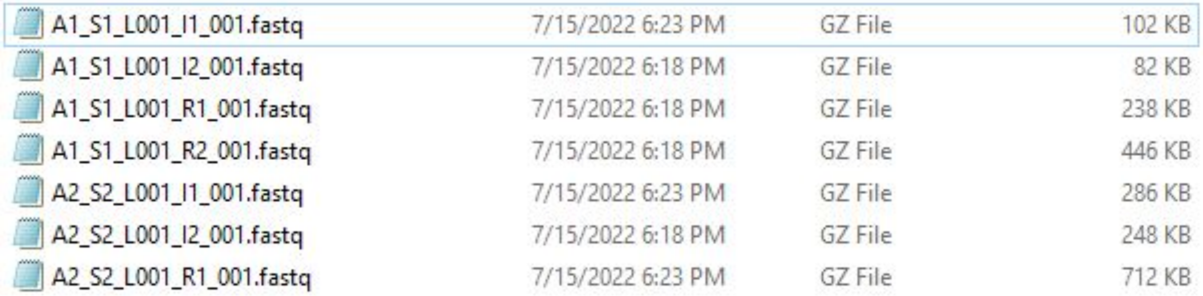
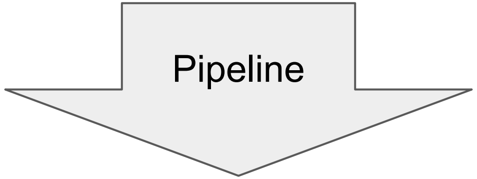
SCI-lite Analysis Pipeline: from FASTQ to cell-level heteroplasmy



A1_S1_L001_I1_001.fastq	7/15/2022 6:23 PM	GZ File	102 KB
A1_S1_L001_I2_001.fastq	7/15/2022 6:18 PM	GZ File	82 KB
A1_S1_L001_R1_001.fastq	7/15/2022 6:18 PM	GZ File	238 KB
A1_S1_L001_R2_001.fastq	7/15/2022 6:18 PM	GZ File	446 KB
A2_S2_L001_I1_001.fastq	7/15/2022 6:23 PM	GZ File	286 KB
A2_S2_L001_I2_001.fastq	7/15/2022 6:18 PM	GZ File	248 KB
A2_S2_L001_R1_001.fastq	7/15/2022 6:23 PM	GZ File	712 KB

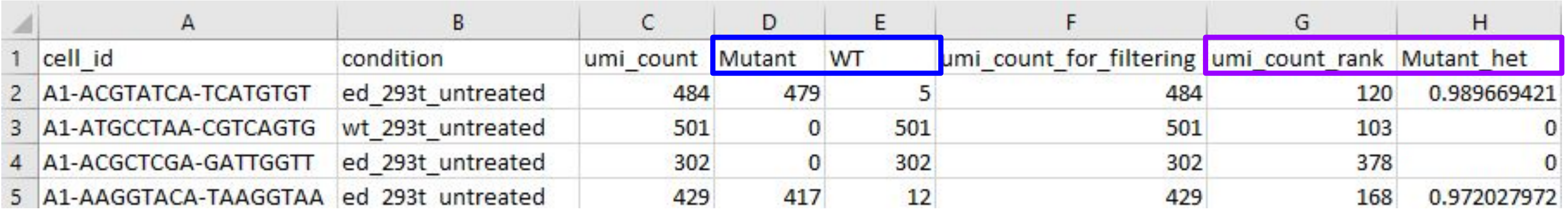


Pipeline

one row = one cell

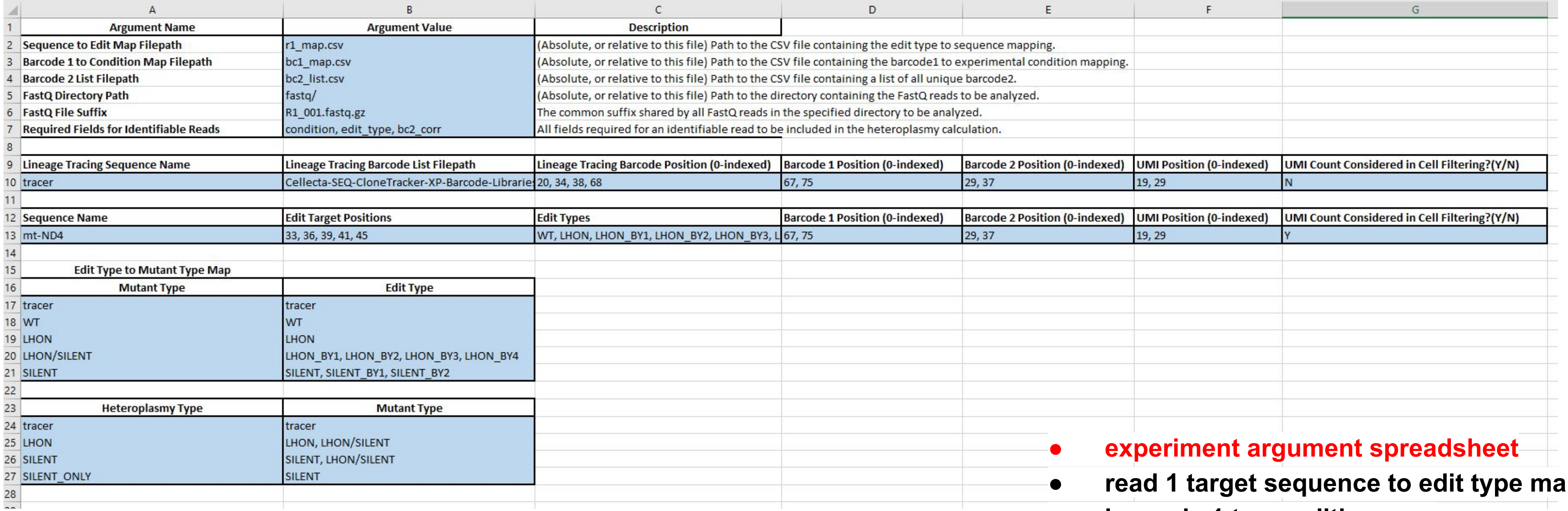
# of UMIs for each UMI type

Heteroplasmy level based on each UMI type



	A	B	C	D	E	F	G	H
1	cell_id	condition	umi_count	Mutant	WT	umi_count_for_filtering	umi_count_rank	Mutant_het
2	A1-ACGTATCA-TCATGTGT	ed_293t_untreated	484	479	5	484	120	0.989669421
3	A1-ATGCCTAA-CGTCAGTG	wt_293t_untreated	501	0	501	501	103	0
4	A1-ACGCTCGA-GATTGGTT	ed_293t_untreated	302	0	302	302	378	0
5	A1-AAGGTACA-TAAGGTAA	ed_293t_untreated	429	417	12	429	168	0.972027972

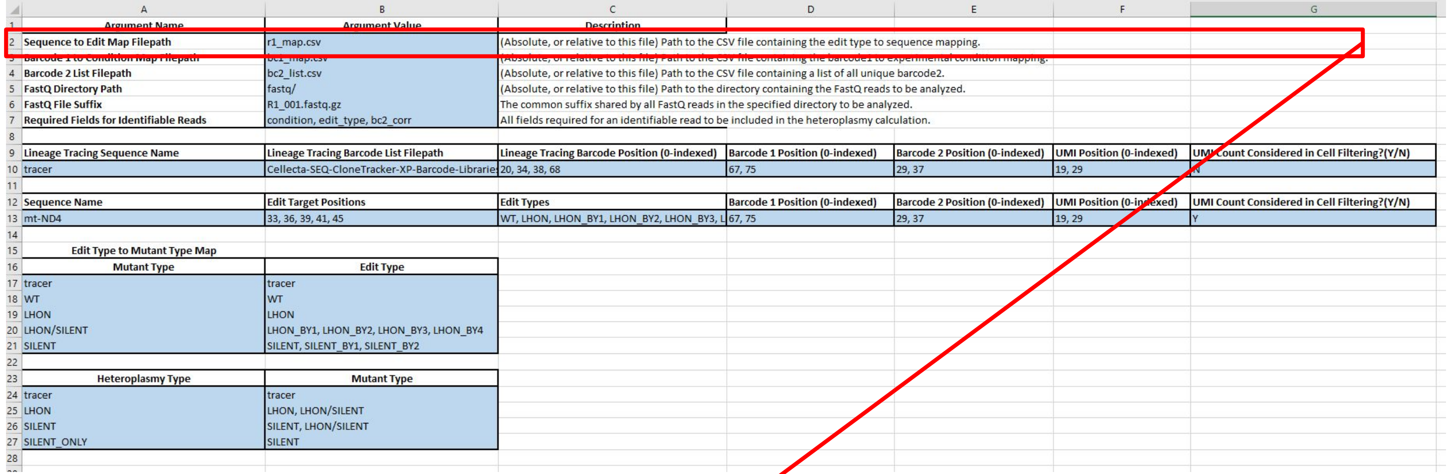
Required Input Files



	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Lineage Tracing Sequence Name**	**Lineage Tracing Barcode List Filepath**	**Lineage Tracing Barcode Position (0-indexed)**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	**Edit Type to Mutant Type Map**						
16	**Mutant Type**	**Edit Type**					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	**Heteroplasmy Type**	**Mutant Type**					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- construct-specific information
- edit -> mutation -> heteroplasmy type map

Required Input Files



Argument Name	Argument Value	Description
Sequence to Edit Map Filepath	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.
Barcode 1 to Condition Map Filepath	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.
Barcode 2 List Filepath	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.
FastQ Directory Path	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.
FastQ File Suffix	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.
Required Fields for Identifiable Reads	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.

Lineage Tracing Sequence Name	Lineage Tracing Barcode List Filepath	Lineage Tracing Barcode Position (0-indexed)	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N

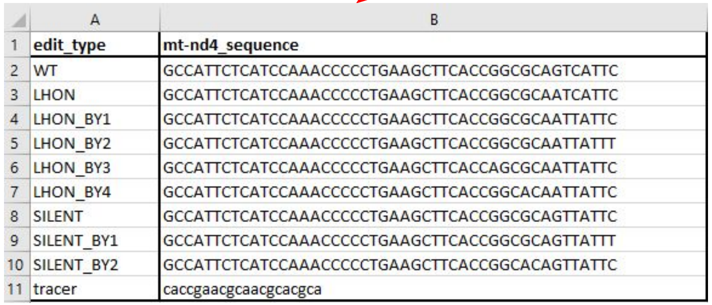
Sequence Name	Edit Target Positions	Edit Types	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y

Edit Type to Mutant Type Map

Mutant Type	Edit Type
tracer	tracer
WT	WT
LHON	LHON
LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4
SILENT	SILENT, SILENT_BY1, SILENT_BY2

Heteroplasmy Type	Mutant Type
tracer	tracer
LHON	LHON, LHON/SILENT
SILENT	SILENT, LHON/SILENT
SILENT_ONLY	SILENT

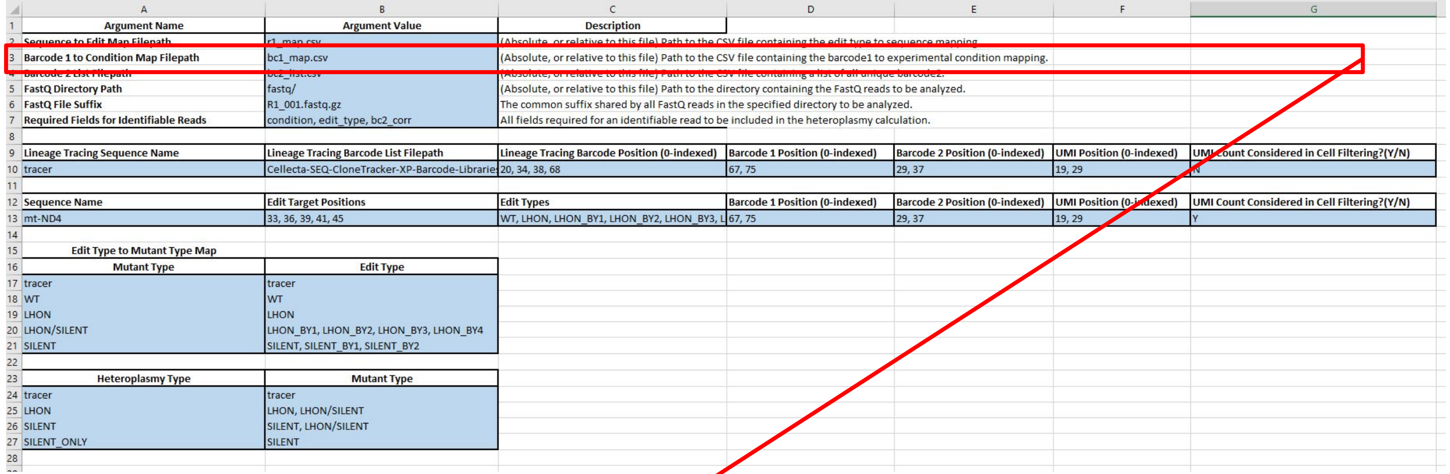
The first row does not have to literally be "edit_type, mt-nd4_sequence" – anything works as long as the edit type-sequence mapping starts at the second row, and the two columns follow the correct order



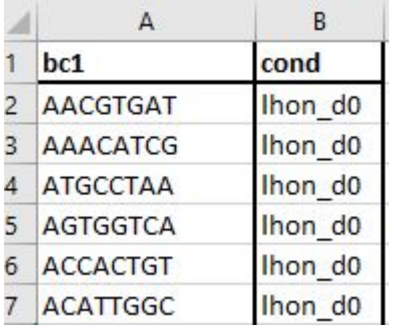
edit_type	mt-nd4_sequence
WT	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTCATTC
LHON	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATCATTC
LHON_BY1	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTC
LHON_BY2	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTT
LHON_BY3	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCAGCGCAATTATTC
LHON_BY4	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAATTATTC
SILENT	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTC
SILENT_BY1	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTT
SILENT_BY2	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAGTTATTC
tracer	caccgaacgcaacgcacgca

- experiment argument spreadsheet
- **read 1 target sequence to edit type map**
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- construct-specific information
- edit -> mutation -> heteroplasmy type map

Required Input Files



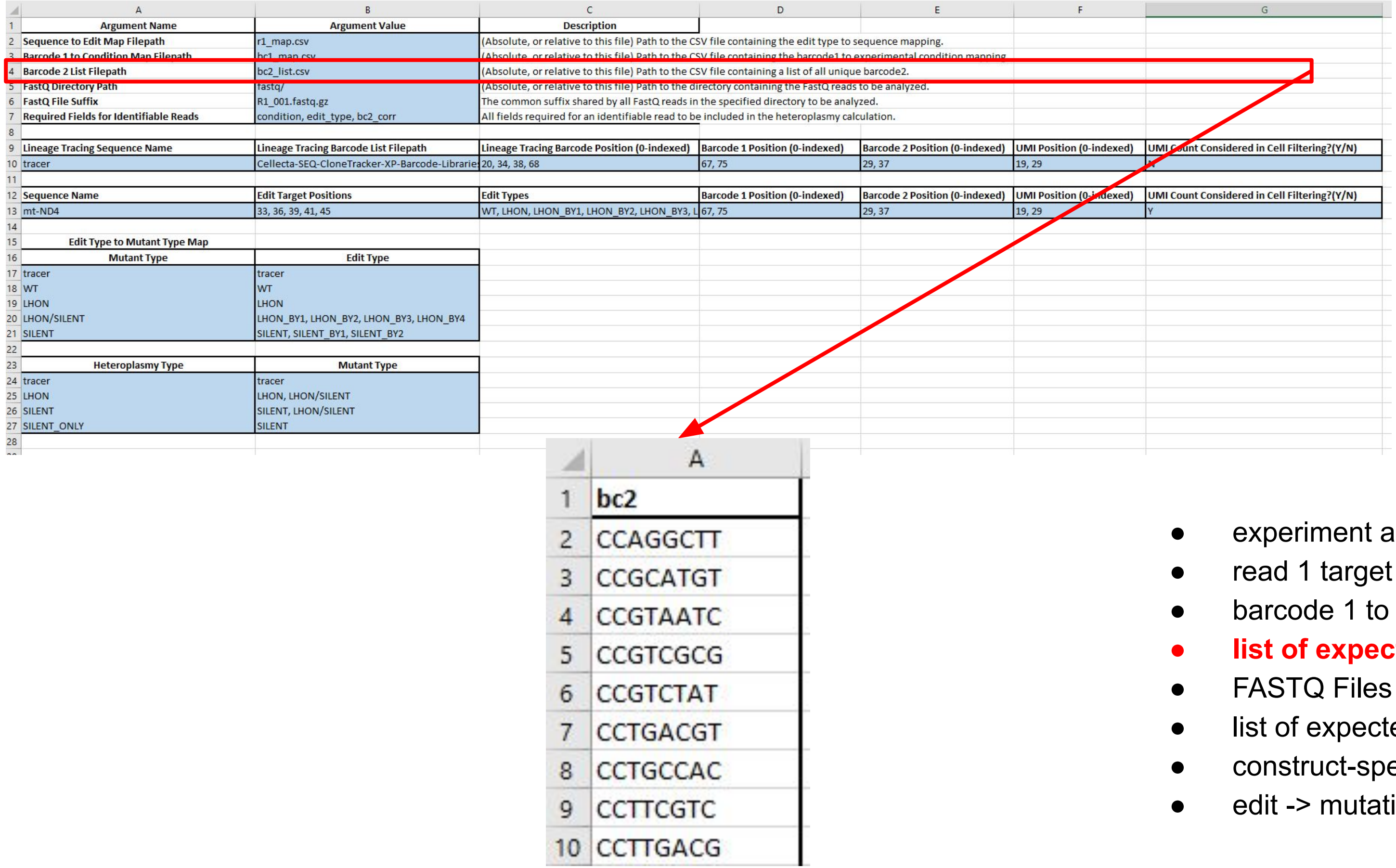
	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Lineage Tracing Sequence Name**	**Lineage Tracing Barcode List Filepath**	**Lineage Tracing Barcode Position (0-indexed)**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	**Edit Type to Mutant Type Map**						
16	**Mutant Type**	**Edit Type**					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	**Heteroplasmy Type**	**Mutant Type**					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							



	A	B
1	**bc1**	**cond**
2	AACGTGAT	lhon_d0
3	AAACATCG	lhon_d0
4	ATGCCTAA	lhon_d0
5	AGTGGTCA	lhon_d0
6	ACCACTGT	lhon_d0
7	ACATTGGC	lhon_d0

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- **barcode 1 to condition map**
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- construct-specific information
- edit -> mutation -> heteroplasmy type map

Required Input Files

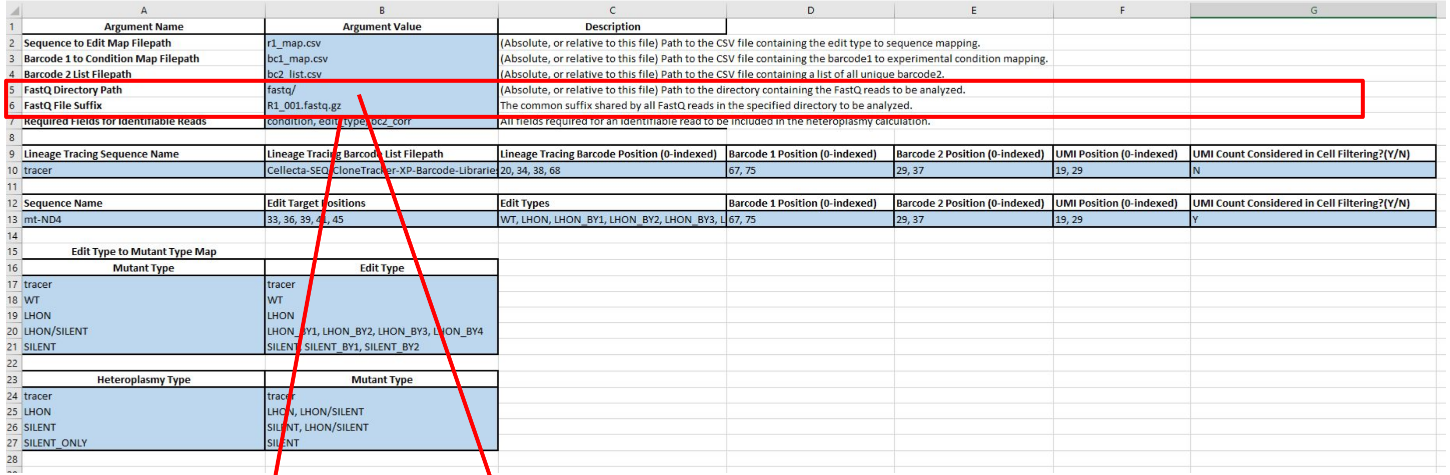


	A	B	C	D	E	F	G
1	Argument Name	Argument Value	Description				
2	Sequence to Edit Map Filepath	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	Barcode 1 to Condition Map Filepath	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	Barcode 2 List Filepath	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	FastQ Directory Path	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	FastQ File Suffix	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	Required Fields for Identifiable Reads	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	Lineage Tracing Sequence Name	Lineage Tracing Barcode List Filepath	Lineage Tracing Barcode Position (0-indexed)	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	Sequence Name	Edit Target Positions	Edit Types	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	Edit Type to Mutant Type Map						
16	Mutant Type	Edit Type					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	Heteroplasmy Type	Mutant Type					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							

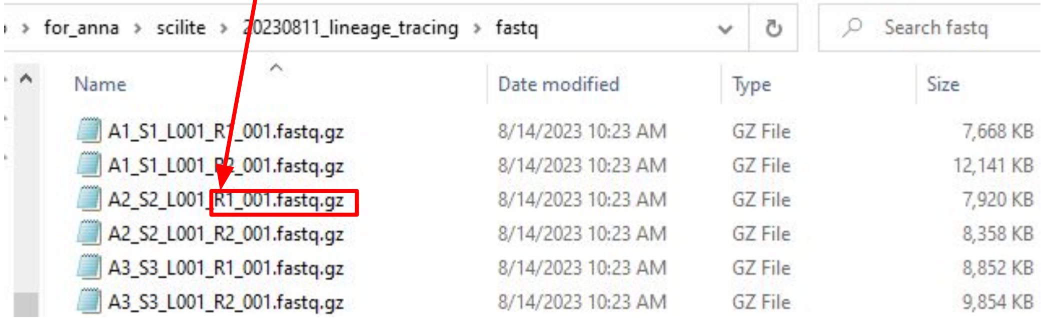
	A
1	bc2
2	CCAGGCTT
3	CCGCATGT
4	CCGTAATC
5	CCGTCGCG
6	CCGTCTAT
7	CCTGACGT
8	CCTGCCAC
9	CCTTCGTC
10	CCTTGACG

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- **list of expected barcode 2s**
- FASTQ Files
- list of expected ORF barcodes (if any)
- construct-specific information
- edit -> mutation -> heteroplasmy type map

Required Input Files



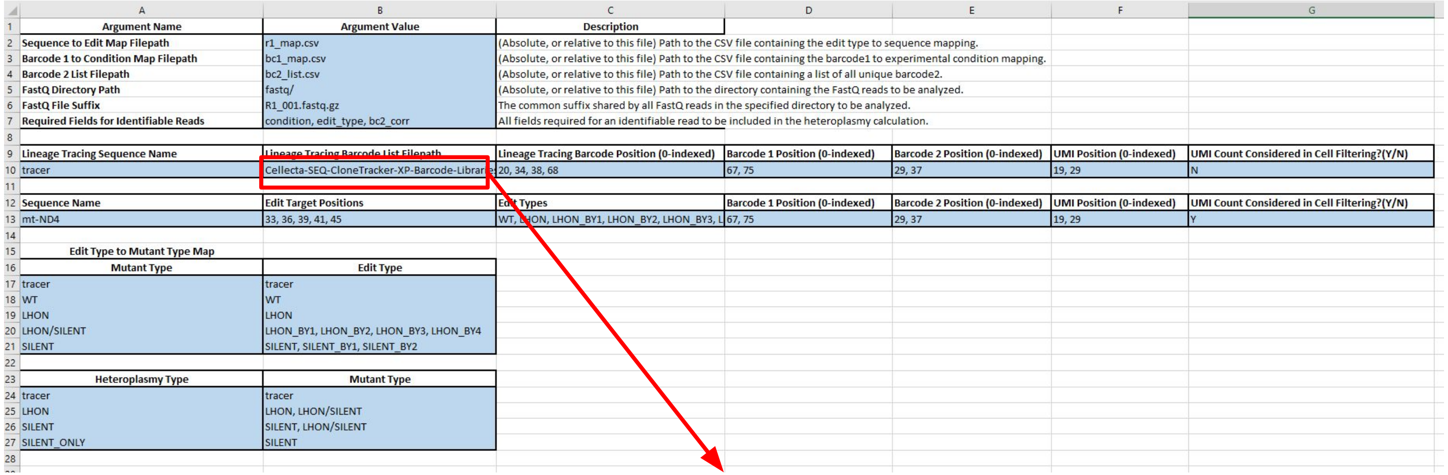
	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Lineage Tracing Sequence Name**	**Lineage Tracing Barcode List Filepath**	**Lineage Tracing Barcode Position (0-indexed)**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
10	tracer	Cellecta-SEQ CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
13	mt-ND4	33, 36, 39, 4[illegible], 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	**Edit Type to Mutant Type Map**						
16	**Mutant Type**	**Edit Type**					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	**Heteroplasmy Type**	**Mutant Type**					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							



Name	Date modified	Type	Size
A1_S1_L001_R1_001.fastq.gz	8/14/2023 10:23 AM	GZ File	7,668 KB
A1_S1_L001_R2_001.fastq.gz	8/14/2023 10:23 AM	GZ File	12,141 KB
A2_S2_L001_R1_001.fastq.gz	8/14/2023 10:23 AM	GZ File	7,920 KB
A2_S2_L001_R2_001.fastq.gz	8/14/2023 10:23 AM	GZ File	8,358 KB
A3_S3_L001_R1_001.fastq.gz	8/14/2023 10:23 AM	GZ File	8,852 KB
A3_S3_L001_R2_001.fastq.gz	8/14/2023 10:23 AM	GZ File	9,854 KB

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- **FASTQ Files**
- list of expected ORF barcodes (if any)
- construct-specific information
- edit -> mutation -> heteroplasmy type map

Required Input Files



Argument Name	Argument Value	Description
Sequence to Edit Map Filepath	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.
Barcode 1 to Condition Map Filepath	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.
Barcode 2 List Filepath	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.
FastQ Directory Path	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.
FastQ File Suffix	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.
Required Fields for Identifiable Reads	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.

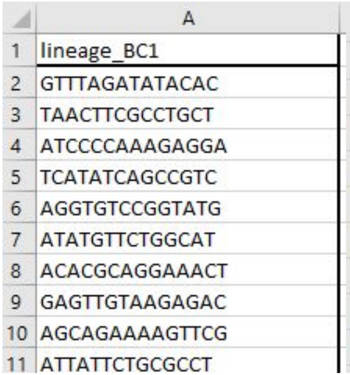
Lineage Tracing Sequence Name	Lineage Tracing Barcode List Filepath	Lineage Tracing Barcode Position (0-indexed)	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librar	20, 34, 38, 68	67, 75	29, 37	19, 29	N

Sequence Name	Edit Target Positions	Edit Types	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y

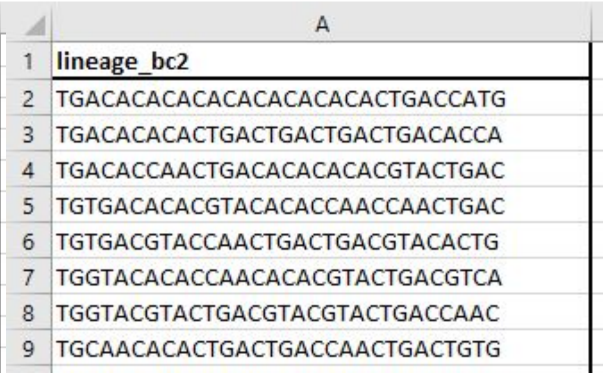
Edit Type to Mutant Type Map

Mutant Type	Edit Type
tracer	tracer
WT	WT
LHON	LHON
LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4
SILENT	SILENT, SILENT_BY1, SILENT_BY2

Heteroplasmy Type	Mutant Type
tracer	tracer
LHON	LHON, LHON/SILENT
SILENT	SILENT, LHON/SILENT
SILENT_ONLY	SILENT



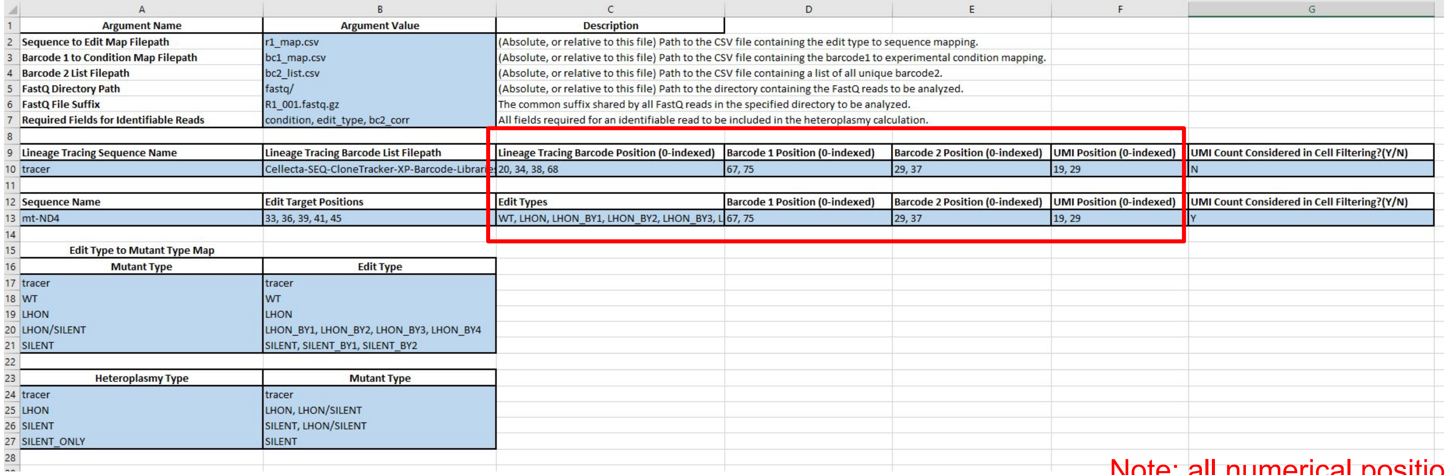
lineage_BC1
GTTTAGATATACAC
TAACTTCGCCTGCT
ATCCCCAAAGAGGA
TCATATCAGCCGTC
AGGTGTCCGGTATG
ATATGTTCTGGCAT
ACACGCAGGAAACT
GAGTTGTAAGAGAC
AGCAGAAAAGTTCG
ATTATTCTGCGCCT



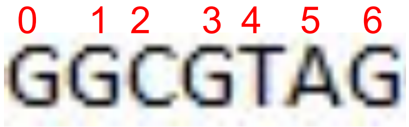
lineage_bc2
TGACACACACACACACACACTGACCATG
TGACACACACTGACTGACTGACTGACACCA
TGACACCAACTGACACACACACGTACTGAC
TGTGACACACGTACACACCAACCAACTGAC
TGTGACGTACCAACTGACTGACGTACACTG
TGGTACACACCAACACACGTACTGACGTCA
TGGTACGTACTGACGTACGTACTGACCAAC
TGCAACACACTGACTGACCAACTGACTGTG

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- **list of expected lineage tracing barcodes (if any)**
- construct-specific information
- edit -> mutation -> heteroplasmy type map

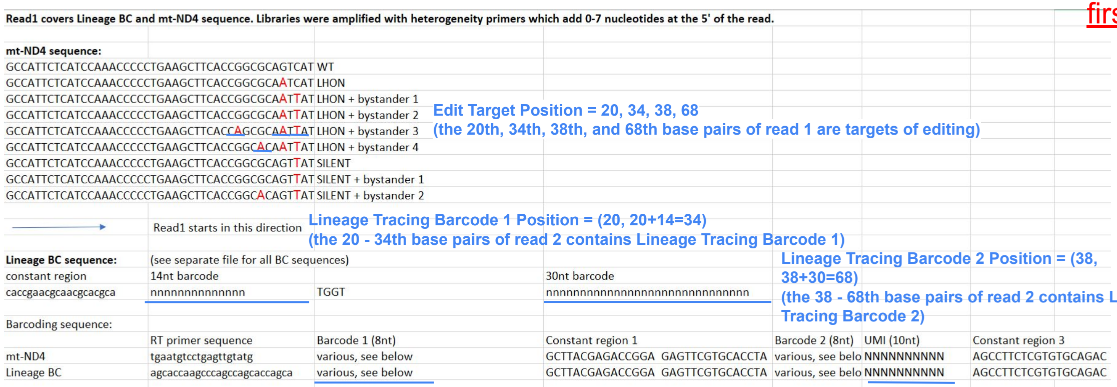
Required Input Files



	A	B	C	D	E	F	G
1	Argument Name	Argument Value	Description				
2	Sequence to Edit Map Filepath	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	Barcode 1 to Condition Map Filepath	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	Barcode 2 List Filepath	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	FastQ Directory Path	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	FastQ File Suffix	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	Required Fields for Identifiable Reads	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	Lineage Tracing Sequence Name	Lineage Tracing Barcode List Filepath	Lineage Tracing Barcode Position (0-indexed)	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	Sequence Name	Edit Target Positions	Edit Types	Barcode 1 Position (0-indexed)	Barcode 2 Position (0-indexed)	UMI Position (0-indexed)	UMI Count Considered in Cell Filtering?(Y/N)
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	Edit Type to Mutant Type Map						
16	Mutant Type	Edit Type					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	Heteroplasmy Type	Mutant Type					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							



Note: all numerical positions are 0-indexed – meaning the first base pair is indexed as the 0th, the second base pair is the 1st, the third is the 2nd, and so on…



Read1 covers Lineage BC and mt-ND4 sequence. Libraries were amplified with heterogeneity primers which add 0-7 nucleotides at the 5' of the read.

mt-ND4 sequence:

GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTCAT	WT
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATCAT	LHON
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTAT	LHON + bystander 1
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTAT	LHON + bystander 2
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCAGCGCAATTAT	LHON + bystander 3
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAATTAT	LHON + bystander 4
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTAT	SILENT
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTAT	SILENT + bystander 1
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAGTTAT	SILENT + bystander 2

Edit Target Position = 20, 34, 38, 68
(the 20th, 34th, 38th, and 68th base pairs of read 1 are targets of editing)

Read1 starts in this direction

Lineage Tracing Barcode 1 Position = (20, 20+14=34)
(the 20 - 34th base pairs of read 2 contains Lineage Tracing Barcode 1)

Lineage BC sequence:	(see separate file for all BC sequences)		
constant region	14nt barcode		30nt barcode
caccgaacgcaacgcacgca	nnnnnnnnnnnnnn	TGGT	nnnnnnnnnnnnnnnnnnnnnnnnnnnnnn

Lineage Tracing Barcode 2 Position = (38, 38+30=68)
(the 38 - 68th base pairs of read 2 contains Lineage Tracing Barcode 2)

Barcoding sequence:

	RT primer sequence	Barcode 1 (8nt)	Constant region 1	Barcode 2 (8nt)	UMI (10nt)	Constant region 3
mt-ND4	tgaatgtcctgagttgtatg	various, see below	GCTTACGAGACCGGA GAGTTCGTGCACCTA	various, see belo	NNNNNNNNNN	AGCCTTCTCGTGTGCAGAC
Lineage BC	agcaccaagcccagccagcaccagca	various, see below	GCTTACGAGACCGGA GAGTTCGTGCACCTA	various, see belo	NNNNNNNNNN	AGCCTTCTCGTGTGCAGAC

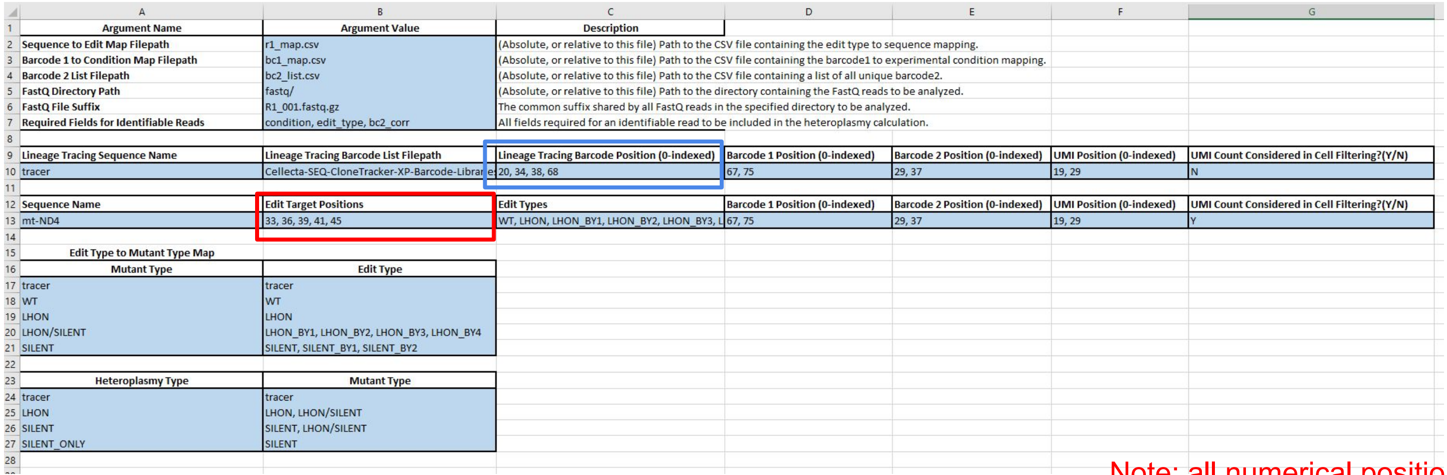
Read 2 starts in this diretion

Barcode 1 Position = (67, 67+8=75)
(the 67 - 75th base pairs of read 2 contains Barcode 1)

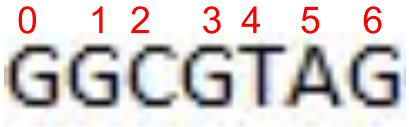
UMI Position = (19, 19+10=29)
(the 19 - 29th base pairs of read 2 contains UMI)

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- **construct-specific information**
- edit -> mutation -> heteroplasmy type map

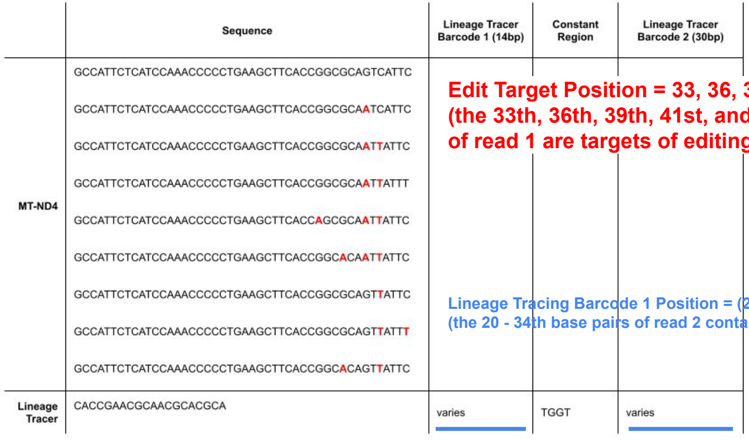
Required Input Files



	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Lineage Tracing Sequence Name**	**Lineage Tracing Barcode List Filepath**	**Lineage Tracing Barcode Position (0-indexed)**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librar	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	**Edit Type to Mutant Type Map**						
16	**Mutant Type**	**Edit Type**					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	**Heteroplasmy Type**	**Mutant Type**					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							



Note: all numerical positions are 0-indexed – meaning the first base pair is indexed as the 0th, the second base pair is the 1st, the third is the 2nd, and so on…



	Sequence	Lineage Tracer Barcode 1 (14bp)	Constant Region	Lineage Tracer Barcode 2 (30bp)
MT-ND4	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTCATTC GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATCATTC GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTC GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTT GCCATTCTCATCCAAACCCCCTGAAGCTTCACCAGCGCAATTATTC GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAATTATTC GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTC GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTT GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAGTTATTC			
Lineage Tracer	CACCGAACGCAACGCACGCA	varies	TGGT	varies

Edit Target Position = 33, 36, 39, 41, 45 (the 33th, 36th, 39th, 41st, and 45th base pairs of read 1 are targets of editing)

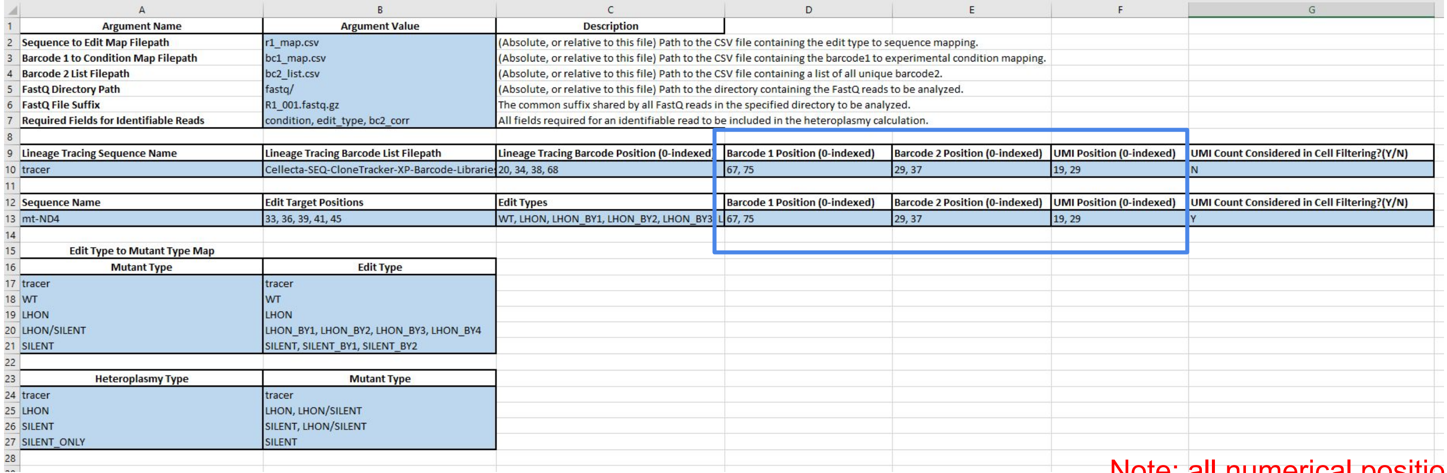
Lineage Tracing Barcode 1 Position = (20, 20+14=34) (the 20 - 34th base pairs of read 2 contains Lineage Tracing Barcode 1)

Lineage Tracing Barcode 2 Position = (38, 38+30=68) (the 38 - 68th base pairs of read 2 contains Lineage Tracing Barcode 2)

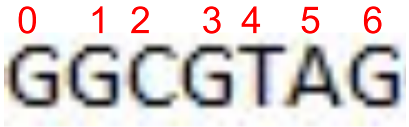
Read 1 starts in this direction

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- **construct-specific information**
- edit -> mutation -> heteroplasmy type map

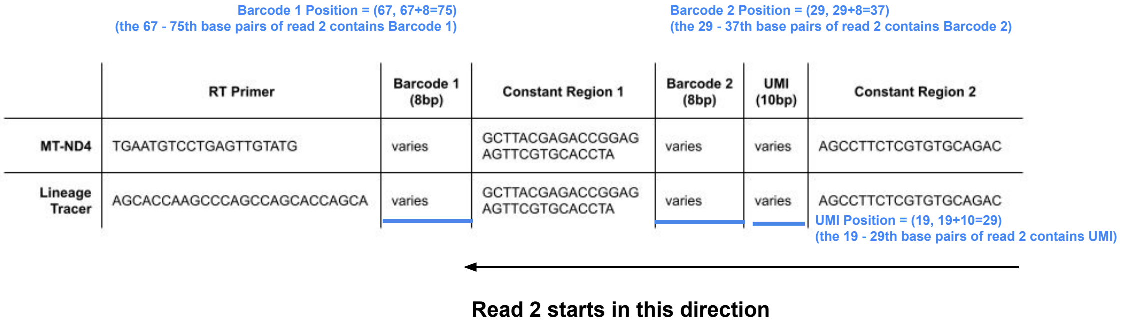
Required Input Files



	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Lineage Tracing Sequence Name**	**Lineage Tracing Barcode List Filepath**	**Lineage Tracing Barcode Position (0-indexed)**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3 L	67, 75	29, 37	19, 29	Y
14							
15	**Edit Type to Mutant Type Map**						
16	**Mutant Type**	**Edit Type**					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	**Heteroplasmy Type**	**Mutant Type**					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							



Note: all numerical positions are 0-indexed – meaning the first base pair is indexed as the 0th, the second base pair is the 1st, the third is the 2nd, and so on…



Barcode 1 Position = (67, 67+8=75)
(the 67 - 75th base pairs of read 2 contains Barcode 1)

Barcode 2 Position = (29, 29+8=37)
(the 29 - 37th base pairs of read 2 contains Barcode 2)

	RT Primer	Barcode 1 (8bp)	Constant Region 1	Barcode 2 (8bp)	UMI (10bp)	Constant Region 2
MT-ND4	TGAATGTCCTGAGTTGTATG	varies	GCTTACGAGACCGGAG AGTTCGTGCACCTA	varies	varies	AGCCTTCTCGTGTGCAGAC
Lineage Tracer	AGCACCAAGCCCAGCCAGCACCAGCA	varies	GCTTACGAGACCGGAG AGTTCGTGCACCTA	varies	varies	AGCCTTCTCGTGTGCAGAC

UMI Position = (19, 19+10=29)
(the 19 - 29th base pairs of read 2 contains UMI)

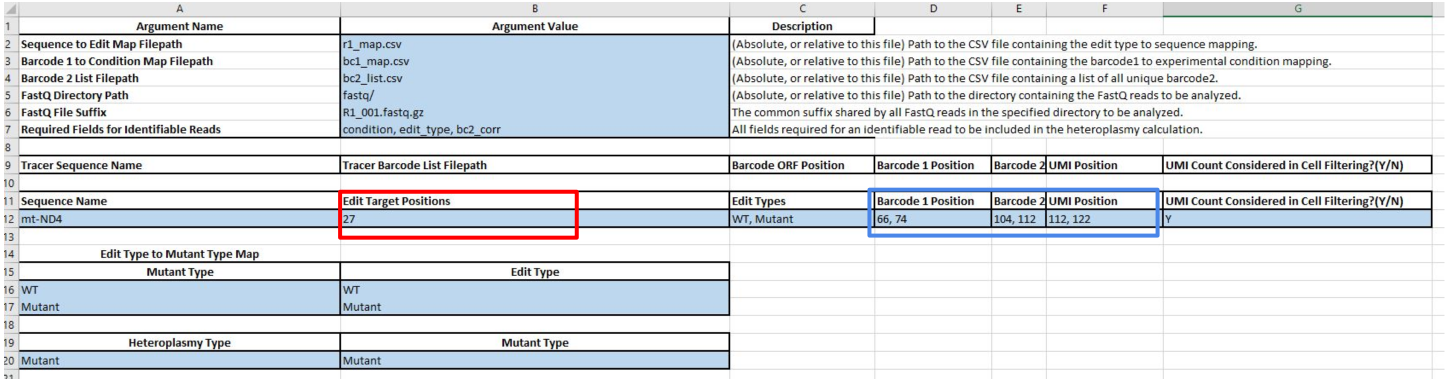
Read 2 starts in this direction

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- **construct-specific information**
- edit -> mutation -> heteroplasmy type map

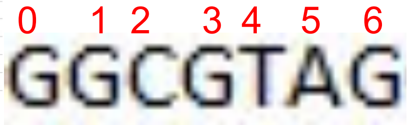
	Sequence	**Lineage Tracer Barcode 1 (14bp)**	**Constant Region**	**Lineage Tracer Barcode 2 (30bp)**
MT-ND4	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTCATTC			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCA**A**TCATTC			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCA**A**T**T**ATTC			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCA**A**T**T**ATTT			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACC**A**GCGCA**A**T**T**ATTC			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGC**A**CA**A**T**T**ATTC			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGT**T**ATTC			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGT**T**ATT**T**			
	GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGC**A**CAGT**T**ATTC			
Lineage Tracer	CACCGAACGCAACGCACGCA	varies	TGGT	varies

	RT Primer	Barcode 1 (8bp)	Constant Region 1	Barcode 2 (8bp)	UMI (10bp)	Constant Region 2
MT-ND4	TGAATGTCCTGAGTTGTATG	varies	GCTTACGAGACCGGAGAGTTCGTGCACCTA	varies	varies	AGCCTTCTCGTGTGCAGAC
Lineage Tracer	AGCACCAAGCCCAGCCAGCACCAGCA	varies	GCTTACGAGACCGGAGAGTTCGTGCACCTA	varies	varies	AGCCTTCTCGTGTGCAGAC

Required Input Files



	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Tracer Sequence Name**	**Tracer Barcode List Filepath**	**Barcode ORF Position**	**Barcode 1 Position**	**Barcode 2**	**UMI Position**	**UMI Count Considered in Cell Filtering?(Y/N)**
10							
11	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position**	**Barcode 2**	**UMI Position**	**UMI Count Considered in Cell Filtering?(Y/N)**
12	mt-ND4	27	WT, Mutant	66, 74	104, 112	112, 122	Y
13							
14	**Edit Type to Mutant Type Map**						
15	**Mutant Type**	**Edit Type**					
16	WT	WT					
17	Mutant	Mutant					
18							
19	**Heteroplasmy Type**	**Mutant Type**					
20	Mutant	Mutant					



Barcode 2 Position = (104, 104+8=112)
(the 104 - 112th base pairs of read 1 contains Barcode 2)

Note: all numerical positions are 0-indexed – meaning the first base pair is indexed as the 0th, the second base pair is the 1st, the third is the 2nd, and so on…

Edit Target Position = 27
(the 27th base pair of read 1 are targets of editing)

Sequence	RT Primer Sequence	Barcode 1 (8bp)	Constant Region 1	Barcode 2 (8bp)	UMI (10bp)	Constant Region 2
TCTCTGTGCTAGTAACCACGTTCTCCTGATCAAATATCACTCTCCT	ACTTACAGGA CTCAACATAC	varies	CGAATGCT CTGGCCTC TCAAGCAC GTGGAT	varies	varies	TCGGAAGAGC ACACGTCTG
TCTCTGTGCTAGTAACCACGTTCTCCTAATCAAATATCACTCTCCT						

Barcode 1 Position = (66, 66+8=74)
(the 66 - 74th base pairs of read 1 contains Barcode 1)

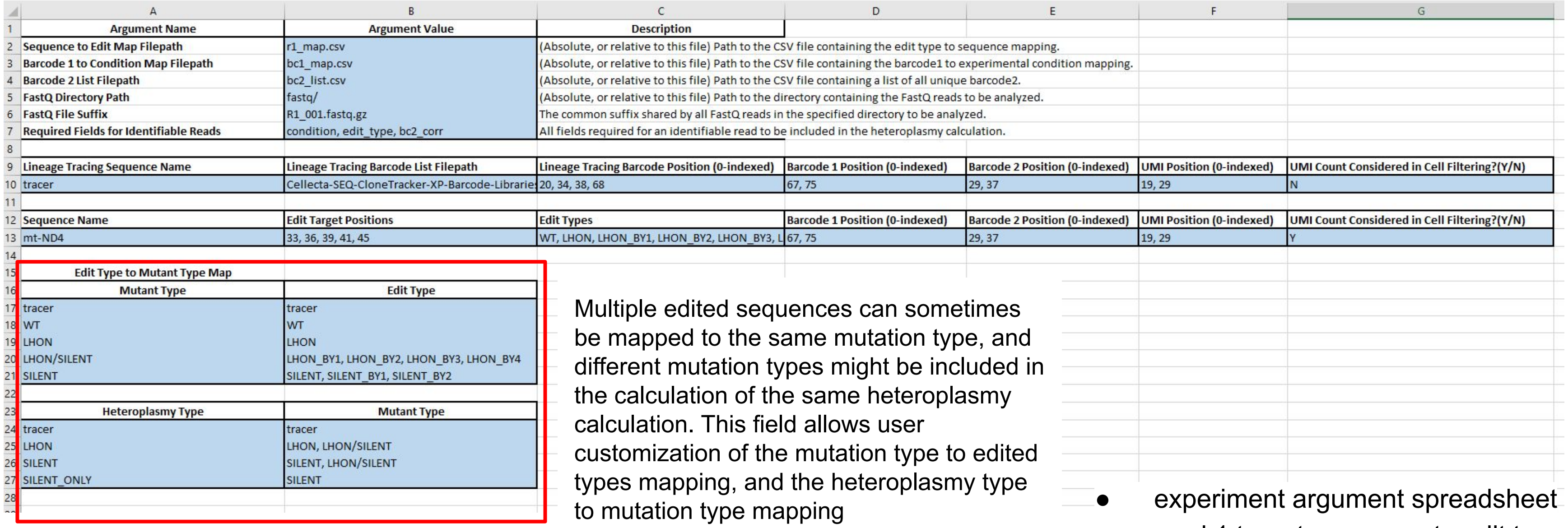
UMI Position = (112, 112+10=122)
(the 112 - 122th base pairs of read 1 contains UMI)

→ **Read 1 starts in this direction**

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- **construct-specific information**
- edit -> mutation -> heteroplasmy type map

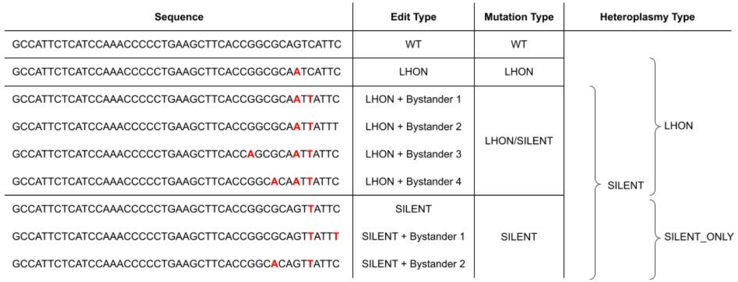
Sequence	RT Primer Sequence	Barcode 1 (8bp)	Constant Region 1	Barcode 2 (8bp)	UMI (10bp)	Constant Region 2
TCTCTGTGCTAGTAACCACGTTCTCCT**G**ATCAAATATCACTCTCCT	ACTTACAGGACTCAACATAC	varies	CGAATGCTCTGGCCTCTCAAGCACGTGGAT	varies	varies	TCGGAAGAGCACACGTCTG
TCTCTGTGCTAGTAACCACGTTCTCCT**A**ATCAAATATCACTCTCCT						

Required Input Files



	A	B	C	D	E	F	G
1	**Argument Name**	**Argument Value**	**Description**				
2	**Sequence to Edit Map Filepath**	r1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the edit type to sequence mapping.				
3	**Barcode 1 to Condition Map Filepath**	bc1_map.csv	(Absolute, or relative to this file) Path to the CSV file containing the barcode1 to experimental condition mapping.				
4	**Barcode 2 List Filepath**	bc2_list.csv	(Absolute, or relative to this file) Path to the CSV file containing a list of all unique barcode2.				
5	**FastQ Directory Path**	fastq/	(Absolute, or relative to this file) Path to the directory containing the FastQ reads to be analyzed.				
6	**FastQ File Suffix**	R1_001.fastq.gz	The common suffix shared by all FastQ reads in the specified directory to be analyzed.				
7	**Required Fields for Identifiable Reads**	condition, edit_type, bc2_corr	All fields required for an identifiable read to be included in the heteroplasmy calculation.				
8							
9	**Lineage Tracing Sequence Name**	**Lineage Tracing Barcode List Filepath**	**Lineage Tracing Barcode Position (0-indexed)**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
10	tracer	Cellecta-SEQ-CloneTracker-XP-Barcode-Librarie	20, 34, 38, 68	67, 75	29, 37	19, 29	N
11							
12	**Sequence Name**	**Edit Target Positions**	**Edit Types**	**Barcode 1 Position (0-indexed)**	**Barcode 2 Position (0-indexed)**	**UMI Position (0-indexed)**	**UMI Count Considered in Cell Filtering?(Y/N)**
13	mt-ND4	33, 36, 39, 41, 45	WT, LHON, LHON_BY1, LHON_BY2, LHON_BY3, L	67, 75	29, 37	19, 29	Y
14							
15	**Edit Type to Mutant Type Map**						
16	**Mutant Type**	**Edit Type**					
17	tracer	tracer					
18	WT	WT					
19	LHON	LHON					
20	LHON/SILENT	LHON_BY1, LHON_BY2, LHON_BY3, LHON_BY4					
21	SILENT	SILENT, SILENT_BY1, SILENT_BY2					
22							
23	**Heteroplasmy Type**	**Mutant Type**					
24	tracer	tracer					
25	LHON	LHON, LHON/SILENT					
26	SILENT	SILENT, LHON/SILENT					
27	SILENT_ONLY	SILENT					
28							

Multiple edited sequences can sometimes be mapped to the same mutation type, and different mutation types might be included in the calculation of the same heteroplasmy calculation. This field allows user customization of the mutation type to edited types mapping, and the heteroplasmy type to mutation type mapping



Sequence	Edit Type	Mutation Type	Heteroplasmy Type
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTCATTC	WT	WT	
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATCATTC	LHON	LHON	
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTC	LHON + Bystander 1	LHON/SILENT	
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTT	LHON + Bystander 2		LHON
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCAGCGCAATTATTC	LHON + Bystander 3		
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAATTATTC	LHON + Bystander 4		SILENT
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTC	SILENT	SILENT	
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTT	SILENT + Bystander 1		SILENT_ONLY
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAGTTATTC	SILENT + Bystander 2		

- experiment argument spreadsheet
- read 1 target sequence to edit type map
- barcode 1 to condition map
- list of expected barcode 2s
- FASTQ Files
- list of expected ORF barcodes (if any)
- construct-specific information
- **edit -> mutation -> heteroplasmy type map**

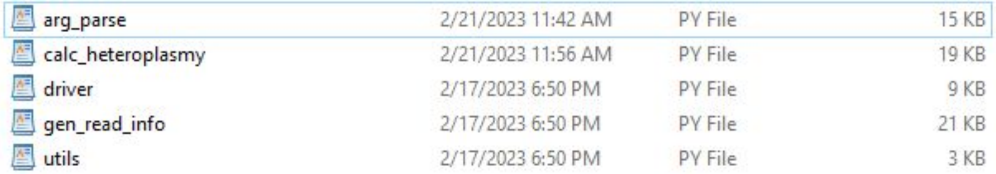
Default Edit -> Mutation -> Heteroplasmy Type Mapping of LHON constructs in Kotrys et al 2023 paper

Sequence	Edit Type	Mutation Type	Heteroplasmy Type
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTCATTC	WT	WT	
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATCATTC	LHON	LHON	LHON
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTC	LHON + Bystander 1	LHON/SILENT	SILENT
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAATTATTT	LHON + Bystander 2		
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCAGCGCAATTATTC	LHON + Bystander 3		
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAATTATTC	LHON + Bystander 4		
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTC	SILENT	SILENT	SILENT_ONLY
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCGCAGTTATTT	SILENT + Bystander 1		
GCCATTCTCATCCAAACCCCCTGAAGCTTCACCGGCACAGTTATTC	SILENT + Bystander 2		

Example Command-line Executions

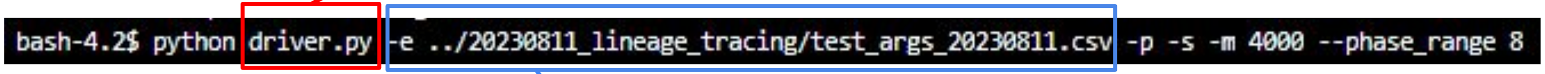
Required Dependencies (all are conda-installable):

- Python >= 3.9.15
- Numpy >= 1.19.5
- Pandas >= 1.5.2
- PySAM >= 0.20.0
- Matplotlib

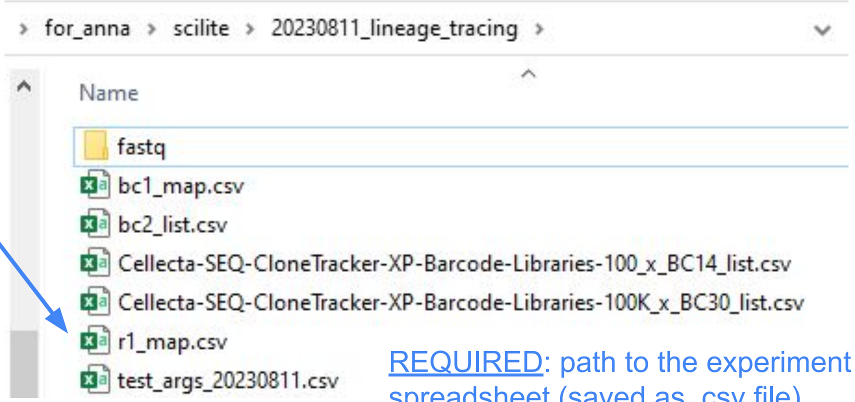


arg_parse	2/21/2023 11:42 AM	PY File	15 KB
calc_heteroplasmy	2/21/2023 11:56 AM	PY File	19 KB
driver	2/17/2023 6:50 PM	PY File	9 KB
gen_read_info	2/17/2023 6:50 PM	PY File	21 KB
utils	2/17/2023 6:50 PM	PY File	3 KB

REQUIRED: path to the main pipeline script (driver.py)



```
bash-4.2$ python driver.py -e ../20230811_lineage_tracing/test_args_20230811.csv -p -s -m 4000 --phase_range 8
```



› for_anna › scilite › 20230811_lineage_tracing ›

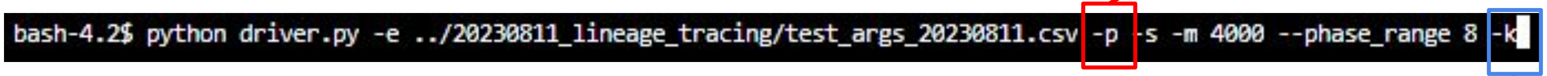
Name

- fastq
- bc1_map.csv
- bc2_list.csv
- Cellecta-SEQ-CloneTracker-XP-Barcode-Libraries-100_x_BC14_list.csv
- Cellecta-SEQ-CloneTracker-XP-Barcode-Libraries-100K_x_BC30_list.csv
- r1_map.csv
- test_args_20230811.csv

REQUIRED: path to the experiment argument spreadsheet (saved as .csv file)

Example Command-line Executions

OPTIONAL: “-p” specifies that this experiment produces paired-reads

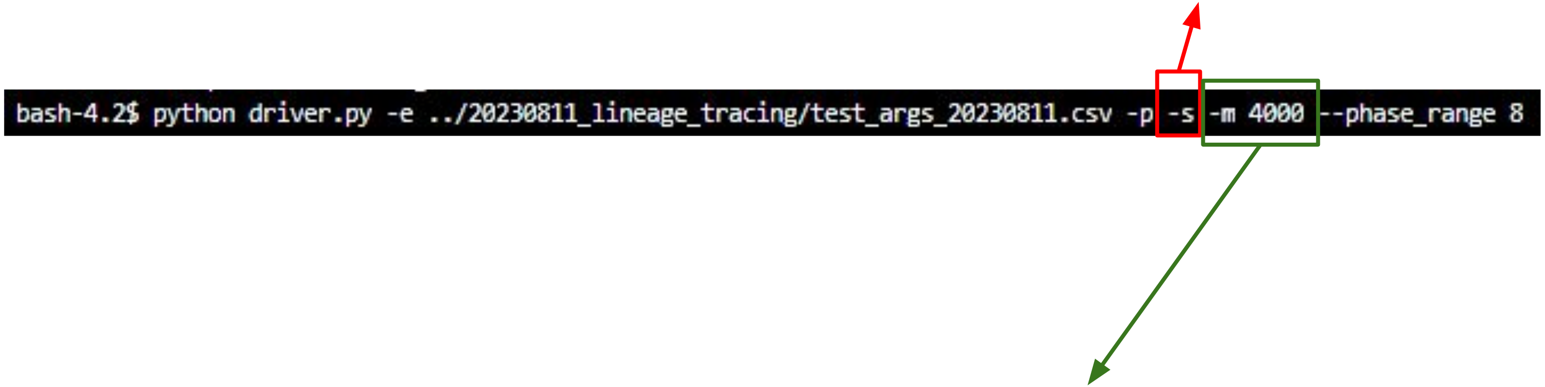


```
bash-4.2$ python driver.py -e ../20230811_lineage_tracing/test_args_20230811.csv -p -s -m 4000 --phase_range 8 -k
```

OPTIONAL: “-k” specifies that the knee plot used for filtering out cells with too few UMIs will NOT be visualized and saved as an output file – “knee_plot.pdf”

Example Command-line Executions

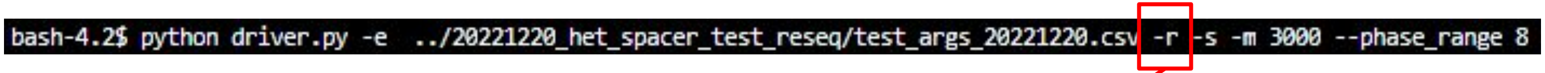
OPTIONAL: “-s” specifies that intermediate read-level and UMI-level spreadsheets will also be saved as output files – “read/UMI_dataframe.csv”



```
bash-4.2$ python driver.py -e ../20230811_lineage_tracing/test_args_20230811.csv -p -s -m 4000 --phase_range 8
```

OPTIONAL: “-m [insert number of cells]” specifies the maximum number of cells allowed to be preserved in the filtering-by-knee-plot process. The final resulting heteroplasmy spreadsheet may have fewer cells than specified, but never more. By default (not having “-m” specified), the maximum number of cells will just be the number of unique cells identified from all reads.

Example Command-line Executions



```
bash-4.2$ python driver.py -e  ../20221220_het_spacer_test_reseq/test_args_20221220.csv -r -s -m 3000 --phase_range 8
```

<u>OPTIONAL</u>: “-r” specifies the ID sequences extracted from the FASTQ reads (barcode 1s, barcode 2s, etc.) should be reverse-complemented to match their expected sequences defined in bc1_map.csv, bc2_list.csv, etc.

This flag needs to be specified in this experiment, because its FASTQ datasets are single-reads while its barcode 1 / 2s were defined in bc1_map.csv / bc2_list.csv as if they were read from the other end like in a paired-end experiment – but this is not necessarily true for all single-read experiments