

Tabular 20 per page Sort by Relevance Send to:

GENE

Was this helpful?

NUP153 – nucleoporin 153

Homo sapiens (human)

Also known as: HNUP153, N153

GeneID: 9972

RefSeq transcripts (8) RefSeq proteins (8) PubMed (146)

Genome Browser BLAST Download

RefSeq transcripts +

RefSeq proteins +

Search results

Items: 1 to 20 of 132

<< First < Prev Page 1 of 7 Next > Last >>

Hide sidebar >>

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]
Homo sapiens (127)
Human gammaherpesvirus 4 (2)
Caenorhabditis remanei (1)
[Candida] auris (1)
[Candida] pseudohaemulonis (1)

Find related data

Database: Select

Find items

Search details

```
(("Homo sapiens"[Organism] OR Homo
sapiens[All Fields]) AND NUP153[All
Fields]) AND alive[prop]
```

Search

See more...

https://www.ncbi.nlm.nih.gov/gene/9972

Download NUP153 sequence

NUP153 nucleoporin 153 [Hom

https://www.ncbi.nlm.nih.gov/gene/9972

Search

NCBI Resources How To Sign in to NCBI

Gene Gene Search

Advanced Help

Full Report Send to: Hide sidebar >>

NUP153 nucleoporin 153 [*Homo sapiens* (human)]

Gene ID: 9972, updated on 13-Feb-2019

Summary

Official Symbol	NUP153 provided by HGNC
Official Full Name	nucleoporin 153 provided by HGNC
Primary source	HGNC:HGNC:8062
See related	Ensembl:ENSG00000124789 MIM:603948
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	N153; HNUP153
Summary	Nuclear pore complexes regulate the transport of macromolecules between the nucleus and cytoplasm. They are composed of at least 100 different polypeptide subunits, many of which belong to the nucleoporin family. Nucleoporins are glycoproteins found in nuclear pores and contain characteristic pentapeptide XFXFG repeats as well as O-linked N-acetylglucosamine residues oriented towards the cytoplasm. The protein encoded by this gene has three distinct domains: a N-terminal region containing a pore targeting and an RNA-binding domain domain, a central region containing multiple zinc finger motifs, and a C-terminal region containing multiple XFXFG repeats. Alternative splicing results in multiple transcript variants of this gene. [provided by RefSeq, May 2013]
Expression	Ubiquitous expression in testis (RPKM 19.5), bone marrow (RPKM 18.1) and 25 other tissues See more
Orthologs	mouse all

Table of contents

Download NUP153 sequence

NUP153 nucleoporin 153 [Hom

https://www.ncbi.nlm.nih.gov/gene/9972

Search

Annotation release	Status	Assembly	Chr	Location
109	current	GRCh38.p12 (GCF_000001405.38)	6	NC_000006.12 (17615035..17706834, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	6	NC_000006.11 (17615266..17707065, complement)

Chromosome 6 - NC_000006.12

[17585894 ▶ [17987623 ▶

LOC102724591 ◀ LOC105374952 ▶ KIF13A ◀

FAM8A1 ▶ RNA5SP204 ◀ LOC105374953 ▶

RNU6-190P ▶

NUP153 ◀

Genomic regions, transcripts, and products

Go to reference sequence details

Genomic Sequence: NC_000006.12 Chromosome 6 Reference GRCh38.p12 Primary Assembly

Go to nucleotide: Graphics FASTA GenBank

NC_000006.12 | Find: | Tools | Tracks

17,720 K | 17,710 K | 17,700 K | 17,690 K | 17,680 K | 17,670 K | 17,660 K | 17,650 K | 17,640 K | 17,630 K | 17,620 K | 17,610 K

Genes, NCBI Homo sapiens Annotation Release 109, 2018-03-27

NUP153

NM_001278210.1 NM_005124.3 NM_001278209.1 XM_017011594.2 XM_005249507.3 XM_006715290.3 XM_011515028.3 XM_006715291.4

NP_001265139.1 NP_005115.2 NP_001265138.1 XP_016867083.1 XP_005249564.1 XP_006715353.1 XP_011513330.1 XP_006715354.1

LOC105374952 NR_134616.1 NR_134618.1

RNU6-190P

FAM8A1 NP_057339.1

Biological regions, aggregate, NCBI NCBI Homo sapiens Annotatio…

Warning: No track data found in this range

Genes, Ensembl release 95

1000 Genomes Browser (GRCh37.p13)

Ensembl

UCSC

Related information

Order cDNA clone

3D structures

BioAssay by Target (List)

BioAssay by Target (Summary)

BioAssay, by Gene target

BioAssays, RNAi Target, Tested

BioProjects

BioSystems

CCDS

ClinVar

Conserved Domains

dbVar

Full text in PMC

Full text in PMC_nucleotide

Functional Class

Gene neighbors

Genes with a similar H3K4me3 profile

Genome

Download NUP153 sequence

Homo sapiens chromosome 6,

https://www.ncbi.nlm.nih.gov/nuccore/NC_000006.12?report=genbank&from

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Advanced Help

The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

GenBank Send to:

Homo sapiens chromosome 6, GRCh38.p12 Primary Assembly

NCBI Reference Sequence: NC_000006.12

FASTA Graphics

```
LOCUS       NC_000006               91800 bp    DNA     linear   CON 26-MAR-2018
DEFINITION  Homo sapiens chromosome 6, GRCh38.p12 Primary Assembly.
ACCESSION   NC_000006 REGION: complement(17615035..17706834)
VERSION     NC_000006.12
DBLINK      BioProject: PRJNA168
            Assembly: GCF_000001405.38
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 91800)
  CONSRTM   International Human Genome Sequencing Consortium
  TITLE     Finishing the euchromatic sequence of the human genome
  JOURNAL   Nature 431 (7011), 931-945 (2004)
   PUBMED   15496913
REFERENCE   2  (bases 1 to 91800)
  AUTHORS   Mungall,A.J., Palmer,S.A., Sims,S.K., Edwards,C.A., Ashurst,J.L.,
            Wilming,L., Jones,M.C., Horton,R., Hunt,S.E., Scott,C.E.,
            Gilbert,J.G., Clamp,M.E., Bethel,G., Milne,S., Ainscough,R.,
```

Change region shown

- ○ Whole sequence (abbreviated view)
- ◉ Selected region

from: 17615035 to: 17706834

Update View

Customize view

- ○ Abbreviated view
- ◉ Customize

Basic Features

- ○ Default features
- ◉ Gene, RNA, and CDS features only

Features added by NCBI

- ☐ 1 SNP

Display options

- ☑ Show sequence
- ☑ Show reverse complement

Update View

Analyze this sequence

Download NUP153 sequence

Homo sapiens chromosome 6,

https://www.ncbi.nlm.nih.gov/nuccore/NC_000006.12?report=genbank&from

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Advanced Help

The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

GenBank Send to:

Showing 95.80kb region from base 17613035 to 17708834.

Homo sapiens chromosome 6, GRCh38.p12 Primary Assembly

NCBI Reference Sequence: NC_000006.12

FASTA Graphics

```
LOCUS       NC_000006               95800 bp    DNA     linear   CON 26-MAR-2018
DEFINITION  Homo sapiens chromosome 6, GRCh38.p12 Primary Assembly.
ACCESSION   NC_000006 REGION: complement(17613035..17708834)
VERSION     NC_000006.12
DBLINK      BioProject: PRJNA168
            Assembly: GCF_000001405.38
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 95800)
  CONSRTM   International Human Genome Sequencing Consortium
  TITLE     Finishing the euchromatic sequence of the human genome
  JOURNAL   Nature 431 (7011), 931-945 (2004)
   PUBMED   15496913
REFERENCE   2  (bases 1 to 95800)
  AUTHORS   Mungall,A.J., Palmer,S.A., Sims,S.K., Edwards,C.A., Ashurst,J.L.,
```

Change region shown

- ○ Whole sequence (abbreviated view)
- ◉ Selected region

from: 17613035 to: 17708834

Update View

Customize view

- ○ Abbreviated view
- ◉ Customize

Basic Features

- ○ Default features
- ◉ Gene, RNA, and CDS features only

Features added by NCBI

- ☐ 1 SNP

Display options

- ☑ Show sequence
- ☑ Show reverse complement

Update View

Analyze this sequence

Download NUP153 sequence

Homo sapiens chromosome 6,

https://www.ncbi.nlm.nih.gov/nuccore/NC_000006.12?report=genbank&from

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search

Advanced Help

The Nucleotide database will include EST and GSS sequences in early 2019. Read more.

GenBank Send to:

Showing 95.80kb region from base 17613035 to 17708834.

Homo sapiens chromosome 6, GRCh38.p12 Primary Assembly

NCBI Reference Sequence: NC_000006.12

FASTA Graphics

```
LOCUS       NC_000006               95800 bp    DNA     linear   CON 26-MAR-2018
DEFINITION  Homo sapiens chromosome 6, GRCh38.p12 Primary Assembly.
ACCESSION   NC_000006 REGION: complement(17613035..17708834)
VERSION     NC_000006.12
DBLINK      BioProject: PRJNA168
            Assembly: GCF_000001405.38
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 95800)
  CONSRTM   International Human Genome Sequencing Consortium
  TITLE     Finishing the euchromatic sequence of the human genome
  JOURNAL   Nature 431 (7011), 931-945 (2004)
   PUBMED   15496913
REFERENCE   2  (bases 1 to 95800)
  AUTHORS   Mungall,A.J., Palmer,S.A., Sims,S.K., Edwards,C.A., Ashurst,J.L.,
```

Complete Record
Coding Sequences
Gene Features

Choose Destination
File Clipboard
Collections Analysis Tool

Download 1 item.
Format
GenBank
Show GI
Create File

n shown
nce (abbreviated view)
on
to: 17708834
Update View

ew
iew
es
tures
, and CDS features only
ded by NCBI

Display options
Show sequence
Show reverse complement
Update View

Analyze this sequence

Download NUP153 sequence (Ensembl)

Homo sapiens chromosome 6 | Gene: NUP153 (ENSG00000124789

https://uswest.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG

Search

- Paralogues
- Ensembl protein families
- Ontologies
  - GO: Biological process
  - GO: Molecular function
  - GO: Cellular component
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Transcripts

Show transcript table

Summary

Name	NUP153 (HGNC Symbol)
CCDS	This gene is a member of the Human CCDS set: CCDS4541.1, CCDS64359.1, CCDS75407.1
UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: P49790
RefSeq	Overlapping RefSeq Gene ID 9972 matches and has similar biotype of protein_coding
Ensembl version	ENSG00000124789.11
Other assemblies	This gene maps to 17,615,266-17,707,065 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000124789
Gene type	Protein coding
Annotation method	Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Drag/Select:

111.80 kb — Forward strand

17.62Mb | 17.64Mb | 17.66Mb | 17.68Mb | 17.70Mb

Genes (Comprehensive set...): FAM8A1-201 > protein coding; RNU6-190P-201 > snRNA; AL138724.1-201 < antisense

Contigs: AL138824.19 > | AL157776.15 > | AL138724.12 >

Genes (Comprehensive set...): < NUP153-202 protein coding; < NUP153-203 protein coding; < NUP153-201 protein coding

Regulatory Build

https://uswest.ensembl.org/Homo_sapiens/Export/Configure/Gene?db=core;g=ENSG00000124789;r=6:17615035-17706834

Download NUP153 sequence (Ensembl)

Homo sapiens chromosome 6
Gene: NUP153 (ENSG00000124789
https://uswest.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG
Search

Paralogues
Ensembl protein families
Transcripts
Show transcript table

Export data

Export Configuration - Feature List

Tip

For sequence export, please go to the relevant sequence page (see lefthand menu) and use the new "Download sequence" button

Gene to export: ENSG00000124789 (NUP153)

Output: FASTA sequence *

Strand:

5' Flanking sequence (upstream): m of 1000000)

3' Flanking sequence (downstream): m of 1000000)

Fields marked * are required

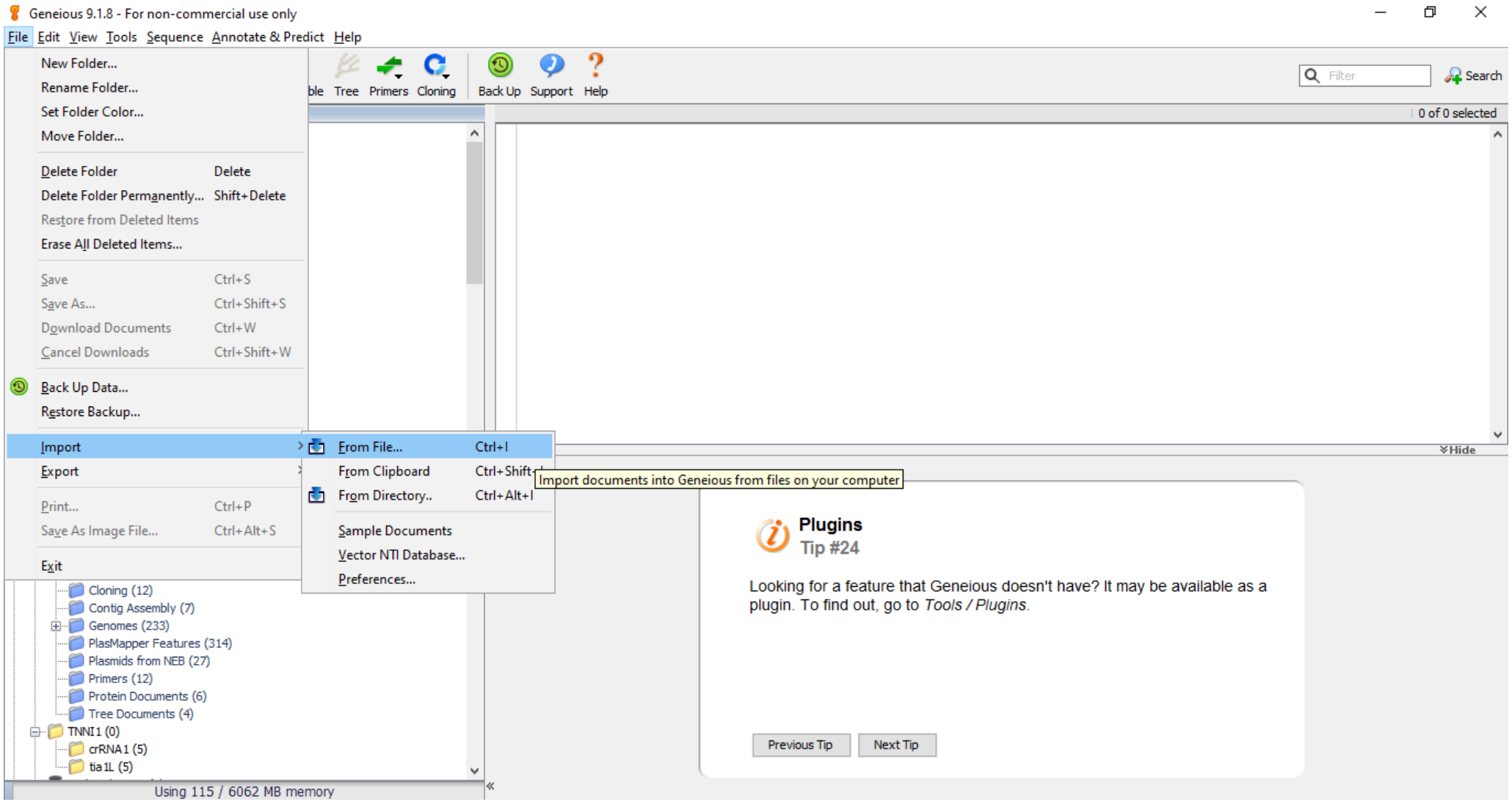
Options for FASTA sequence

Genomic:

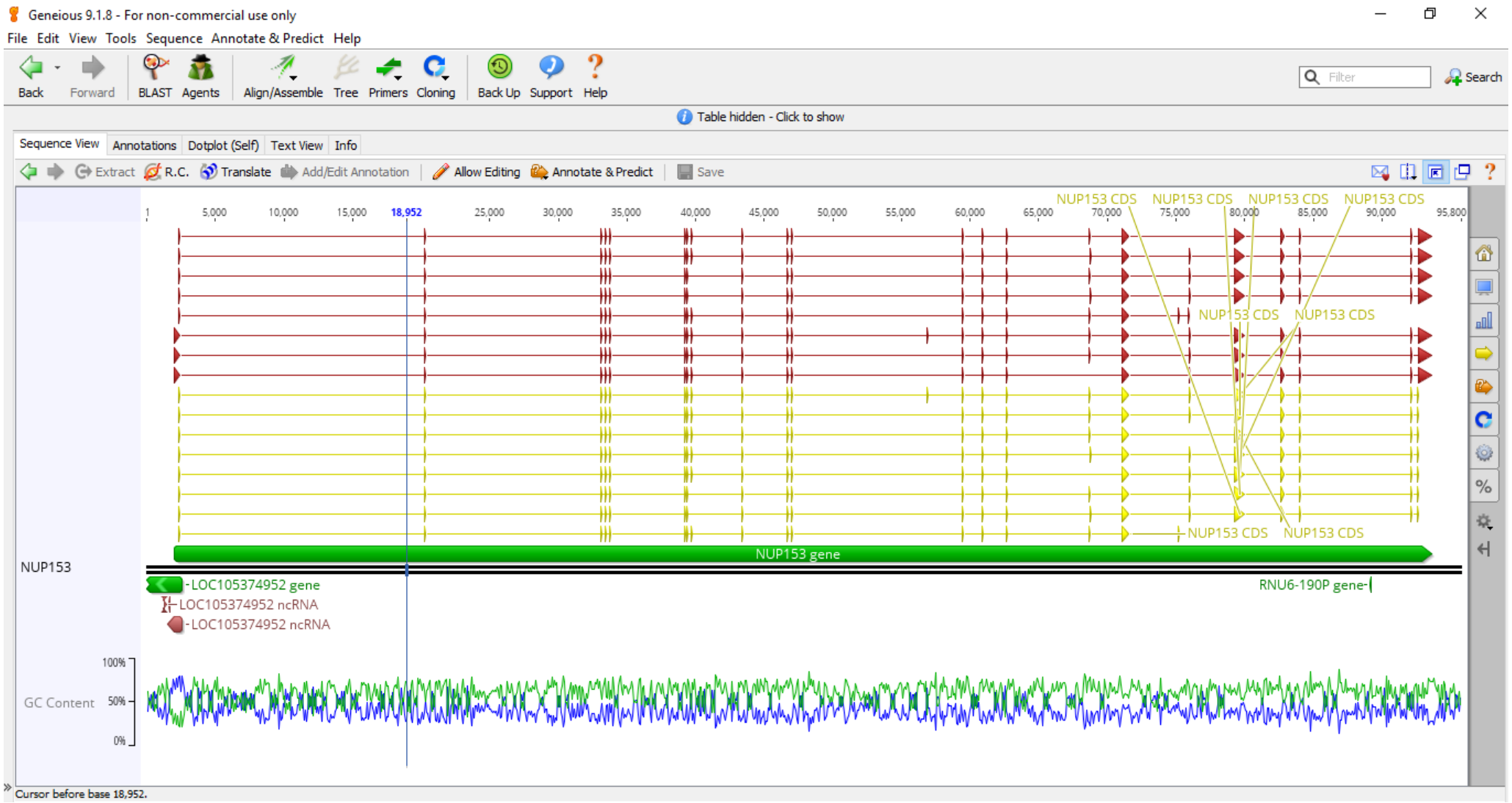
- **Bed Format**
  - BED Format
- **FASTA sequence**
  - FASTA sequence
- **Feature File**
  - CSV (Comma separated values)
  - Tab separated values
  - GTF (Gene Transfer Format)
  - GFF (Generic Feature Format)
  - GFF3 (Generic Feature Format Version 3)
- **Flat File**
  - EMBL
  - GenBank
- **PIP (%age identity plot)**
  - Pipmaker / zPicture format
  - Vista Format

Regulatory Build

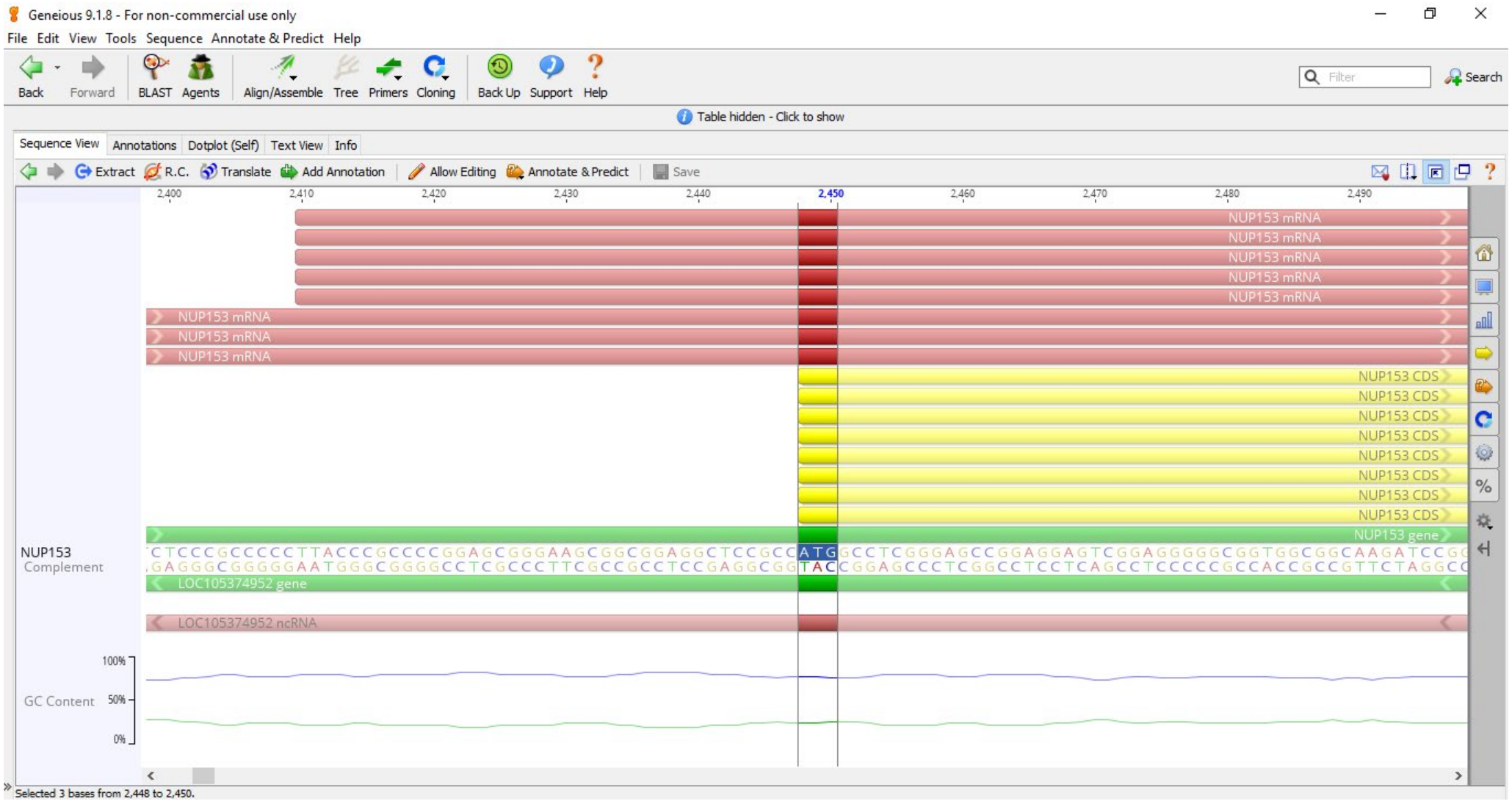
Editing design: Import sequence in Geneious



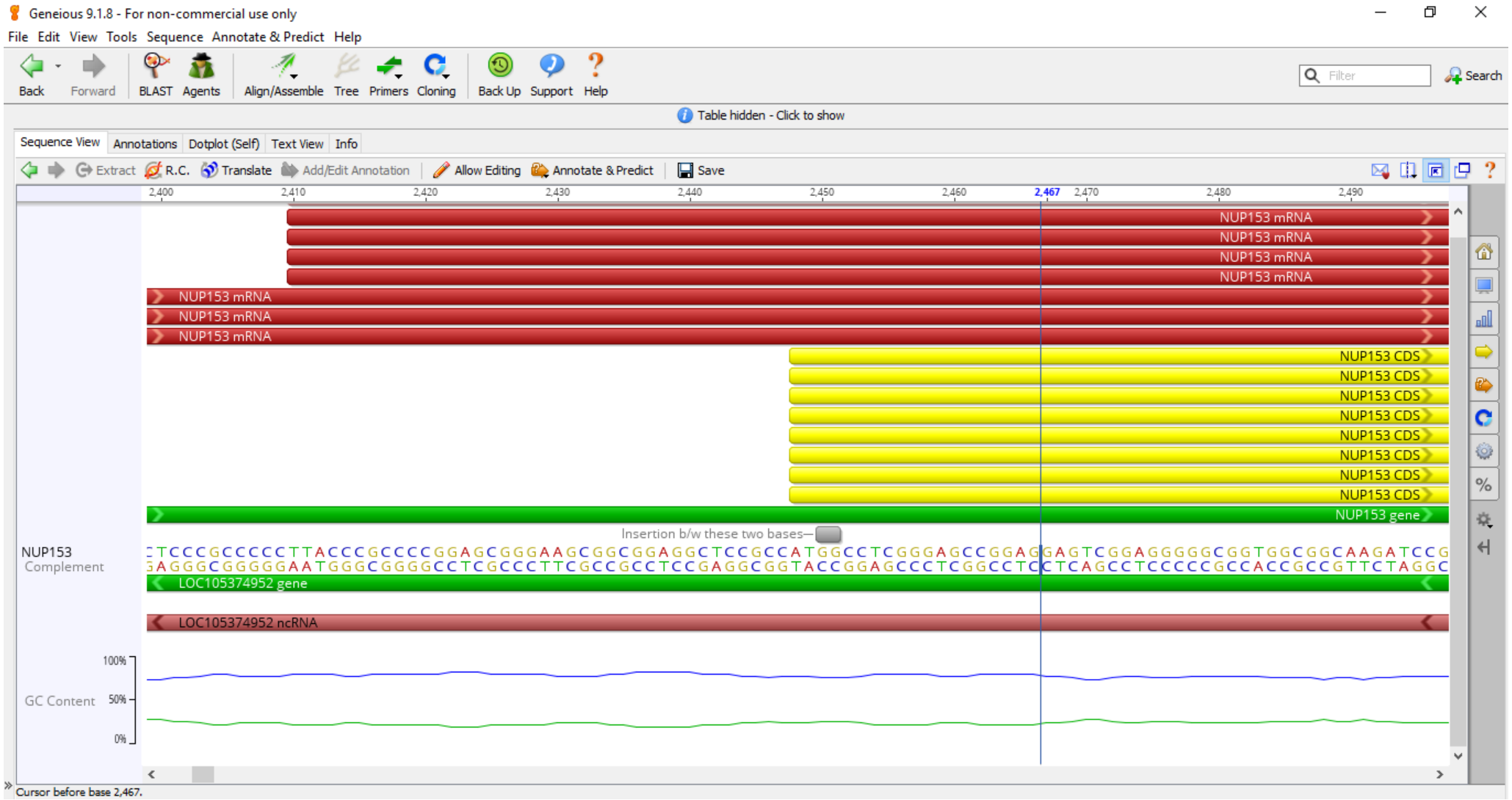
Editing design: Zoom to terminus you want to tag



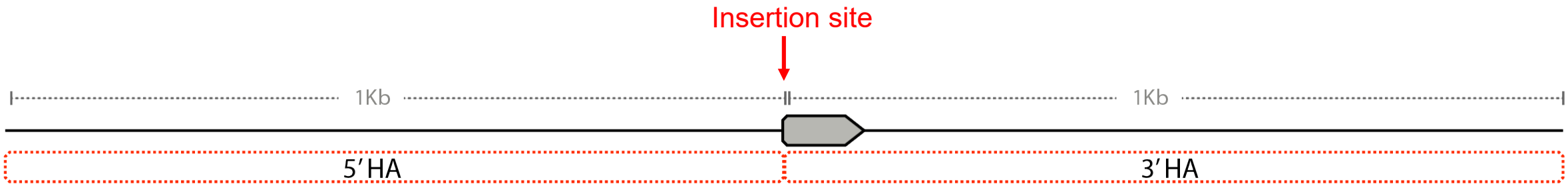
Editing design: Zoom to terminus you want to tag



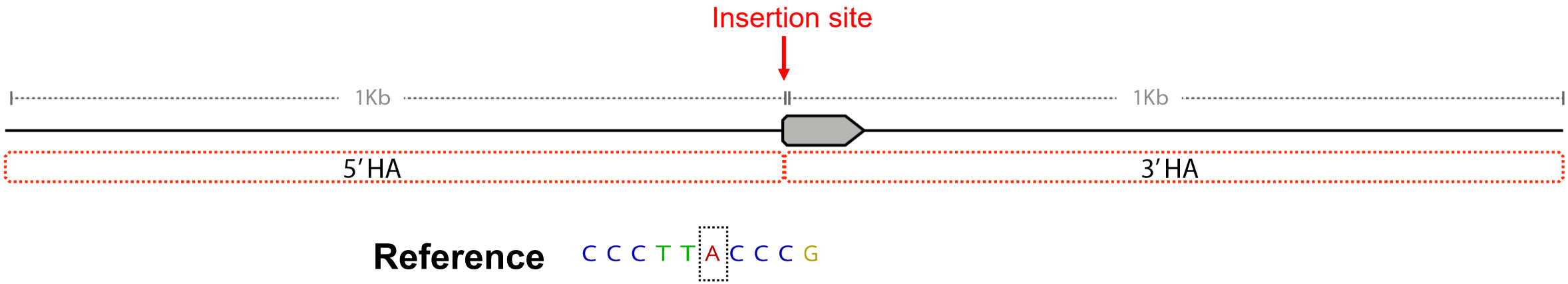
Editing design: Zoom to terminus you want to tag



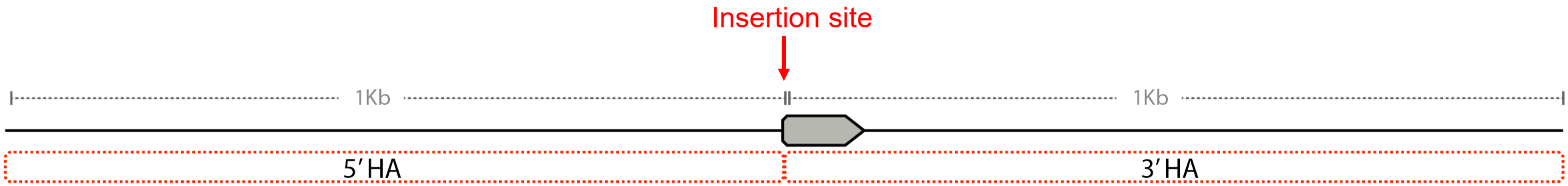
Editing design: Include cell line specific variants



Editing design: Include cell line specific variants



Editing design: Include cell line specific variants

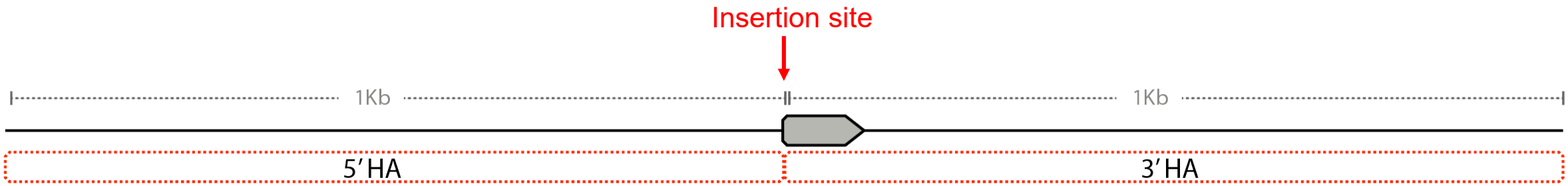


Reference C C C T T A C C C G

Homozygous variants
C C C T T G C C C G
C C C T T G C C C G

Include variant in template

Editing design: Include cell line specific variants



Reference C C C T T A C C C G

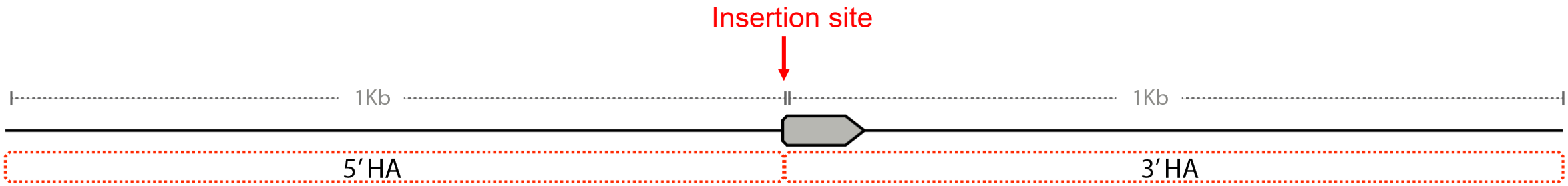
Homozygous variants
C C C T T G C C C G
C C C T T G C C C G

Include variant in template

Heterozygous variants
C C C T T A C C C G
C C C T T G C C C G

Optional inclusion of variant in template

Editing design: Include cell line specific variants



Reference C C C T T A C C C G

Homozygous variants
C C C T T G C C C G
C C C T T G C C C G

Include variant in template

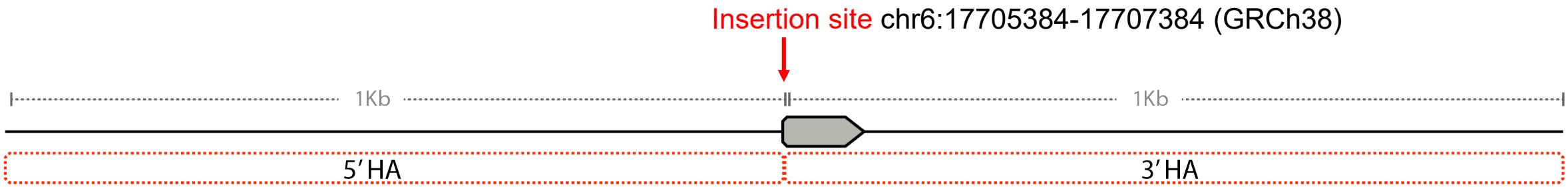
Heterozygous variants
C C C T T A C C C G
C C C T T G C C C G

Optional inclusion of variant in template

Heterozygous variants
C C C T T C C C C G
C C C T T T C C C G

Select one variant to include in template

Editing design: Include cell line specific variants



Reference C C C T T A C C C G

Homozygous variants
C C C T T G C C C G
C C C T T G C C C G

Include variant in template

Heterozygous variants
C C C T T A C C C G
C C C T T G C C C G

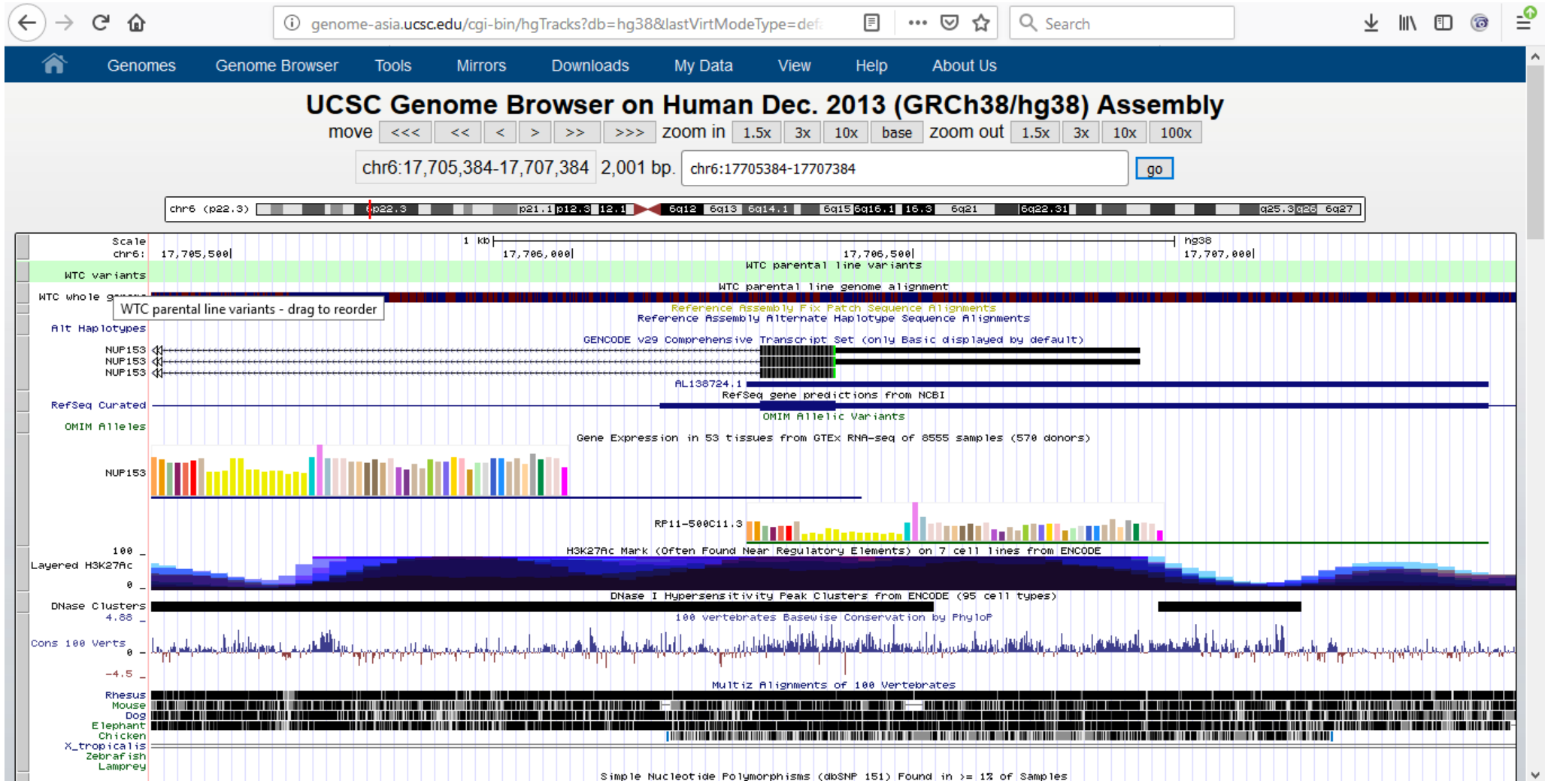
Optional inclusion of variant in template

Heterozygous variants
C C C T T C C C C G
C C C T T T C C C G

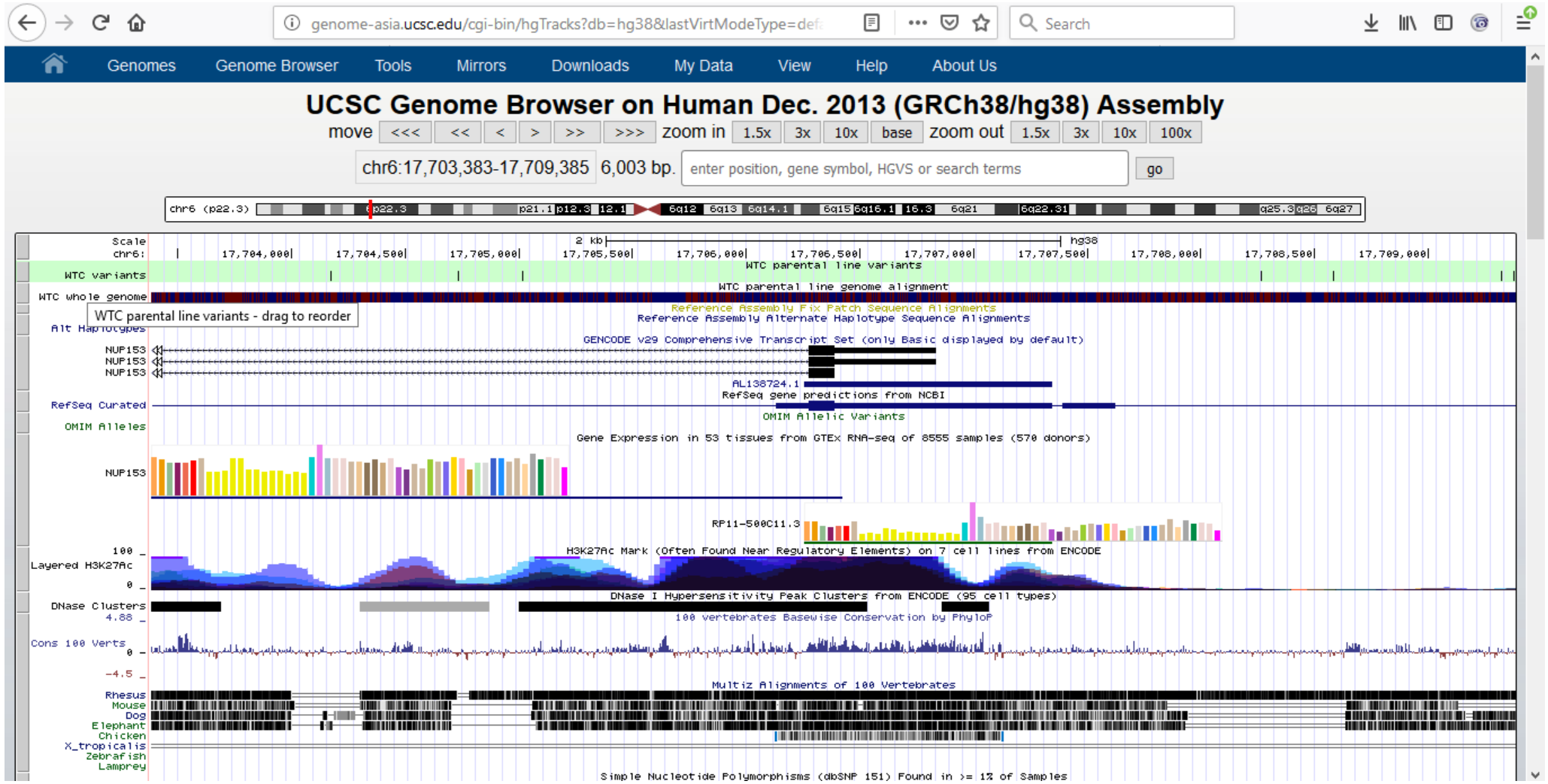
Select one variant to include in template

WTC-11 whole genome variants: https://www.allencell.org/genomics.html

Editing design: Include cell line specific variants

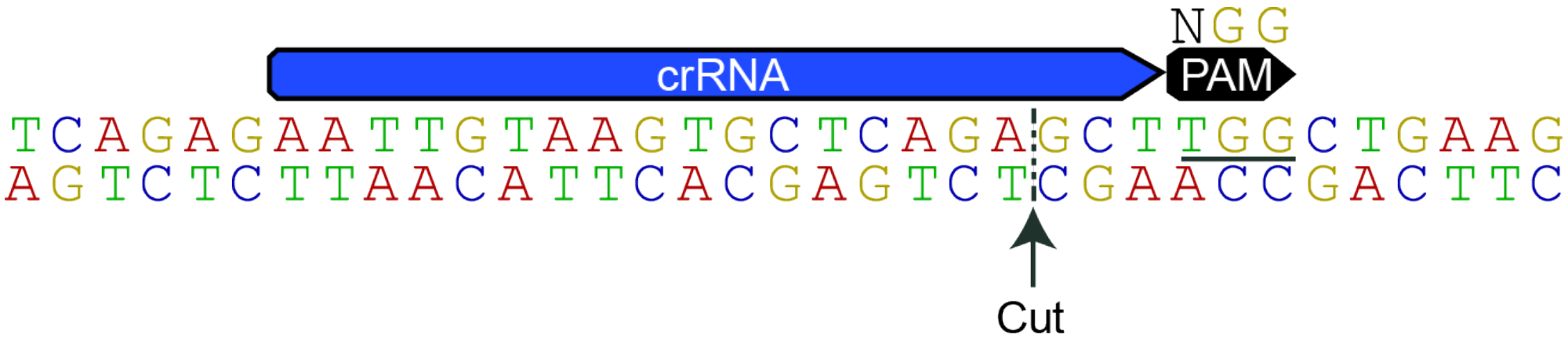


Editing design: Include cell line specific variants



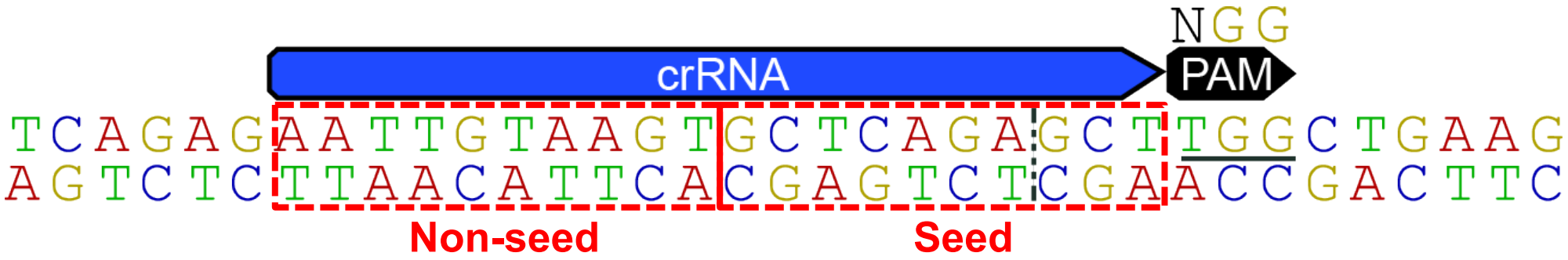
Editing design: Find CRISPR guides

Guide is 20nt sequence adjacent to NGG PAM (SpCas9)



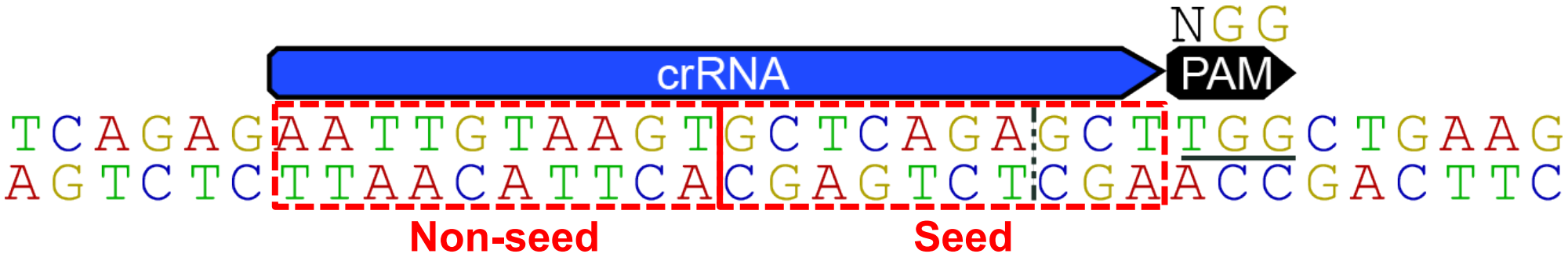
Editing design: Find CRISPR guides

Guide is 20nt sequence adjacent to NGG PAM (SpCas9)



Editing design: Selecting CRISPR guides

- Select guides by balancing:
  - Off-targets



Editing design: Selecting CRISPR guides (Avoiding off-targets)

crRNA PAM

Avoid

On-target AATTGTAAGTGCTCAGAGCTTGG

Off-target AATTGTAAGTGCTCAGAGCTTGG

Avoid

On-target AATTGTAAGTGCTCAGAGCTTGG

Off-target AATTGTAAGTGCTCAGAGCTTAG *

Avoid guides with perfect off-target matches (adjacent to NGG or NAG PAM)

Editing design: Selecting CRISPR guides (Avoiding off-targets)

crRNA PAM

Avoid

On-target AATTGTAAGTGCTCAGAGCTTGG

Off-target AATTGTAAGTGCTCAGAGCTTGG

Avoid

On-target AATTGTAAGTGCTCAGAGCTTGG

Off-target AATTGTAAGTGCTCAGAGCTTAG

Avoid guides with perfect off-target matches (adjacent to NGG or NAG PAM)

On-target AATTGTAAGTGCTCAGAGCTTGG

Off-target AATGGTTAGTGCTCAGAGCTTGG

Better

On-target AATTGTAAGTGCTCAGAGCTTGG

Off-target AATTGTAAGTGCTCAGTGCCTGG

Prioritize guides that have no off-targets with perfectly matching seed regions

Editing design: Selecting CRISPR guides (Avoiding off-targets)

crRNA PAM

Avoid

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target AATTGTAAGTGCTCAGAGCTTGG

Avoid

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target AATTGTAAGTGCTCAGAGCTTAG

Avoid guides with perfect off-target matches (adjacent to NGG or NAG PAM)

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target AATGGTTAGTGCTCAGAGCTTGG

Better

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target AATTGTAAGTGCTCAGTGCCTGG

Prioritize guides that have no off-targets with perfectly matching seed regions

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target GATGGTAAATGCTCAGAGCTTGG

Better

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target AATGGTAAATGCTCAGTGCTTGG

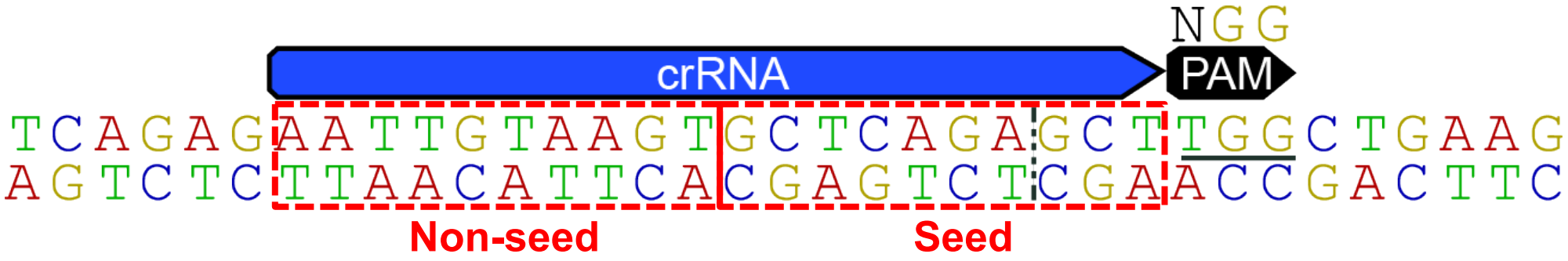
Better

On-target AATTGTAAGTGCTCAGAGCTTGG
Off-target AATTGTAAATGCTCAGTGCCTGG

Prioritize guides with off-targets w/ more than 2 mismatches

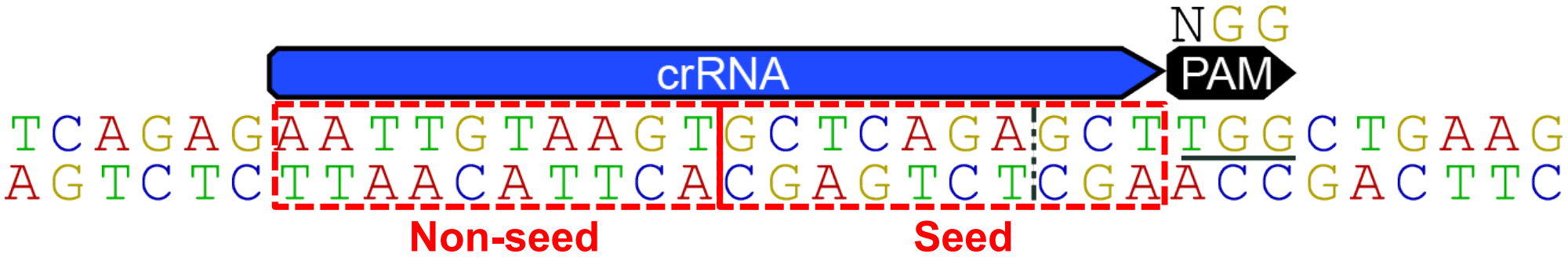
Editing design: Selecting CRISPR guides

- Select guides by balancing:
  - Off-targets
  - Cleavage efficiency



Editing design: Selecting CRISPR guides

- Select guides by balancing:
  - Off-targets
  - Cleavage efficiency
  - Distance from insertion site

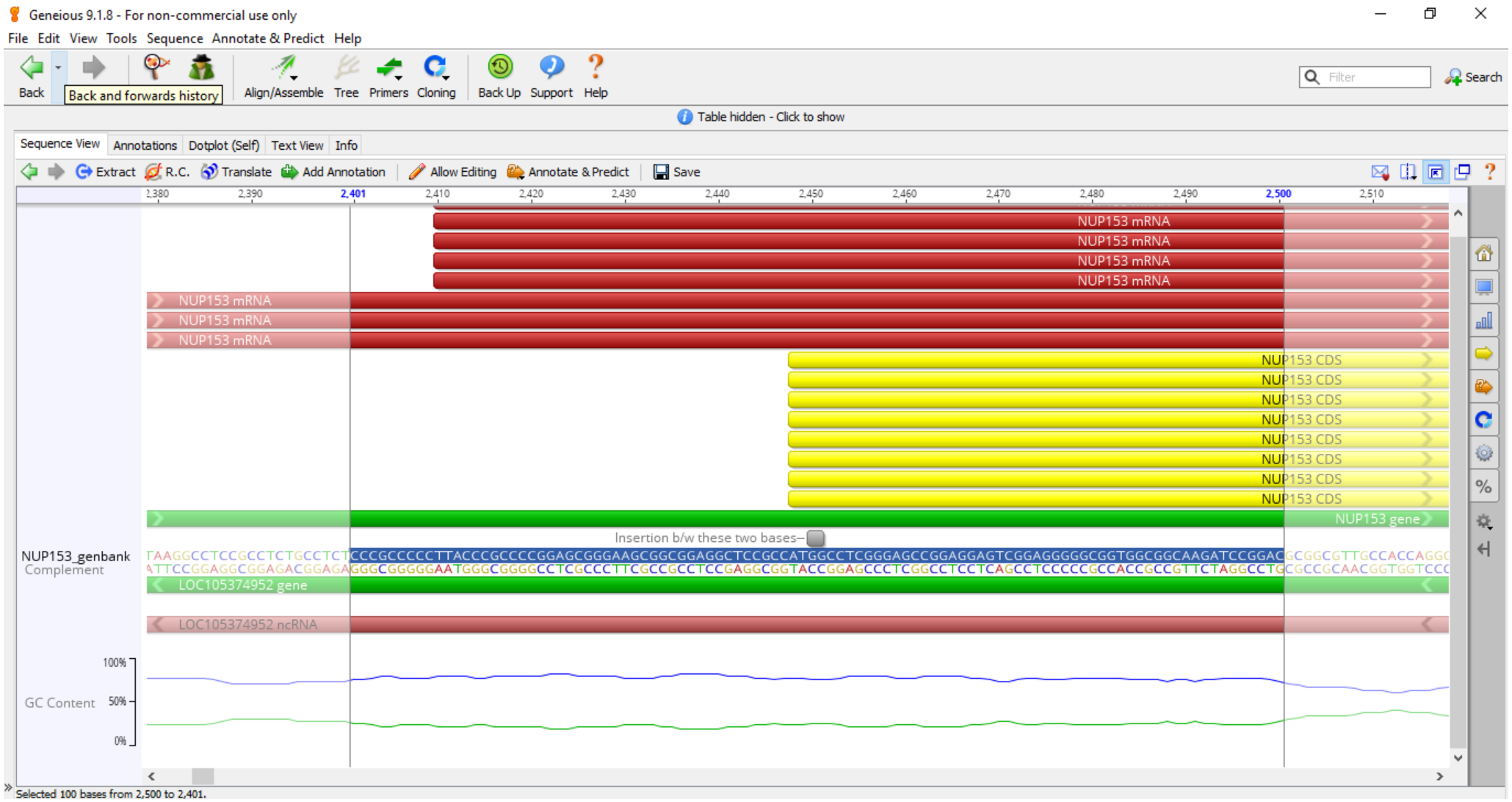


Editing design: Selecting CRISPR guides

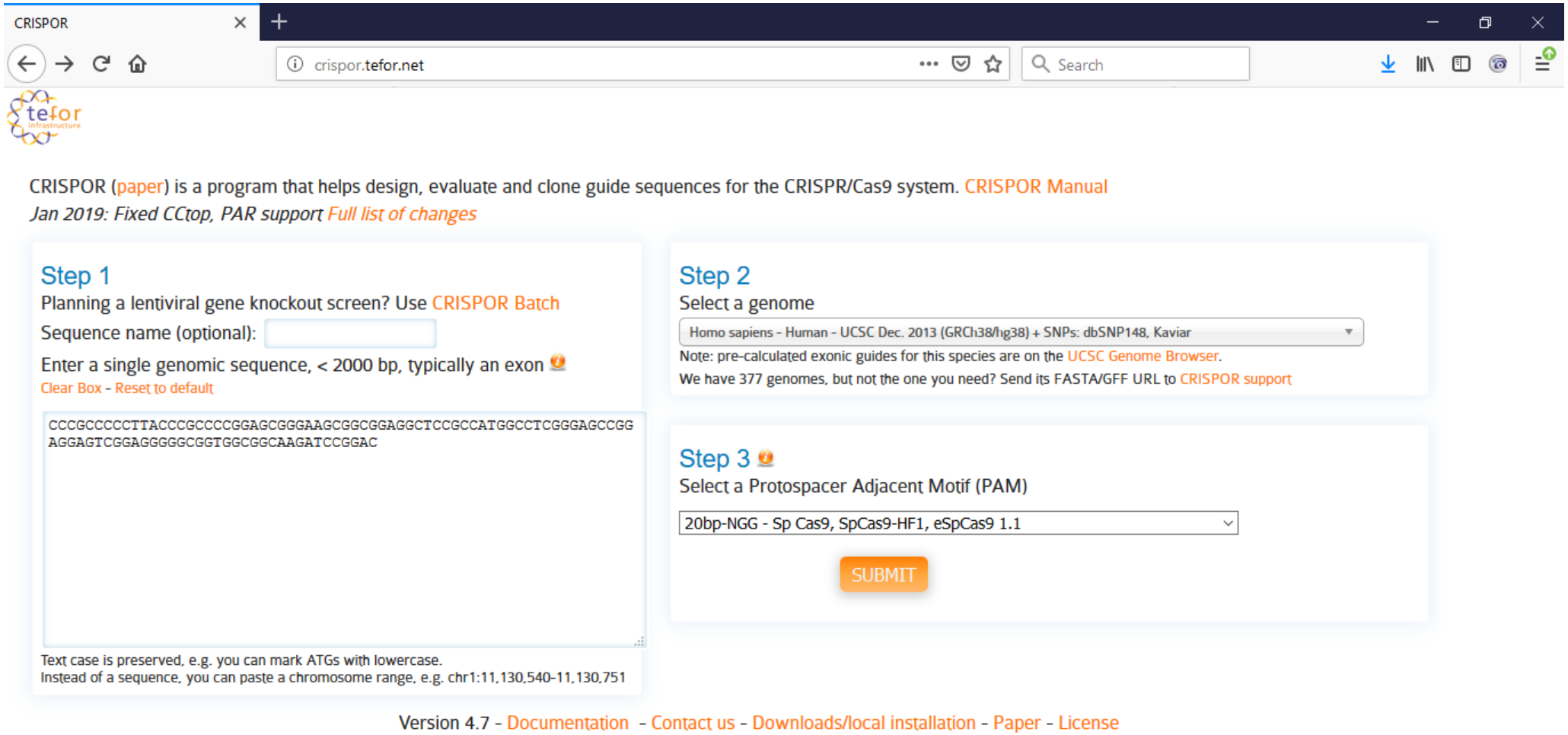
- There are **MANY** tools for CRISPR guide design tools
  - https://en.wikipedia.org/wiki/CRISPR/Cas_Tools

- Some useful tools:
  - CRISPOR
  - UCSC genome browser pre-computed CRISPR tracks for human and mouse

Editing design: Selecting guides w/ CRISPOR



Editing design: Selecting guides w/ CRISPOR



http://crispor.tefor.net/

Editing design: Selecting guides w/ CRISPOR

CRISPOR: hg38

crispor.tefor.net/crispor.py?batchId=eR4LRfwZYWMgiaeyxvat

Homo sapiens (hg38), chr6:17706335-17706434, reverse genomic strand

Found 34 possible guide sequences in input (100 bp). Click on a PAM NGG match to show its 20 bp guide sequence.
Shown below are the PAM site and the expected cleavage position located -3bp 5' of the PAM site.
Colors green, yellow and red indicate high, medium and low specificity of the PAM's guide sequence in the genome.

Variant database: dbSNP 148 Common SNPs Update Missing a variant database? We can probably add it.

```
Position  0         10        20        30        40        50        60        70        80        90
Variants  ..................................................................................................
Sequence  CCCGCCCCCTTACCCGCCCCGGAGCGGGAAGCGGCGGAGGCTCCGCCATGGCCTCGGGAGCCGGAGGAGTCGGAGGGGGCGGTGGCGGCAAGATCCGGAC
          CCC---      CCC---   ---CGG    ---CGG     CCG---   CCT---   CCG---    ---AGG   ---TGG       ---CGG
           CCG---      CCG---   ---GGG      ---AGG     CCA---        ---CGG   ---CGG   ---CGG
              CCC---      CCC---      ---CGG            ---TGG          ---AGG    ---GGG     ---CGG
               CCC---      CCC---                             ---CGG              ---GGG
                CCC---      CCG---                             ---GGG               ---GGG
                 CCT---
```

Download for: SerialCloner (free) - ApE (free) - GenomeCompiler - Benchling - SnapGene - Geneious - Vector NTI - LaserGene - Genbank - FASTA - Copy sequence to clipboard

Predicted guide sequences for PAMs

Ranked by default from highest to lowest specificity score (Hsu et al., Nat Biot 2013). Click on a column title to rank by a score.
If you use this website, please cite our CRISPOR paper in Gen Biol 2016. Too much information? Look at the CRISPOR manual.

Download as Excel tables: Guides / Off-targets / Saturating mutagenesis assistant

Position/ Strand	Guide Sequence + *PAM* + Restriction Enzymes + Variants	Specificity Score	Predicted Efficiency Show all scores	Out-of-Frame score	Off-targets for 0-1-2-3-4 mismatches	Genome Browser links to matches sorted by CFD off-target score ☐ exons only ☐ chr6 only

Editing design: Selecting guides w/ CRISPOR

CRISPOR

crispor.tefor.net/crispor.py?batchId=eR4LRfwZYWMgiaeyxvat&showAllSco

Download as Excel tables: Guides / Off-targets / Saturating mutagenesis assistant

Position/ Strand	Guide Sequence + *PAM* + Restriction Enzymes + Variants ☐ Only G- ☐ Only GG- ☐ Only A-	Specificity Score	Predicted Efficiency Main scores: Doench '16	Doench '16-Old	Chari	Xu	Doench '14	Wang	Moreno-Mateos	Azimuth in-vitro	CCTop	Out-of-Frame Click score for details	Off-targets for 0-1-2-3-4 mismatches + next to PAM	Genome Browser links to matches sorted by CFD off-target score ☐ exons only ☐ chr6 only
8 / rev	TTCCCGCTCCGGGGCGGGTA *AGG* .... **Cloning / PCR primers**	94	34	29	48	-0.40		25	63	53	64	59	0 - 0 - 0 - 1 - 51 0 - 0 - 0 - 0 - 0 52 off-targets	4:intergenic:TSPAN6-SRPX2 4:intergenic:RP11-544P1.3-USP6NL 4:intergenic:APOBEC3A-APOBEC3B show all...
7 / rev	TCCCGCTCCGGGGCGGGTAA *GGG* .... **Cloning / PCR primers**	92	38	35	12	-0.25		20	54	54	81	59	0 - 0 - 0 - 4 - 50 0 - 0 - 0 - 1 - 2 54 off-targets	4:intron:CH507-9B2.4/CH507-9B2.3 4:intron:C21orf33 4:intergenic:PHF2-MIR4291 show all...
6 / rev	CCCGCTCCGGGGCGGGTAAG *GGG* .... ⚠ High GC content **Cloning / PCR primers**	88	33	28	83	-0.22		26	72	52	69	58	0 - 0 - 2 - 5 - 72 0 - 0 - 1 - 0 - 1 79 off-targets	4:intron:MAP3K15 4:exon:ITGAV 4:intergenic:BIRC6-TTC27 show all...
5 / rev	CCGCTCCGGGGCGGGTAAGG *GGG* .... ⚠ High GC content **Cloning / PCR primers**	85	38	36	94	0.61		49	86	68	72	56	0 - 0 - 0 - 10 - 95 0 - 0 - 0 - 0 - 0	3:intron:MOV10 4:intergenic:RP11-796E10.1-AC106900.6 4:exon:KCNQ5 show all...

Editing design: Selecting guides w/ CRISPOR

crispor.tefor.net/crispor.py?batchId=eR4LRfwZYWMgiaeyxvat&showAllSco

Download as Excel tables: Guides / Off-targets / Saturating mutagenesis assistant

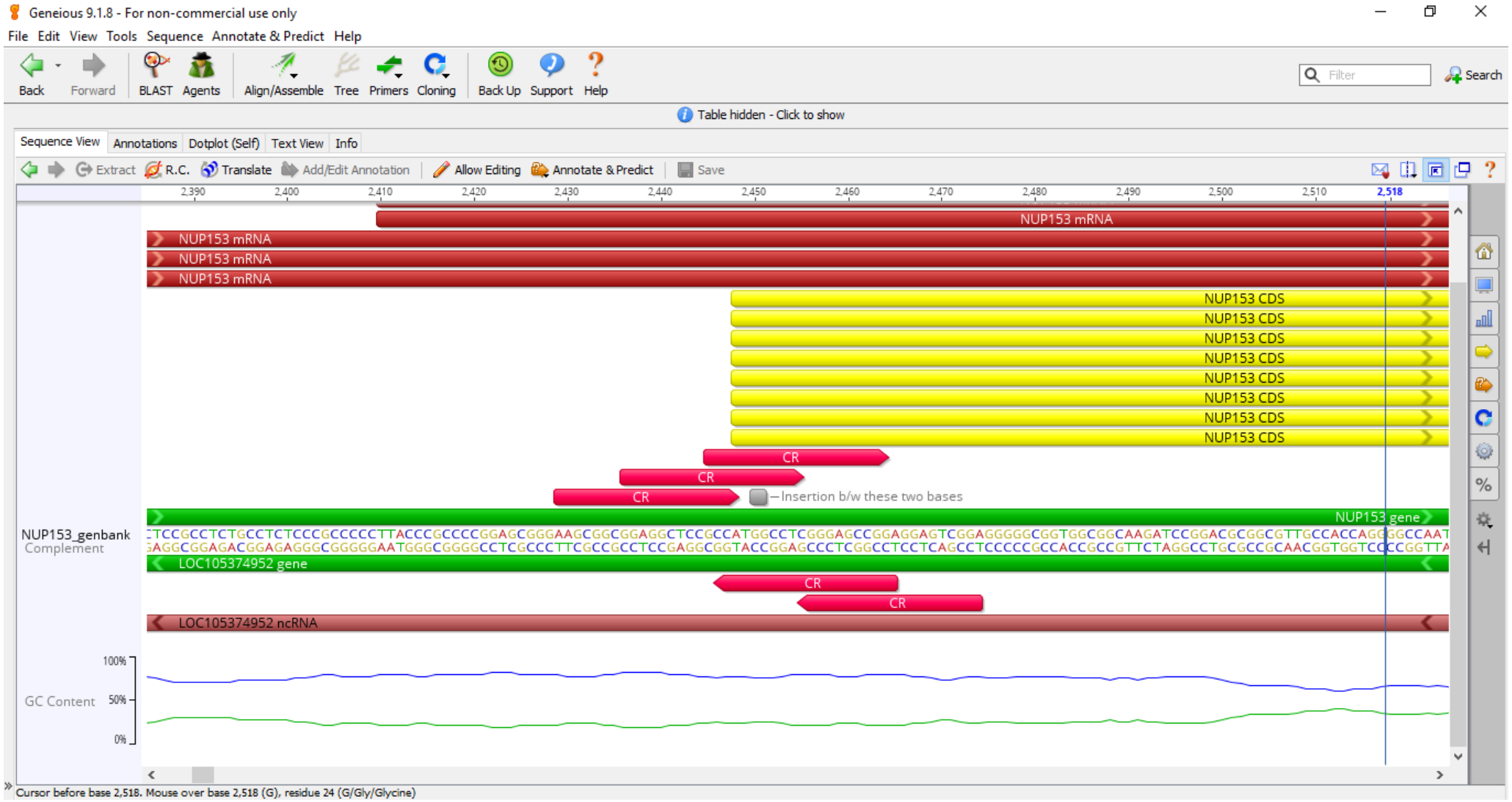
Position/ Strand	Guide Sequence + *PAM* + Restriction Enzymes + Variants ☐ Only G- ☐ Only GG- ☐ Only A-	Specificity Score	Predicted Efficiency Main scores: Doench '16	Doench '16-Old	Chari	Xu	Doench '14	Wang	Moreno-Mateos	Azimuth in-vitro	CCTop	Out-of-Frame Click score for details	Off-targets for 0-1-2-3-4 mismatches + next to PAM	Genome Browser links to matches sorted by CFD off-target score ☐ exons only ☐ chr6 only
8 / rev	TTCCCGCTCCGGGGCGGGTA *AGG* . . . . **Cloning / PCR primers**	94	34	29	48	-0.40		25	63	53	64	59	0 - 0 - 0 - 1 - 51 0 - 0 - 0 - 0 - 0 52 off-targets	4:intergenic:TSPAN6-SRPX2 4:intergenic:RP11-544P1.3-USP6NL 4:intergenic:APOBEC3A-APOBEC3B 4:intron:TEAD3 4:intergenic:SLIT2-SLIT2-IT1 4:intergenic:TRMT61A-RNU7-160P 4:intergenic:RP4-630A11.3-LEPR 4:exon:LRRC4B 4:intron:TAB2 4:intron:PKN3 4:exon:LINC00957 4:intergenic:RP11-16N2.1-LPHN3 4:intron:FAM129A 4:exon:PIGP 4:exon:CACNA1C-AS1/CACNA1C 4:intergenic:PAWR-PPP1R12A 4:intergenic:FUCA2-PHACTR2 4:intergenic:LINC01326-RN7SKP298 4:intergenic:XKR4-RP11-628E19.2 4:intergenic:ATP6V1D/EIF2S1-EIF2S1 4:intergenic:UBE2I-BAIAP3 4:intron:SPATA17

```
guide:       TTCCCGCTCCGGGGCGGGTA AGG
off-target:  TTACCACTCTGGGGCAGGTA GGG
                *  *   *     *

CFD Off-target score: 0.815686
MIT Off-target score: 0.14
Position: chrX:100639314-100639336:-
```

genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&position=chrX:100639314-100639336

Editing design: Selecting guides w/ CRISPOR

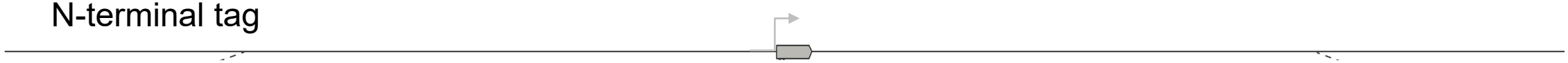


Outline

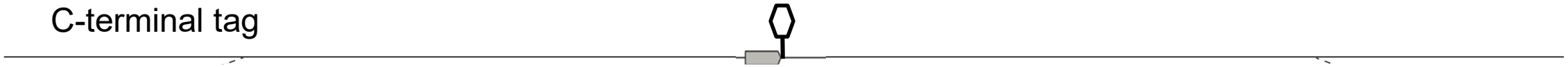
- Intro to the Allen Institute for Cell Science
- Gene editing design
  - Introduction
  - crRNA design
  - **Donor template design**
- Workflows and methods for generating high-quality clonal hiPS cell lines
- In-lab demonstrations
- Roundtable discussion

Basic Overview for Donor Molecule Design

N-terminal tag



C-terminal tag



What you'll need

- ❑ Gene locus file (showing DNA sequence) with crRNA target sites annotated (we are using NUP153)
- ❑ Sequence of FP of choice (we will use mEGFP)
- ❑ Amino acid sequence of linker (we are using YSDLE)
- ❑ Codon frequency table for your species (we are using human)
- ❑ Link to UCSC Genome Browser (with the "conservation" track enabled)

Human codon frequency table

Triplet	Amino acid	Fraction	Frequency/ Thousand	Number	Triplet	Amino acid	Fraction	Frequency/ Thousand	Number
TTT	F	0.45	16.9	336562	TCT	S	0.18	14.6	291040
TTC	F	0.55	20.4	406571	TCC	S	0.22	17.4	346943
TTA	L	0.07	7.2	143715	TCA	S	0.15	11.7	233110
TTG	L	0.13	12.6	249879	TCG	S	0.06	4.5	89429
TAT	Y	0.43	12.0	239268	TGT	C	0.45	9.9	197293
TAC	Y	0.57	15.6	310695	TGC	C	0.55	12.2	243685
TAA	*	0.28	0.7	14322	TGA	*	0.52	1.3	25383
TAG	*	0.20	0.5	10915	TGG	W	1.00	12.8	255512
CTT	L	0.13	12.8	253795	CCT	P	0.28	17.3	343793
CTC	L	0.20	19.4	386182	CCC	P	0.33	20.0	397790
CTA	L	0.07	6.9	138154	CCA	P	0.27	16.7	331944
CTG	L	0.41	40.3	800774	CCG	P	0.11	7.0	139414
CAT	H	0.41	10.4	207826	CGT	R	0.08	4.7	93458
CAC	H	0.59	14.9	297048	CGC	R	0.19	10.9	217130
CAA	Q	0.25	11.8	234785	CGA	R	0.11	6.3	126113
CAG	Q	0.75	34.6	688316	CGG	R	0.21	11.9	235938
ATT	I	0.36	15.7	313225	ACT	T	0.24	12.8	255582
ATC	I	0.48	21.4	426570	ACC	T	0.36	19.2	382050
ATA	I	0.16	7.1	140652	ACA	T	0.28	14.8	294223
ATG	M	1.00	22.3	443795	ACG	T	0.12	6.2	123533
AAT	N	0.46	16.7	331714	AGT	S	0.15	11.9	237404
AAC	N	0.54	19.5	387148	AGC	S	0.24	19.4	385113
AAA	K	0.42	24.0	476554	AGA	R	0.20	11.5	228151
AAG	K	0.58	32.9	654280	AGG	R	0.20	11.4	227281
GTT	V	0.18	10.9	216818	GCT	A	0.26	18.6	370873
GTC	V	0.24	14.6	[illegible]	GCC	A	0.40	28.5	567930
GTA	V	0.11	7.0	[illegible]	GCA	A	0.23	16.0	317338
GTG	V	0.47	28.9	575438	GCG	A	0.11	7.6	150708
GAT	D	0.46	22.3	443369	GGT	G	0.16	10.8	215544
GAC	D	0.54	26.0	517579	GGC	G	0.34	22.8	453917
GAA	E	0.42	29.0	577846	GGA	G	0.25	16.3	325243
GAG	E	0.58	40.8	810842	GGG	G	0.25	16.4	326879

https://www.genscript.com/tools/codon-frequency-table

Human codon frequency table

Triplet	Amino acid	Fraction	Frequency/ Thousand	Number	Triplet	Amino acid	Fraction	Frequency/ Thousand	Number
TTT	F	0.45	16.9	336562	TCT	S	0.18	14.6	291040
TTC	F	0.55	20.4	406571	TCC	S	0.22	17.4	346943
TTA	L	0.07	7.2	143715	TCA	S	0.15	11.7	233110
TTG	L	0.13	12.6	249879	TCG	S	0.06	4.5	89429
TAT	Y	0.43	12.0	239268	TGT	C	0.45	9.9	197293
TAC	Y	0.57	15.6	310695	TGC	C	0.55	12.2	243685
TAA	*	0.28	0.7	14322	TGA	*	0.52	1.3	25383
TAG	*	0.20	0.5	10915	TGG	W	1.00	12.8	255512
CTT	L	0.13	12.8	253795	CCT	P	0.28	17.3	343793
CTC	L	0.20	19.4	386182	CCC	P	0.33	20.0	397790
CTA	L	0.07	6.9	138154	CCA	P	0.27	16.7	331944
CTG	L	0.41	40.3	800774	CCG	P	0.11	7.0	139414
CAT	H	0.41	10.4	207826	CGT	R	0.08	4.7	93458
CAC	H	0.59	14.9	297048	CGC	R	0.19	10.9	217130
CAA	Q	0.25	11.8	234785	CGA	R	0.11	6.3	126113
CAG	Q	0.75	34.6	688316	CGG	R	0.21	11.9	235938
ATT	I	0.36	15.7	313225	ACT	T	0.24	12.8	255582
ATC	I	0.48	21.4	426570	ACC	T	0.36	19.2	382050
ATA	I	0.16	7.1	140652	ACA	T	0.28	14.8	294223
ATG	M	1.00	22.3	443795	ACG	T	0.12	6.2	123533
AAT	N	0.46	16.7	331714	AGT	S	0.15	11.9	237404
AAC	N	0.54	19.5	387148	AGC	S	0.24	19.4	385113
AAA	K	0.42	24.0	476554	AGA	R	0.20	11.5	228151
AAG	K	0.58	32.9	654280	AGG	R	0.20	11.4	227281
GTT	V	0.18	10.9	216818	GCT	A	0.26	18.6	370873
GTC	V	0.24	14.6	290874	GCC	A	0.40	28.5	567930
GTA	V	0.11	7.0	139156	GCA	A	0.23	16.0	317338
GTG	V	0.47	28.9	575438	GCG	A	0.11	7.6	150708
GAT	D	0.46	22.3	443369	GGT	G	0.16	10.8	215544
GAC	D	0.54	26.0	517579	GGC	G	0.34	22.8	453917
GAA	E	0.42	29.0	577846	GGA	G	0.25	16.3	325243
GAG	E	0.58	40.8	810842	GGG	G	0.25	16.4	326879

https://www.genscript.com/tools/codon-frequency-table

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Help | About Us

Human BLAT Search

BLAT Search Genome

Genome: ☐ Search ALL | Assembly: | Query type: | Sort output: | Output type:

Human | Dec. 2013 (GRCh38/hg38) | BLAT's guess | query,score | hyperlink

```
GGGCCGGGCGGCGGGGGAGGCTCTCTAAGGCCTCCGCCTCTGCCTCTCCCGCCCCCTTACCCGCCCCGGAGCGGGAAG
CGGCGGAGGCTCCGCCATG
```

submit | I'm feeling lucky | clear

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence: Choose File No file chosen submit file

Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters.
A valid example is `GTCCTCGGAACCAGGACCTCGGCGTGGCCTAGCG` (human SOD1).

The **Search ALL** checkbox above the Genome drop-down list allows you to search the genomes of the default assemblies for all of our organisms. It also searches any attached hubs' blat servers. This shows you which organisms have the highest homology with your query sequence. The results are ordered so that the organism whose best alignment has the most hits is at the top, and shows the best region found. It makes quick approximate alignments based only on the raw hits, which are a perfectly matching short sub-sequence of a fixed size: 11 for DNA and 4 for protein. The entire alignment, including mismatches and gaps, must score 20 or higher in order to appear in the BLAT output. Having too few hits will often yield no BLAT results. Click the Assembly column link on the results page to see the full BLAT output for that organism.

For locating PCR primers, use In-Silico PCR for best results instead of BLAT.

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Help | About Us

Human (hg38) BLAT Results

BLAT Search Results

Go back to chr6:17706385-17706445 on the Genome Browser.

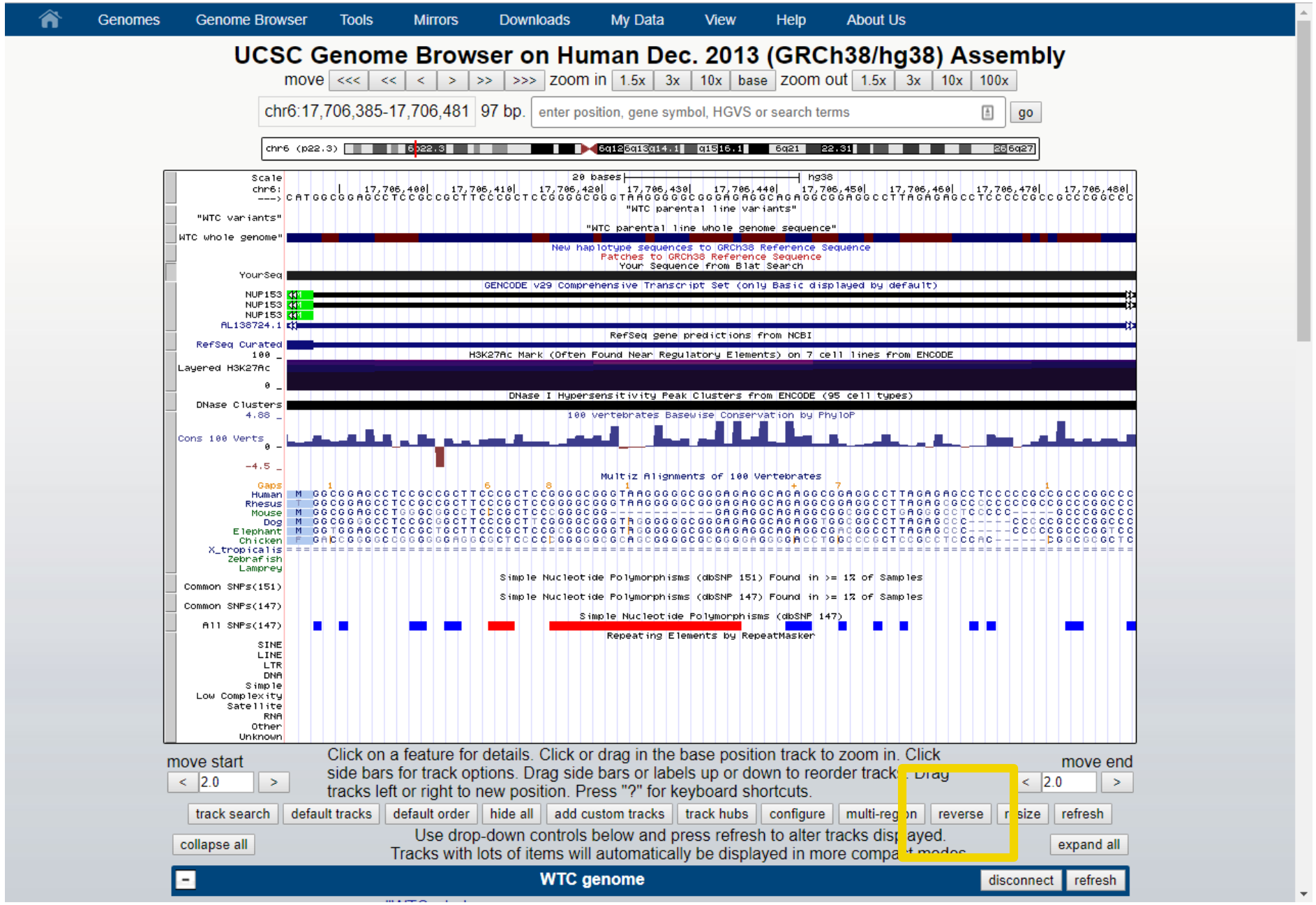
Custom track name: blat YourSeq

Custom track description: blat on YourSeq

Build a custom track with these results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	97	1	97	97	100.0%	chr6	-	17706385	17706481	97
browser details	YourSeq	26	59	94	97	96.6%	chr19	+	590724	590761	38
browser details	YourSeq	23	31	54	97	100.0%	chr9	-	137760135	137760162	28
browser details	YourSeq	21	30	50	97	100.0%	chr16	+	71588821	71588841	21

Missing a match?


Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
View
Help
About Us
UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move
zoom in
zoom out
chr6:17,706,385-17,706,481
97 bp.
enter position, gene symbol, HGVS or search terms
go
"WTC parental line variants"
"WTC parental line whole genome sequence"
New haplotype sequences to GRCh38 Reference Sequence
Patches to GRCh38 Reference Sequence
Your Sequence from Blat Search
GENCODE v29 Comprehensive Transcript Set (only Basic displayed by default)
RefSeq gene predictions from NCBI
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
DNase I Hypersensitivity Peak Clusters from ENCODE (95 cell types)
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Simple Nucleotide Polymorphisms (dbSNP 151) Found in >= 1% of Samples
Simple Nucleotide Polymorphisms (dbSNP 147) Found in >= 1% of Samples
Simple Nucleotide Polymorphisms (dbSNP 147)
Repeating Elements by RepeatMasker
move start
move end
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.
track search
default tracks
default order
hide all
add custom tracks
track hubs
configure
reverse
refresh
collapse all
expand all
Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
WTC genome
disconnect
refresh

mRNA and EST

refresh

P12 Human mRNAs hide	P12 Spliced ESTs hide	P12 Human ESTs hide	P12 Other ESTs hide	P12 Other mRNAs hide	SIB Alt-Splicing hide

Expression

refresh

19 P12 GTEx Gene hide	GTEx Transcript hide	Affy GNF1H hide	Affy U133 hide	Affy U95 hide	GNF Atlas 2 hide
GWIPS-viz Riboseq hide	miRNA Tissue Atlas hide				

Regulation

refresh

19 P12 ENCODE Regulation... show	Updated GeneHancer hide	P12 CpG Islands... hide	ORegAnno hide	RefSeq Func Elems hide

Comparative Genomics

refresh

Conservation full	Cons 7 Verts hide	Cons 20 Mammals hide	Cons 30 Primates hide	Primate Chain/Net hide	Placental Chain/Net hide
Vertebrate Chain/Net hide	Chicken Chain/Net hide				

Variation

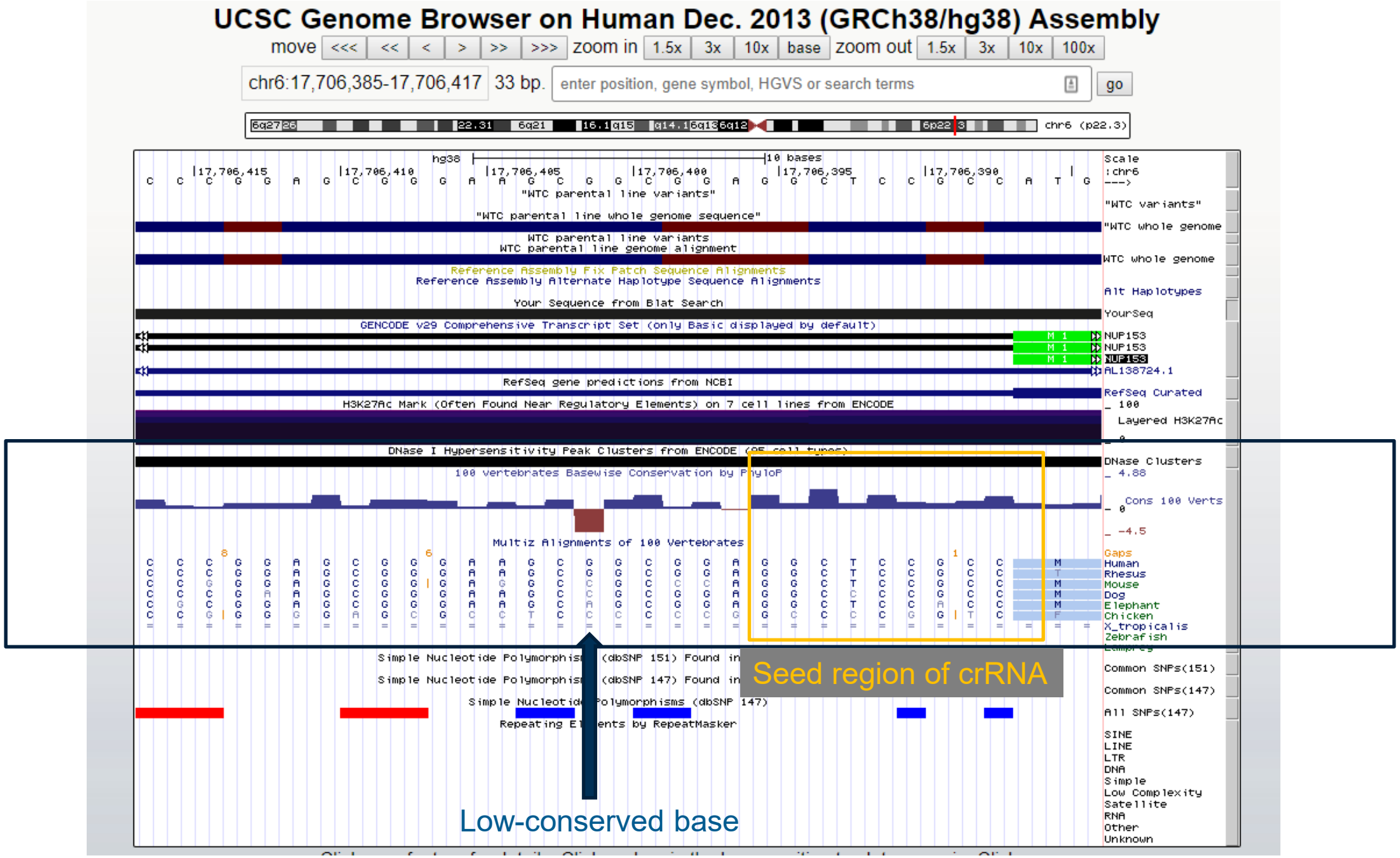
refresh

Common SNPs(151) dense	Common SNPs(150) hide	Common SNPs(147) dense	Common SNPs(146) hide	Common SNPs(144) hide	Common SNPs(142) hide
Common SNPs(141) hide	All SNPs(151) hide	All SNPs(150) hide	All SNPs(147) dense	All SNPs(146) hide	All SNPs(144) hide
All SNPs(142) hide	All SNPs(141) hide	Flagged SNPs(151) hide	Flagged SNPs(150) hide	Flagged SNPs(147) hide	Flagged SNPs(146) hide
Flagged SNPs(144) hide	Flagged SNPs(142) hide	Flagged SNPs(141) hide	Mult. SNPs(151) hide	Mult. SNPs(150) hide	Mult. SNPs(147) hide
Mult. SNPs(146) hide	Mult. SNPs(144) hide	Mult. SNPs(142) hide	Mult. SNPs(141) hide	DGV Struct Var hide	

Repeats

refresh

P12 RepeatMasker	Interrupted Rpts	RepeatMasker Viz.	Microsatellite	Segmental Dups	Self Chain

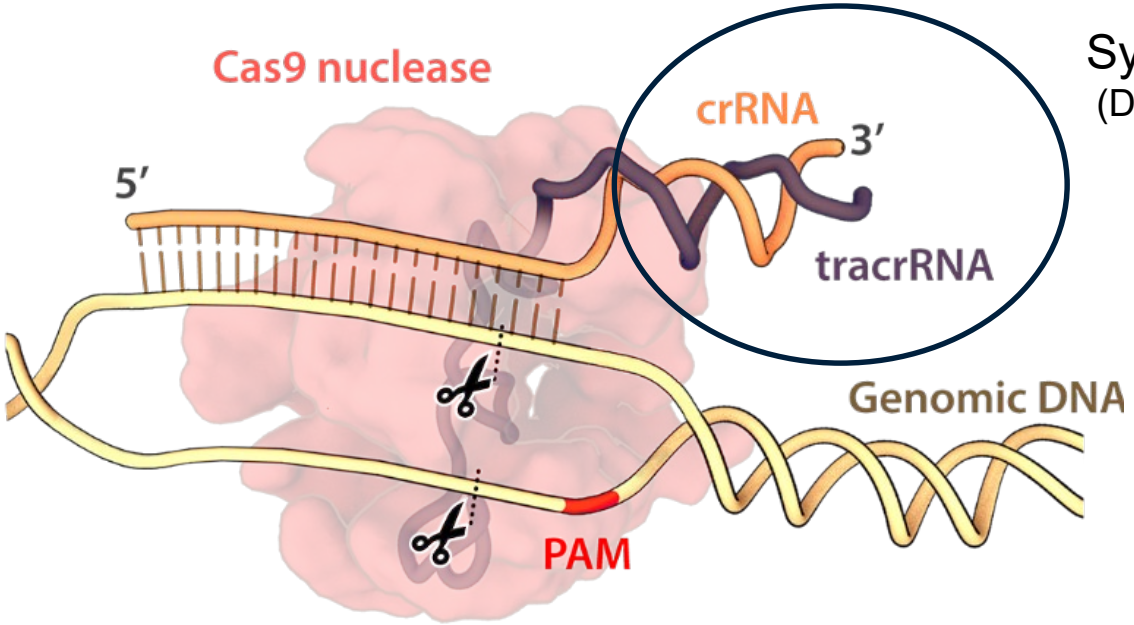

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
chr6:17,706,385-17,706,417 33 bp.
enter position, gene symbol, HGVS or search terms
go
chr6 (p22.3)
hg38
10 bases
"WTC parental line variants"
"WTC parental line whole genome sequence"
WTC parental line variants
WTC parental line genome alignment
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
Your Sequence from Blat Search
GENCODE v29 Comprehensive Transcript Set (only Basic displayed by default)
NUP153
AL138724.1
RefSeq gene predictions from NCBI
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
DNase I Hypersensitivity Peak Clusters from ENCODE
100 vertebrates Basewise Conservation by PhyloP
Multiz Alignments of 100 Vertebrates
Human
Rhesus
Mouse
Dog
Elephant
Chicken
X_tropicalis
Zebrafish
Simple Nucleotide Polymorphisms (dbSNP 151) Found in
Simple Nucleotide Polymorphisms (dbSNP 147) Found in
Simple Nucleotide Polymorphisms (dbSNP 147)
Repeating Elements by RepeatMasker
Seed region of crRNA
Low-conserved base
Scale
chr6
"WTC variants"
"WTC whole genome
WTC whole genome
Alt Haplotypes
YourSeq
RefSeq Curated
Layered H3K27Ac
DNase Clusters
Cons 100 Verts
Gaps
Common SNPs(151)
Common SNPs(147)
All SNPs(147)
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

Human codon frequency table

Triplet	Amino acid	Fraction	Frequency/ Thousand	Number	Triplet	Amino acid	Fraction	Frequency/ Thousand	Number
TTT	F	0.45	16.9	336562	TCT	S	0.18	14.6	291040
TTC	F	0.55	20.4	406571	TCC	S	0.22	17.4	346943
TTA	L	0.07	7.2	143715	TCA	S	0.15	11.7	233110
TTG	L	0.13	12.6	249879	TCG	S	0.06	4.5	89429
TAT	Y	0.43	12.0	239268	TGT	C	0.45	9.9	197293
TAC	Y	0.57	15.6	310695	TGC	C	0.55	12.2	243685
TAA	*	0.28	0.7	14322	TGA	*	0.52	1.3	25383
TAG	*	0.20	0.5	10915	TGG	W	1.00	12.8	255512
CTT	L	0.13	12.8	253795	CCT	P	0.28	17.3	343793
CTC	L	0.20	19.4	386182	CCC	P	0.33	20.0	397790
CTA	L	0.07	6.9	138154	CCA	P	0.27	16.7	331944
CTG	L	0.41	40.3	800774	CCG	P	0.11	7.0	139414
CAT	H	0.41	10.4	207826	CGT	R	0.08	4.7	93458
CAC	H	0.59	14.9	297048	CGC	R	0.19	10.9	217130
CAA	Q	0.25	11.8	234785	CGA	R	0.11	6.3	126113
CAG	Q	0.75	34.6	688316	CGG	R	0.21	11.9	235938
ATT	I	0.36	15.7	313225	ACT	T	0.24	12.8	255582
ATC	I	0.48	21.4	426570	ACC	T	0.36	19.2	382050
ATA	I	0.16	7.1	140652	ACA	T	0.28	14.8	294223
ATG	M	1.00	22.3	443795	ACG	T	0.12	6.2	123533
AAT	N	0.46	16.7	331714	AGT	S	0.15	11.9	237404
AAC	N	0.54	19.5	387148	AGC	S	0.24	19.4	385113
AAA	K	0.42	24.0	476554	AGA	R	0.20	11.5	228151
AAG	K	0.58	32.9	654280	AGG	R	0.20	11.4	227281
GTT	V	0.18	10.9	216818	GCT	A	0.26	18.6	370873
GTC	V	0.24	14.6	290874	GCC	A	0.40	28.5	567930
GTA	V	0.11	7.0	139156	GCA	A	0.23	16.0	317338
GTG	V	0.47	28.9	575438	GCG	A	0.11	7.6	150708
GAT	D	0.46	22.3	443369	GGT	G	0.16	10.8	215544
GAC	D	0.54	26.0	517579	GGC	G	0.34	22.8	453917
GAA	E	0.42	29.0	577846	GGA	G	0.25	16.3	325243
GAG	E	0.58	40.8	810842	GGG	G	0.25	16.4	326879

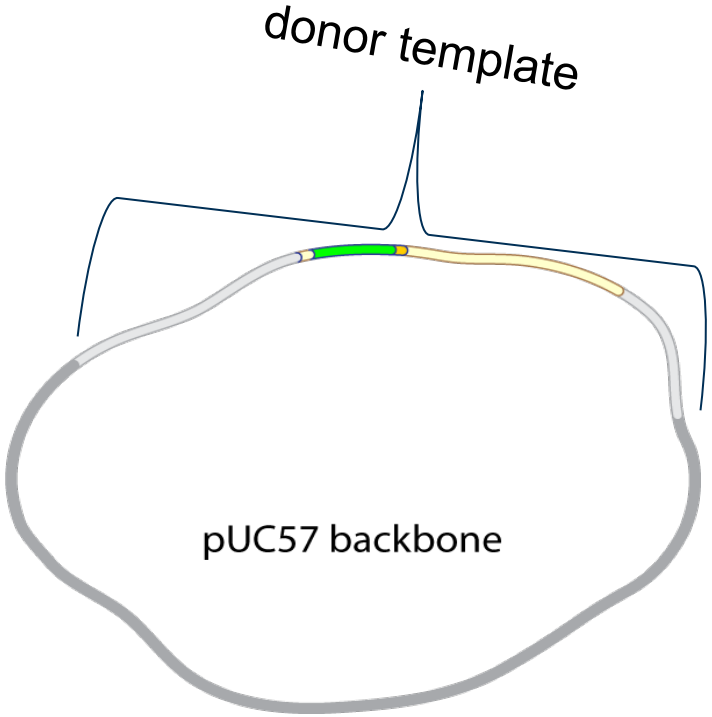
https://www.genscript.com/tools/codon-frequency-table

Ordering Reagents: crRNAs and Donor Templates



Synthetic unmodified 2-part crRNA and tracrRNA
(Dharmacon, Horizon Discovery)

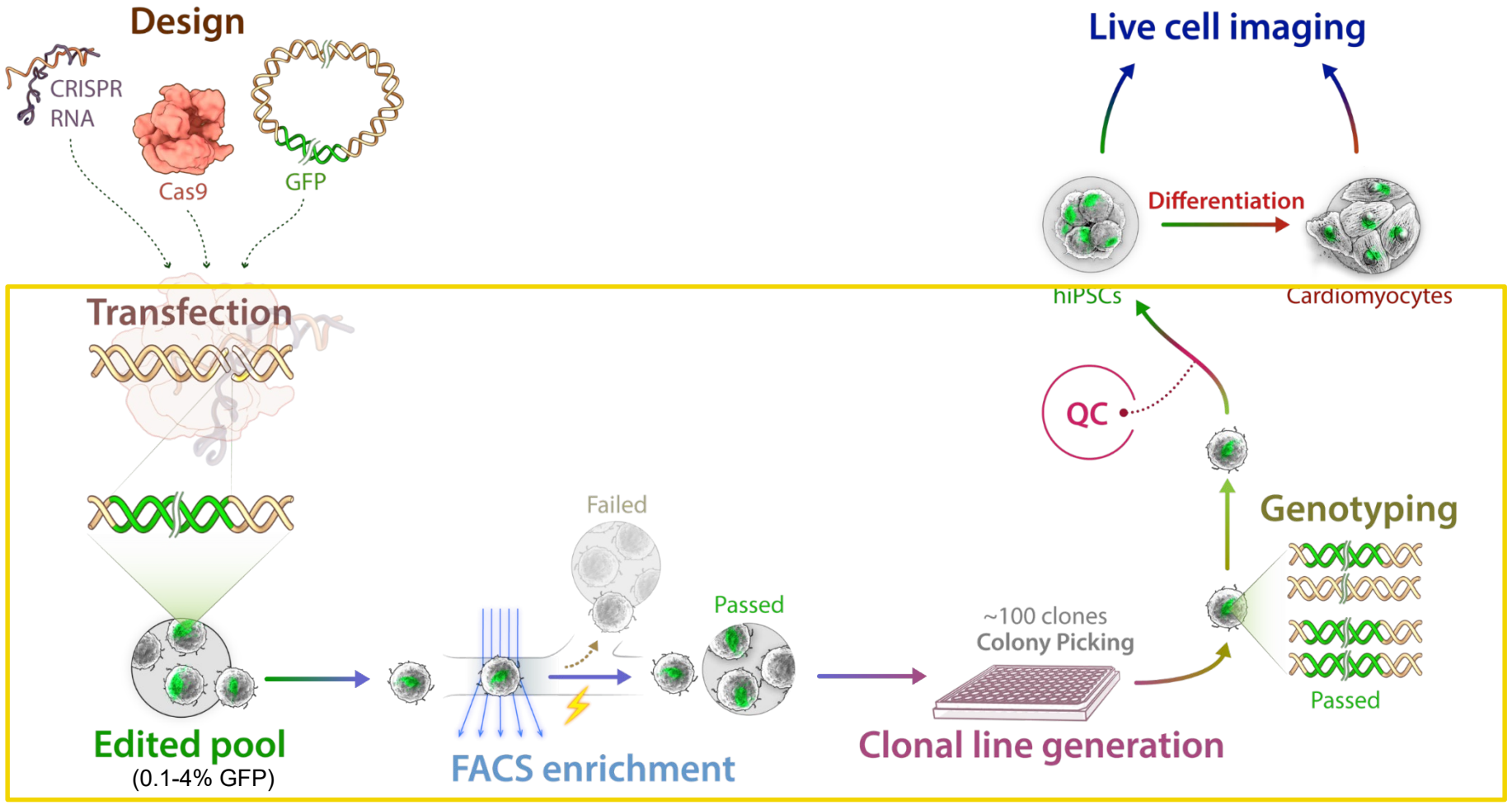
Pros: short-lived in cells, easy to order, pure, fast turnaround time
Cons: have to purchase tracrRNA, too



Plasmid, synthesized commercially
(Genewiz)

Pros: easy to order, sequence-verified, customizable backbone
Cons: costly, takes 6-8 weeks, can randomly integrate in genome, can be toxic to cells

Next!


Design
CRISPR RNA
Cas9
GFP
Transfection
Edited pool
(0.1-4% GFP)
Failed
Passed
FACS enrichment
~100 clones
Colony Picking
Clonal line generation
Genotyping
Passed
QC
hiPSCs
Differentiation
Cardiomyocytes
Live cell imaging