CloudMap: A Cloud-based Pipeline for Analysis of Mutant Genome Sequences

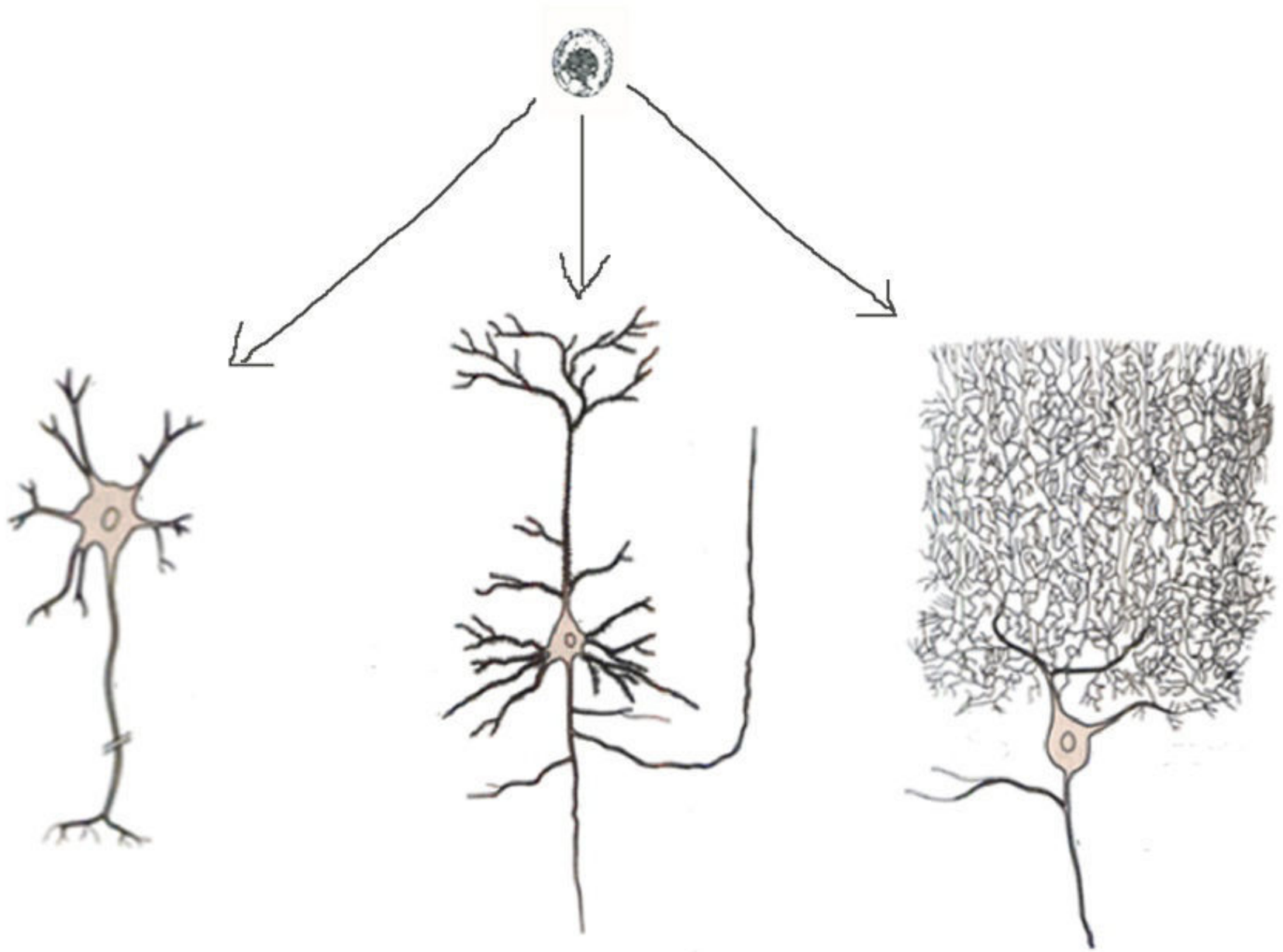
https://test.g2.bx.psu.edu/u/gal40/p/cloudmap

Gregory Minevich[1], Danny S. Park[1], Richard Poole[1], Daniel Blankenberg[2], Anton Nekrutenko[2], Oliver Hobert[1]

[1] Department of Biochemistry and Molecular Biophysics, HHMI Columbia University

[2] Center for Comparative Genomics and Bioinformatics, Penn State University

Identifying molecular mechanisms that create neuronal diversity



Motor neuron of spinal cord

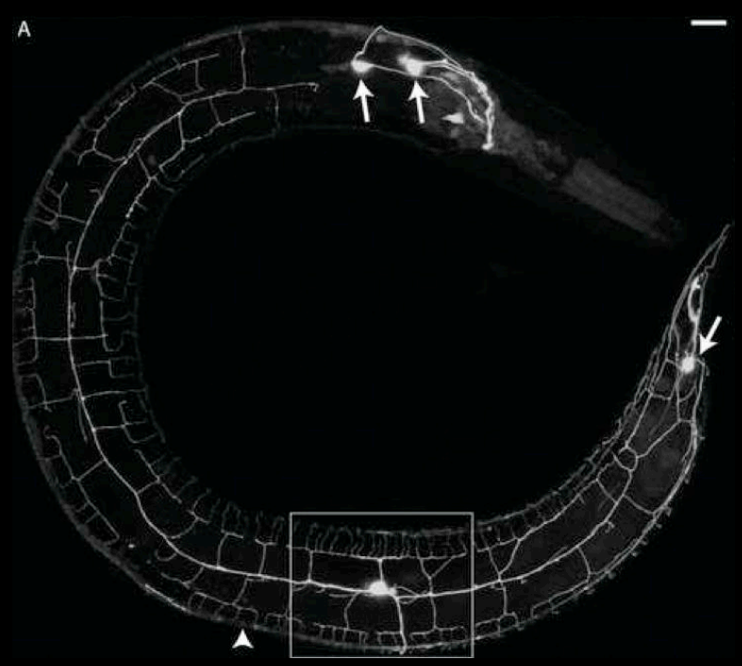
Pyramidal cell of hippocampus

Purkinje cell of cerebellum

adapted from Kandel, 2000

C. elegans: an ideal model organism

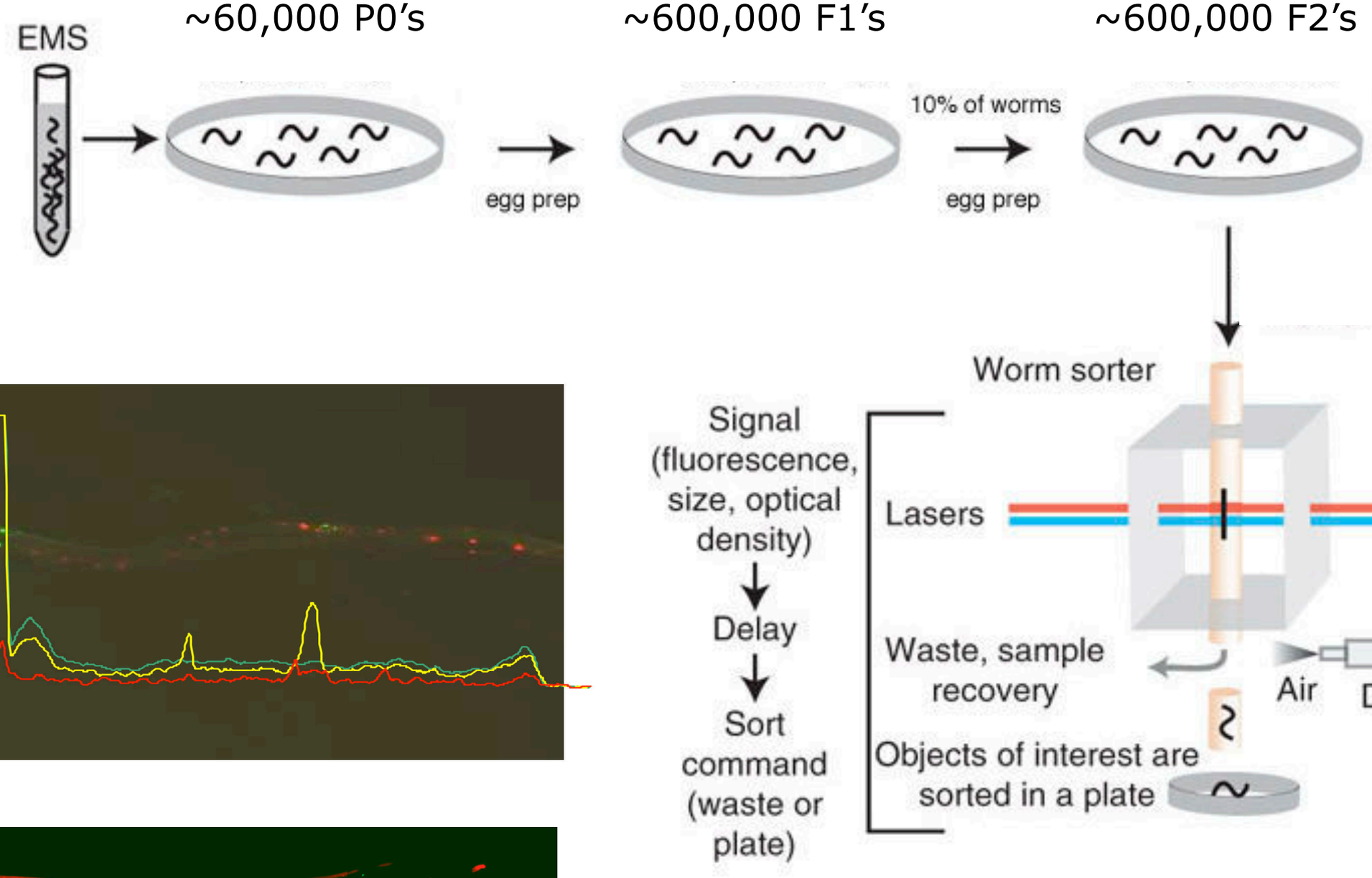
Wormatlas

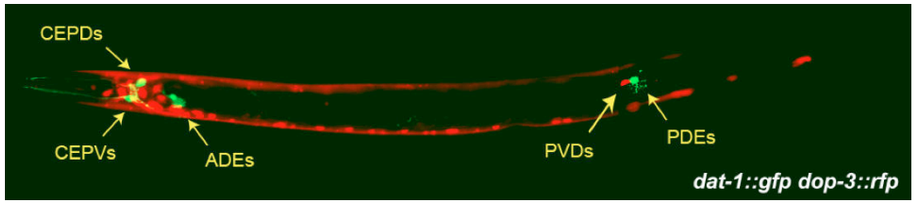


Smith et al. 2010

- Short generation time, 1mm long
- Hermaphrodites (w/ some males)
- Complete wiring diagram of nervous system
- Developmental fate of all 959 somatic cells known, cell position also constant
- Easy to make transgenic animals
- Genome sequenced in >5 closely related species
- Transparency allows for in vivo studies using reporter genes
- 20,470 protein coding genes

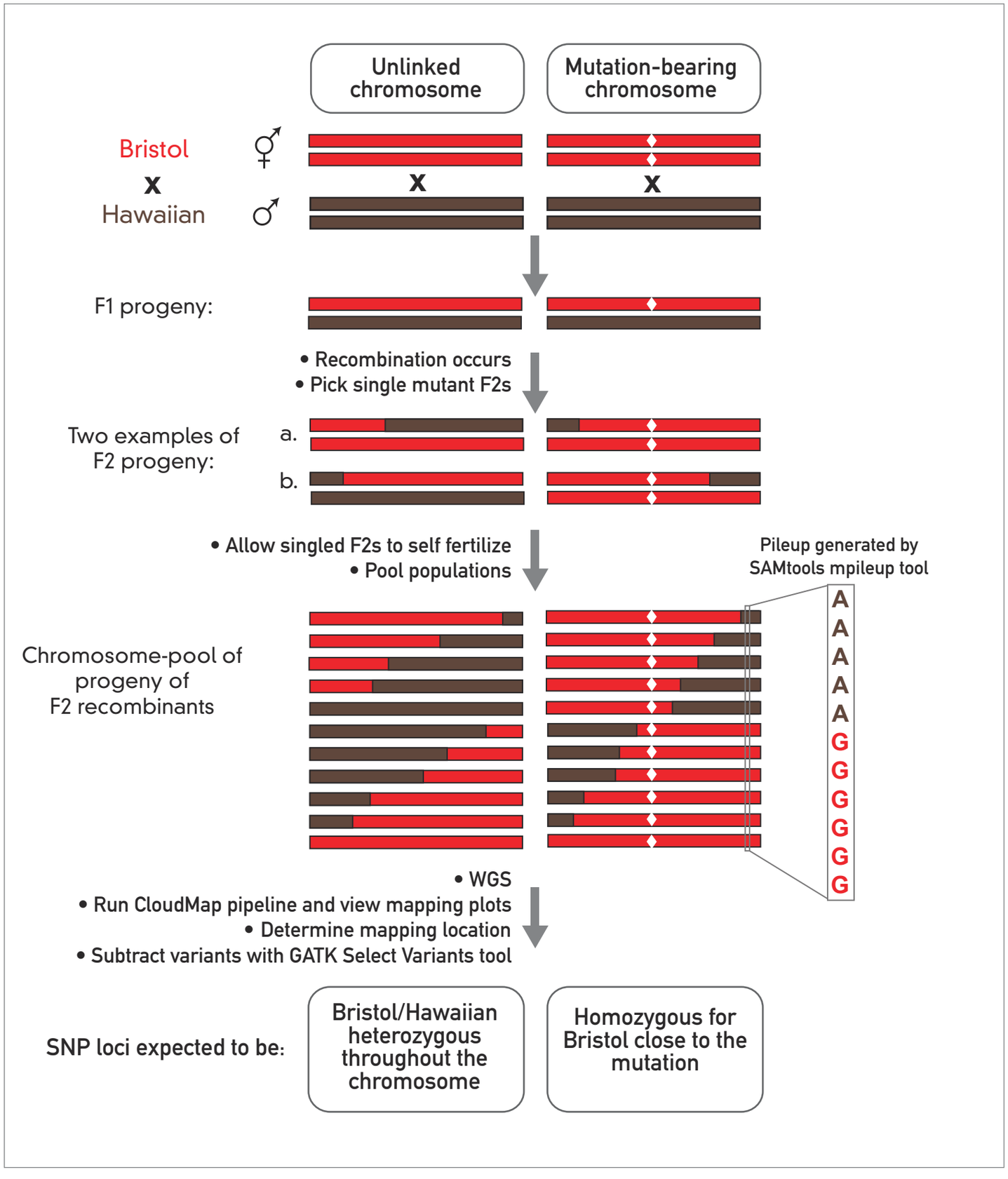
Forward genetic screens using a wormsorter



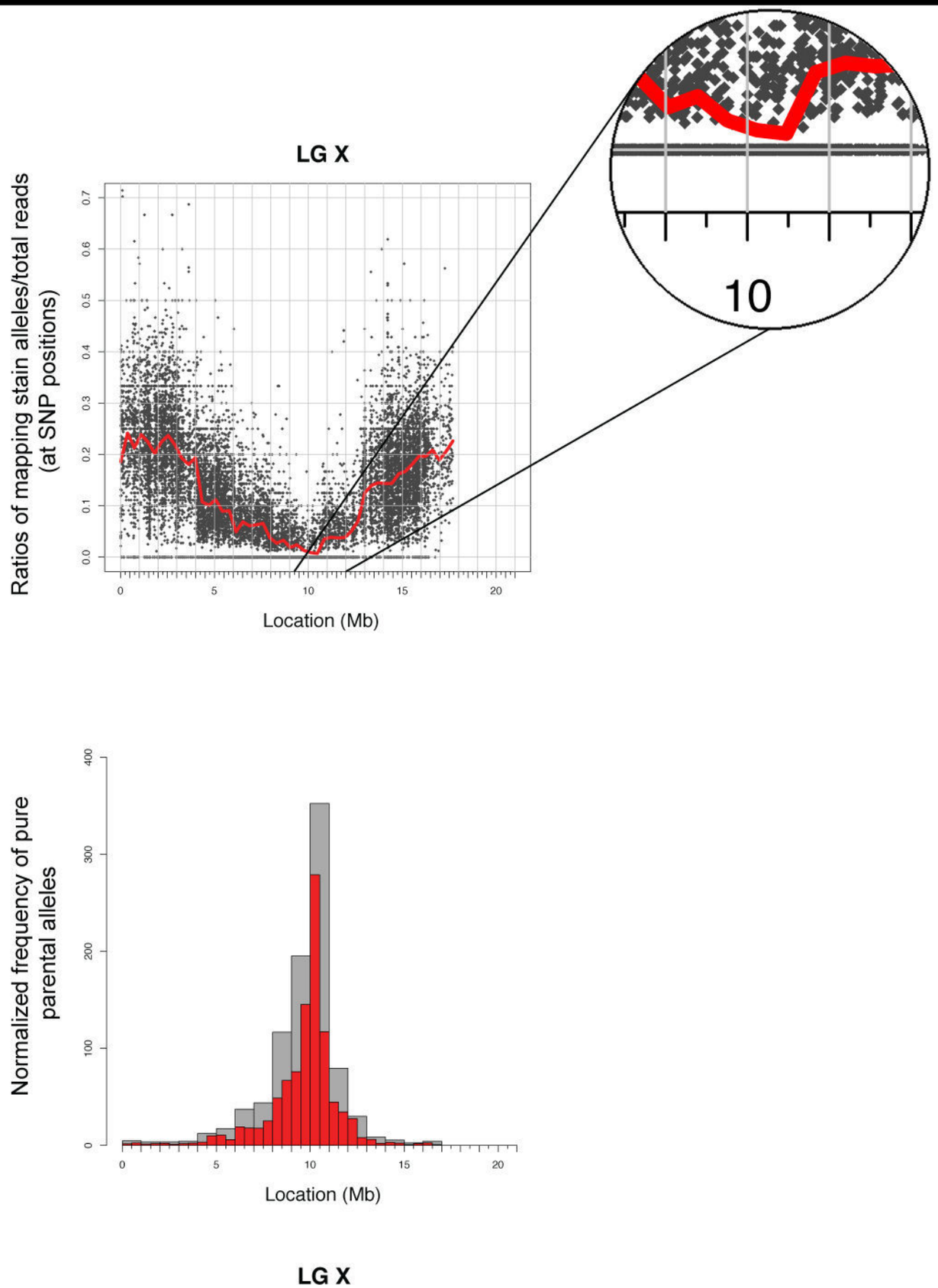


adapted from Doitsidou, 2008

SNP mapping concept



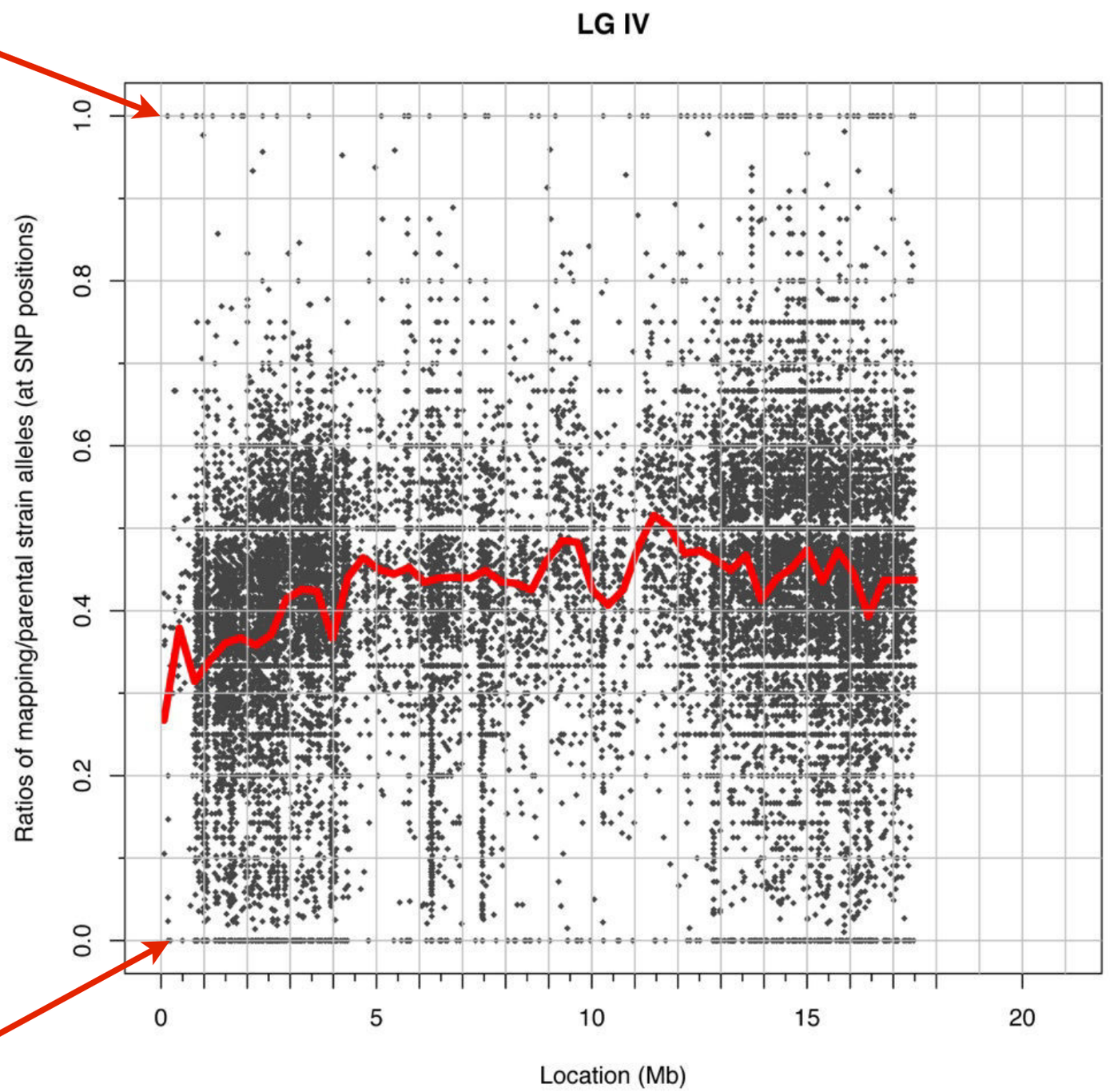
CloudMap's SNP mapping plots



vab-3:LG X 10,503,393-
10,519,348

An unlinked chromosome

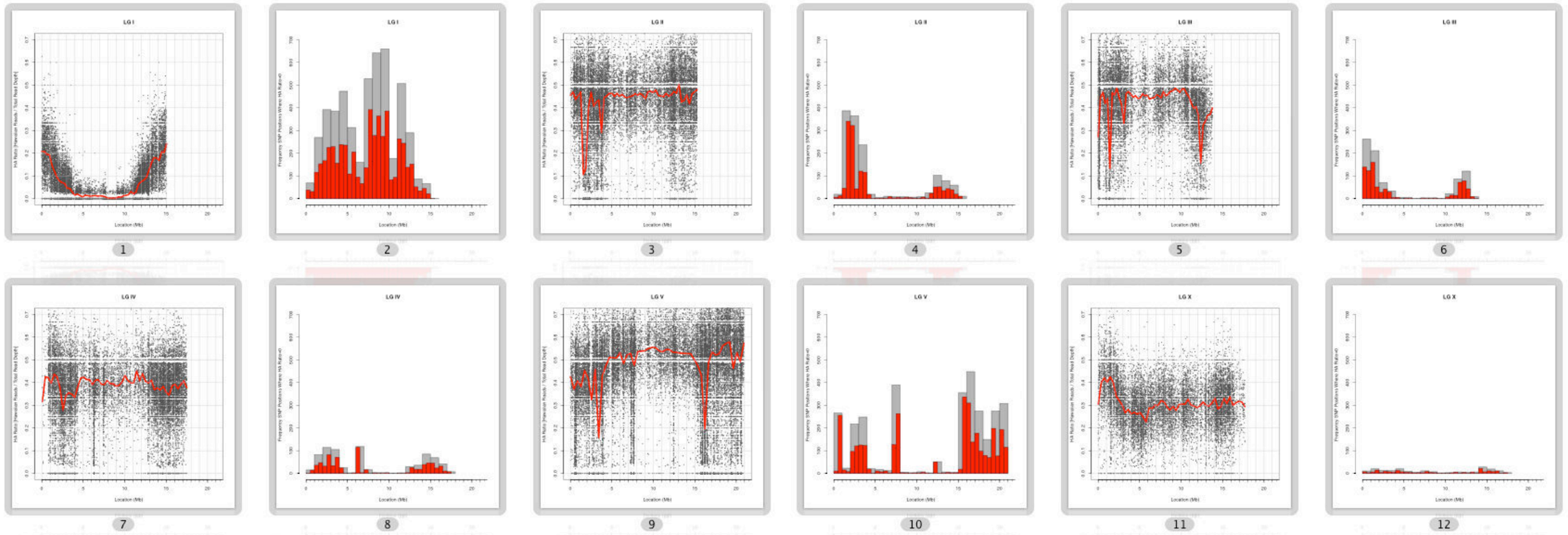
pure Hawaiian ratio



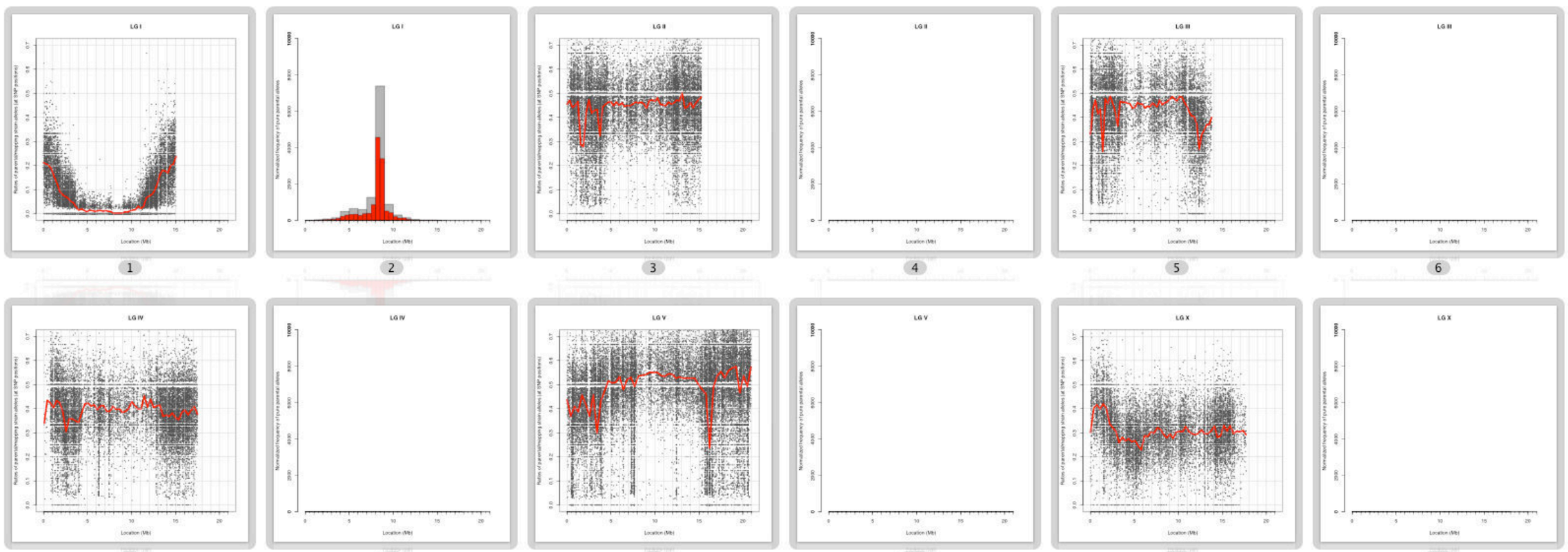
pure parental ratio

SNP mapping plots before/after normalization

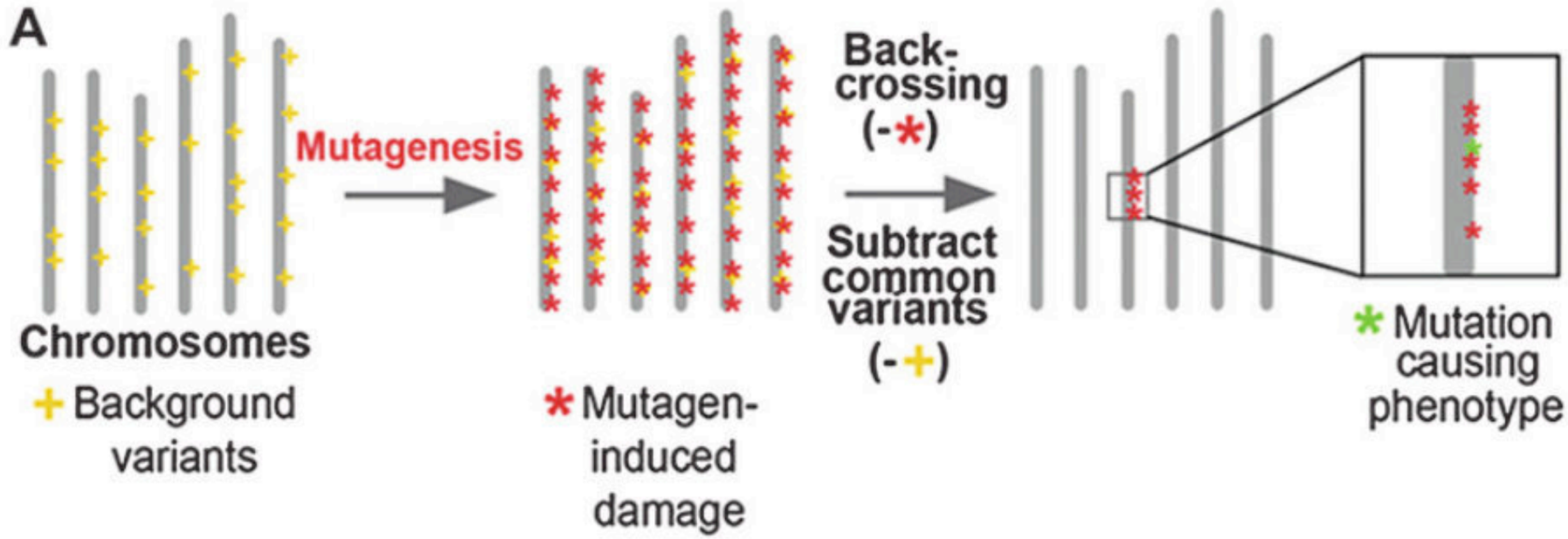
Pre-Normalization:

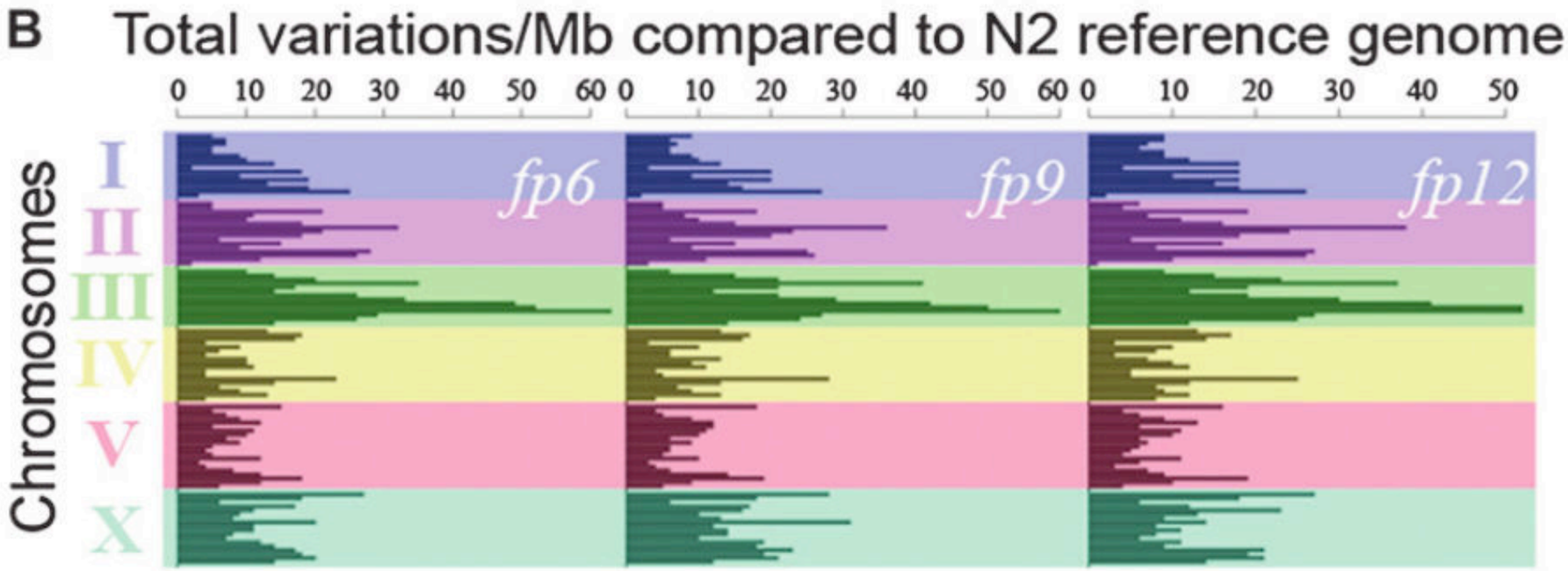


Post-Normalization:



EMS Density Mapping





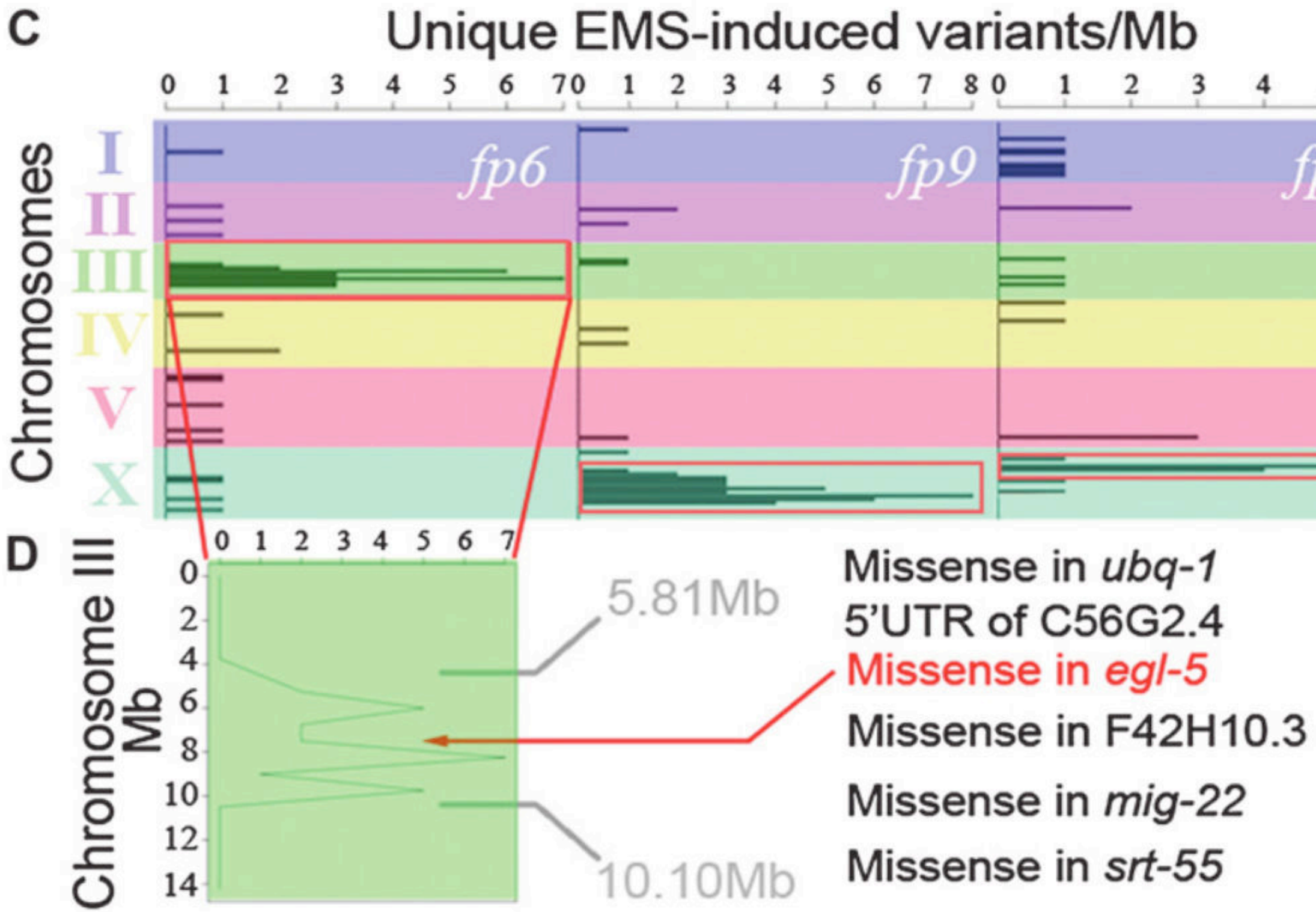
1) Subtraction of common variants

2) Quality filtering

3) Filtering for EMS nucleotide changes

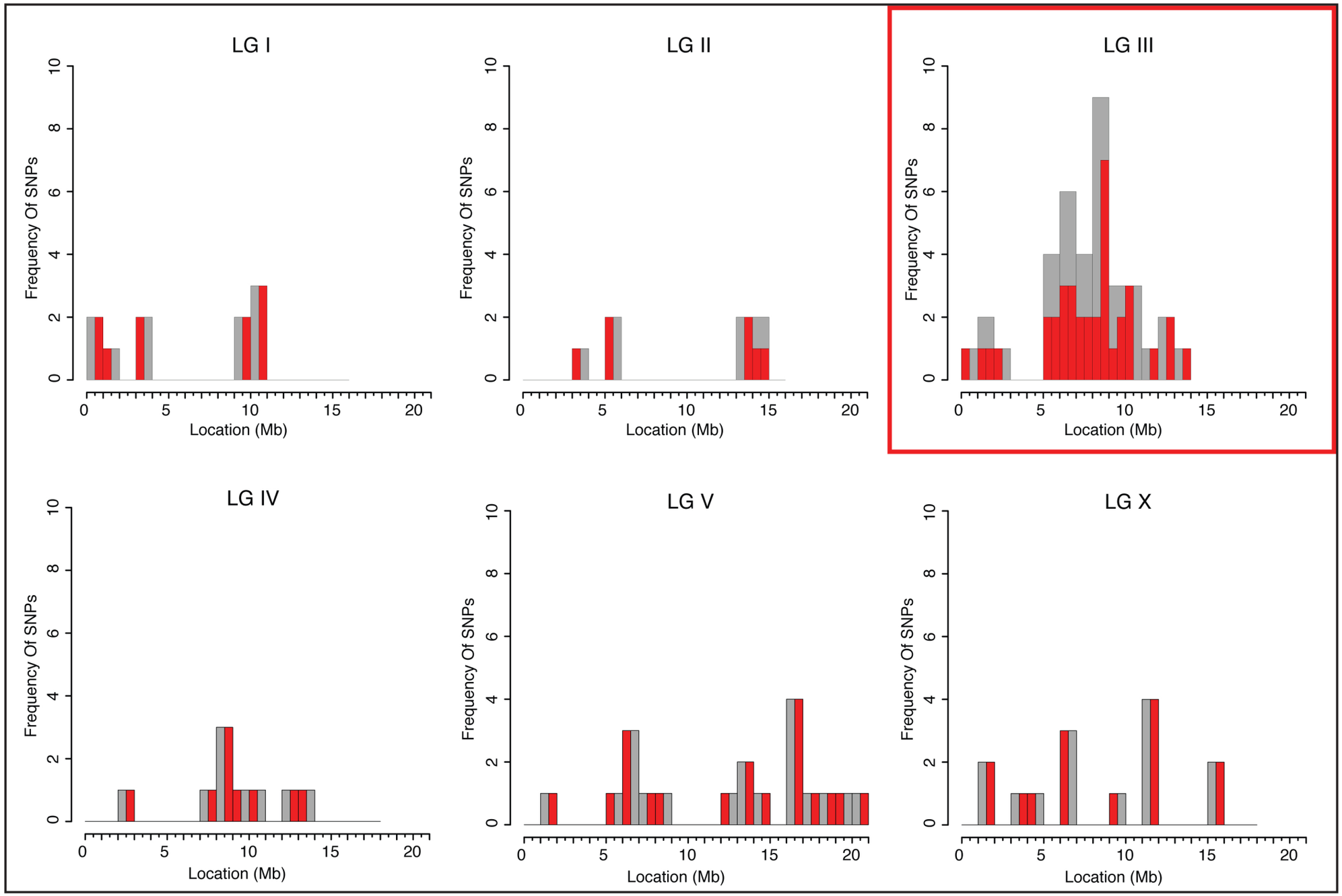
Zuryn et al., 2010

EMS Density Mapping

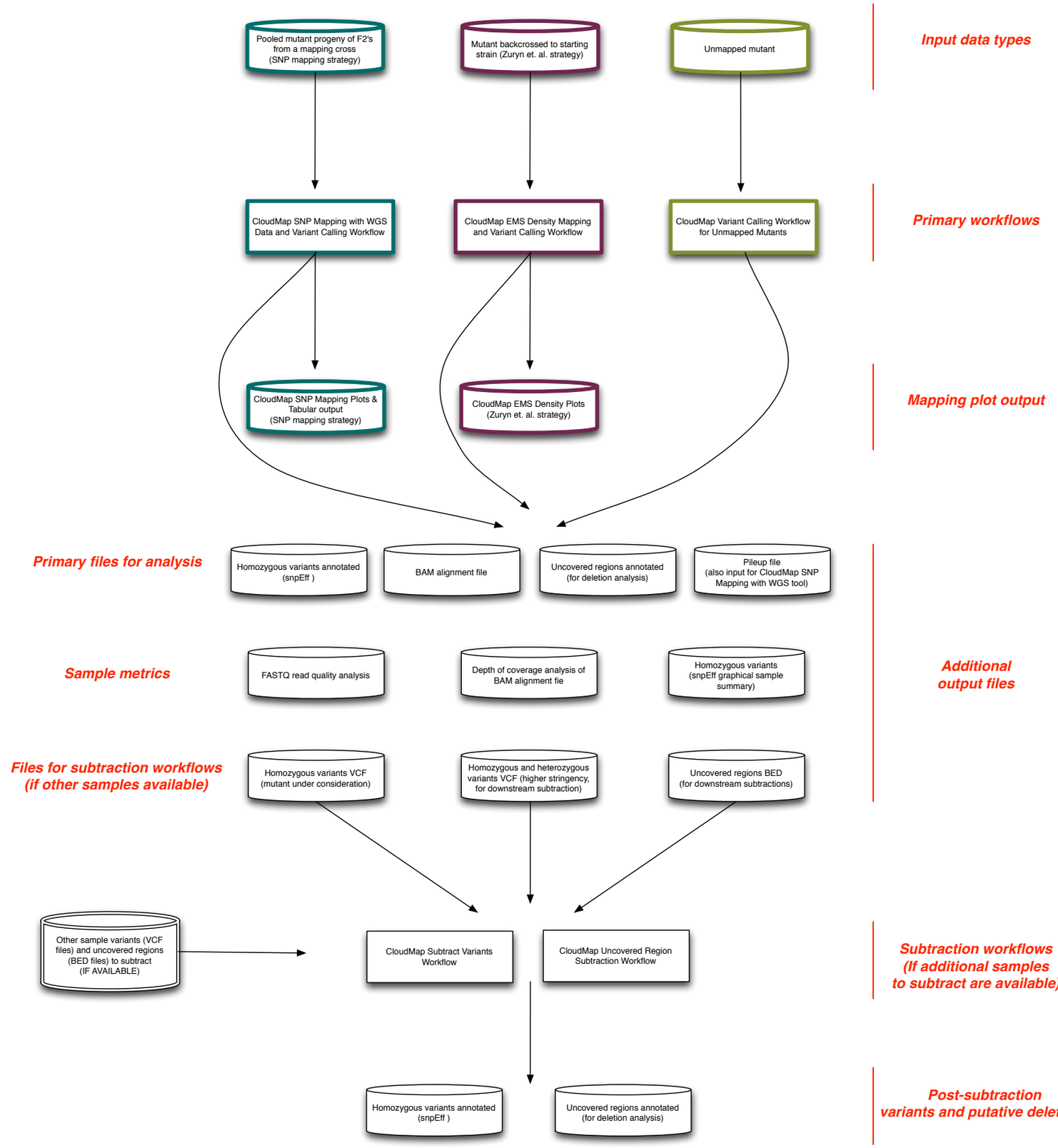


Zuryn et al., 2010

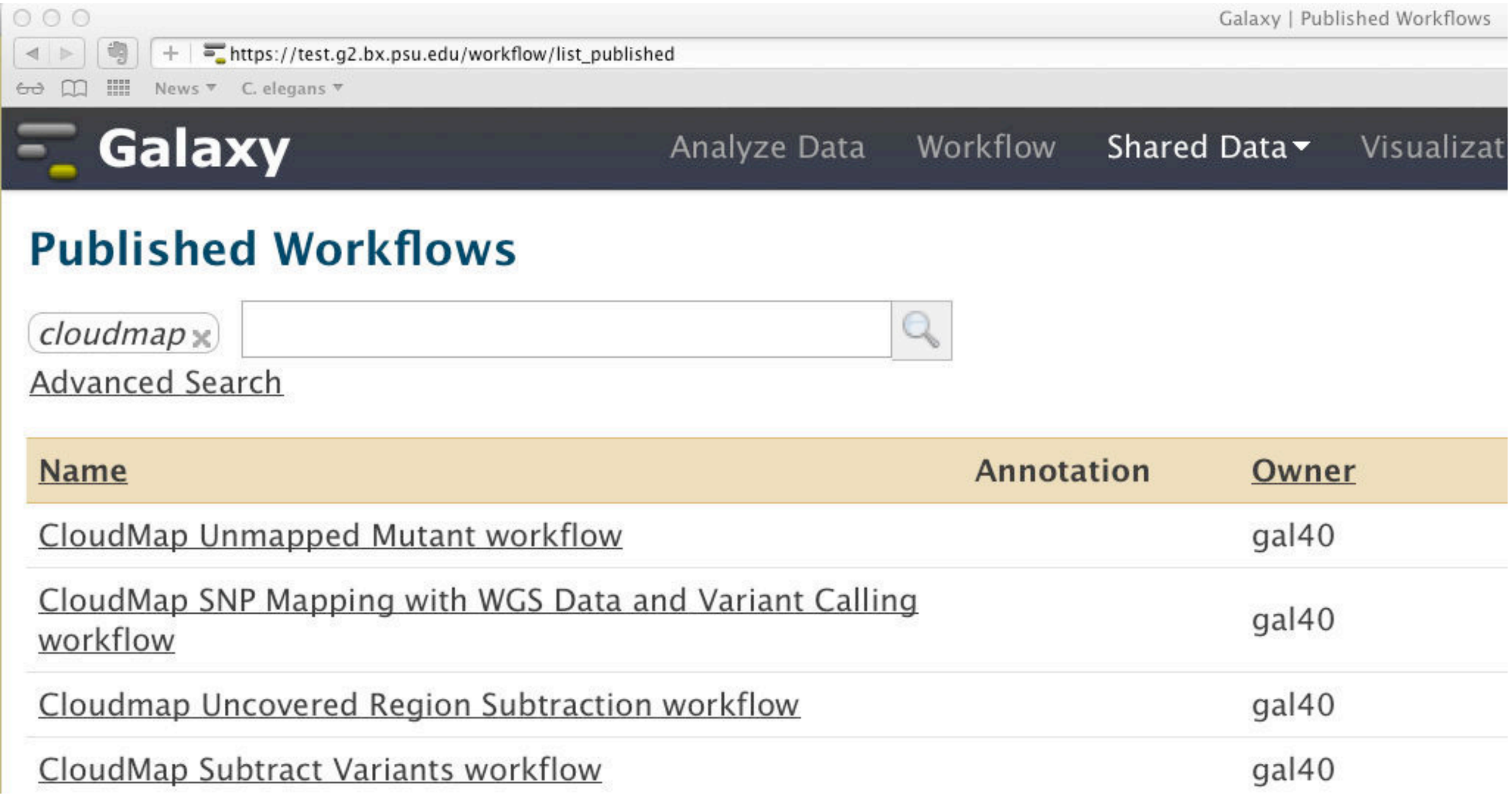
CloudMap: EMS Variants Density Mapping Tool



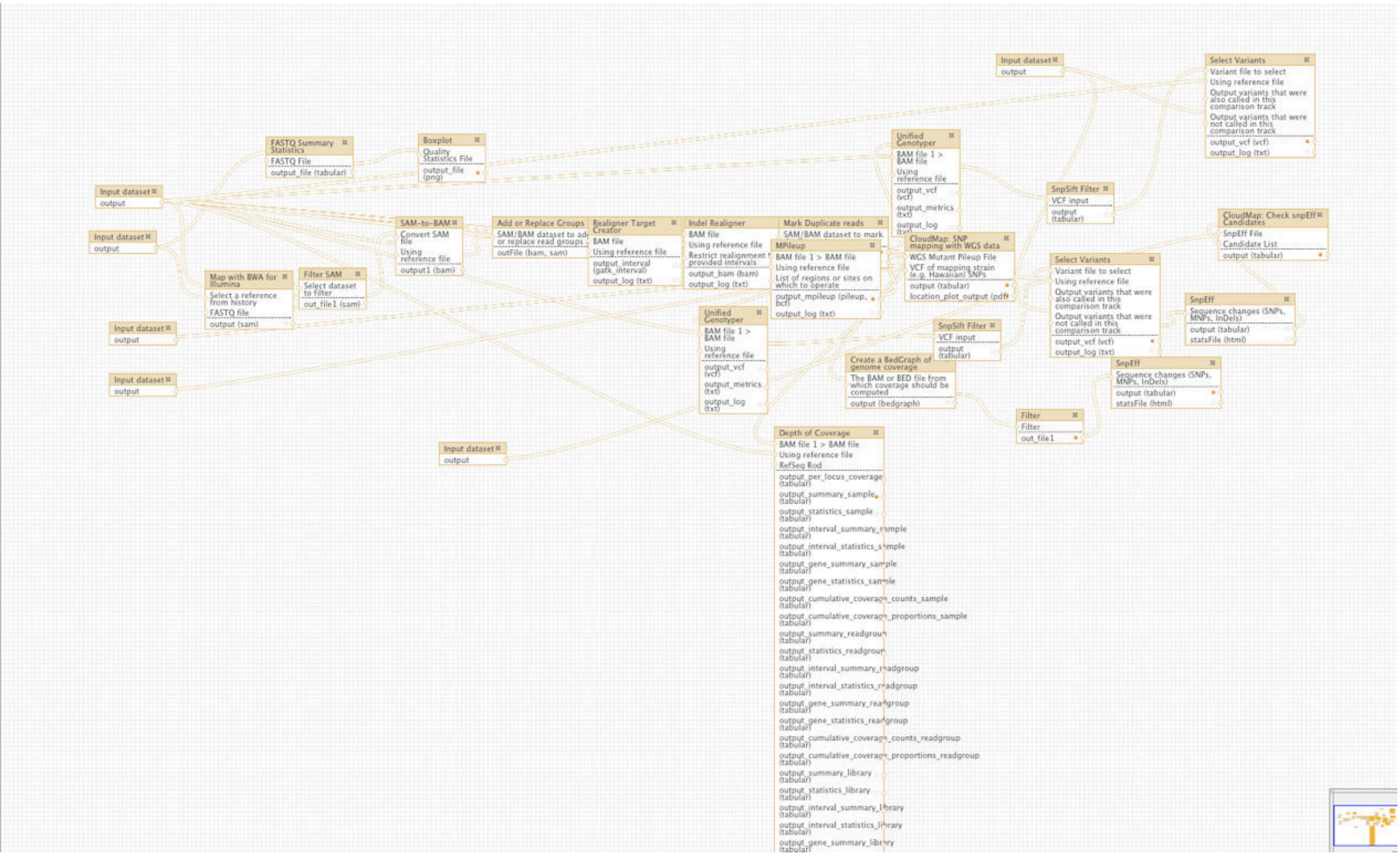
CloudMap provides a set of workflows for analysis of mutant genome sequences



CloudMap provides a set of workflows for analysis of mutant genome sequences



CloudMap provides a set of workflows for analysis of mutant genome sequences



CloudMap data libraries contain a proof of principle dataset and config files

Galaxy | Analyze Data | Workflow | **Shared Data** | Visualization | Cloud | Help | User

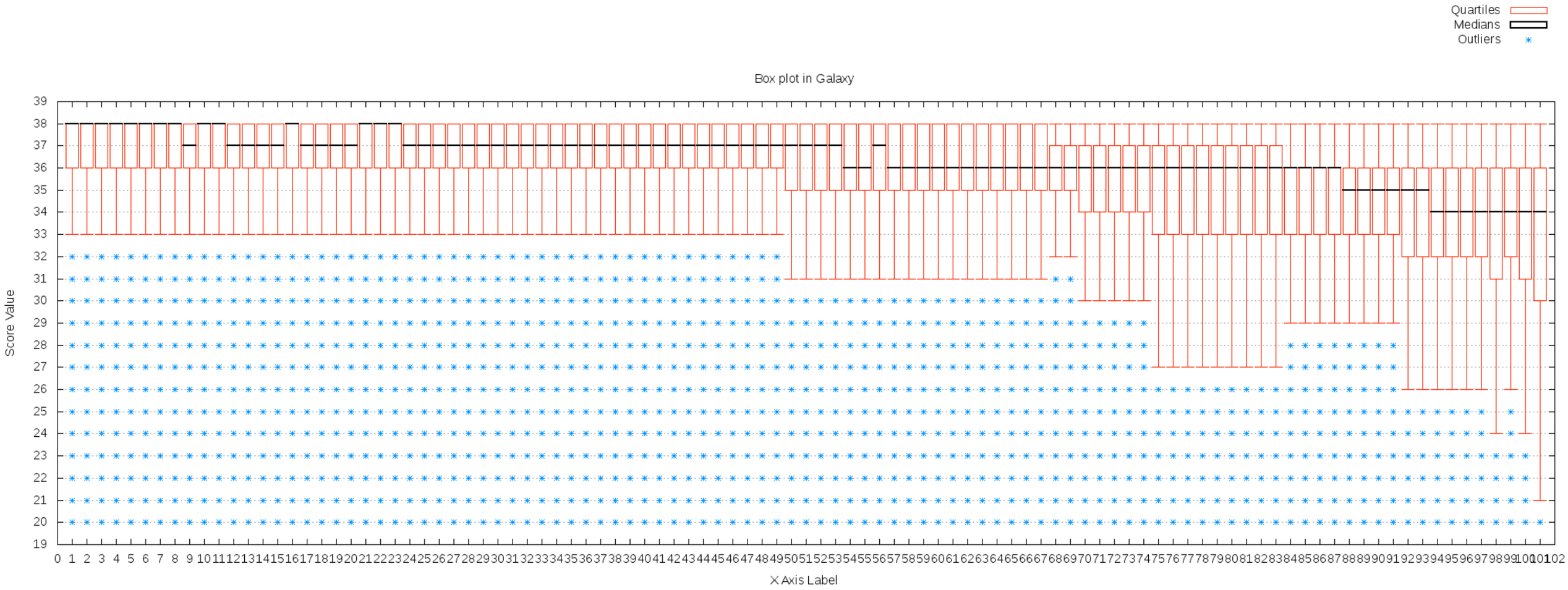
Data Library "CloudMap"

Contains reference and configuration files for the Cloudmap pipeline

Name	Message
Candidate gene lists	Check snpEff output against these candidate genes using CloudMap Check snpEff Candidates tool
CloudMap_ChromatinFactors.txt	
CloudMap_TranscriptionFactors_wTF2.2.txt	List of 937 C. elegans transcription factors from Reece-Hoyes et. al. Nature Methods 2011.
CloudMap user guides	Detailed guides for using the CloudMap pipeline
EMS Variant Density Mapping	Use this dataset to try out the CloudMap EMS Variant Density Mapping tool
ot266 proof of principle dataset	Use these files to run the CloudMap ot266 proof of principle example
Hawaiian SNP reference files unfiltered (WS220.64)	
HA_SNPS_WS220.64_chr.bed	
HA_SNPS_WS220.64_chr.vcf	
ot260 and ot263 VCFs for variant subtraction	Use these VCFs for the CloudMap ot266 proof of principle example
ot266_ProofOfPrinciple_Small.fastqsanger	Sample FASTQ file for ot266 Proof of principle
HA_SNPs_WS220_Filtered_103626_SNPs_chr.bed	Filtered set of Hawaiian SNP positions (used by mpileup tool)
HA_SNPs_WS220_Filtered_103626_SNPs_chr.vcf	Filtered set of Hawaiian SNP variants (used by CloudMap SNP Mapping with WGS tool)
WS220.64_chr.fa	WS220.64 genomic reference file
SNP Mapping with WGS Data Other Species Config Files	Use these config files if you want to use the SNP Mapping with WGS Data for any species other than C.elegans and Arabadopsis

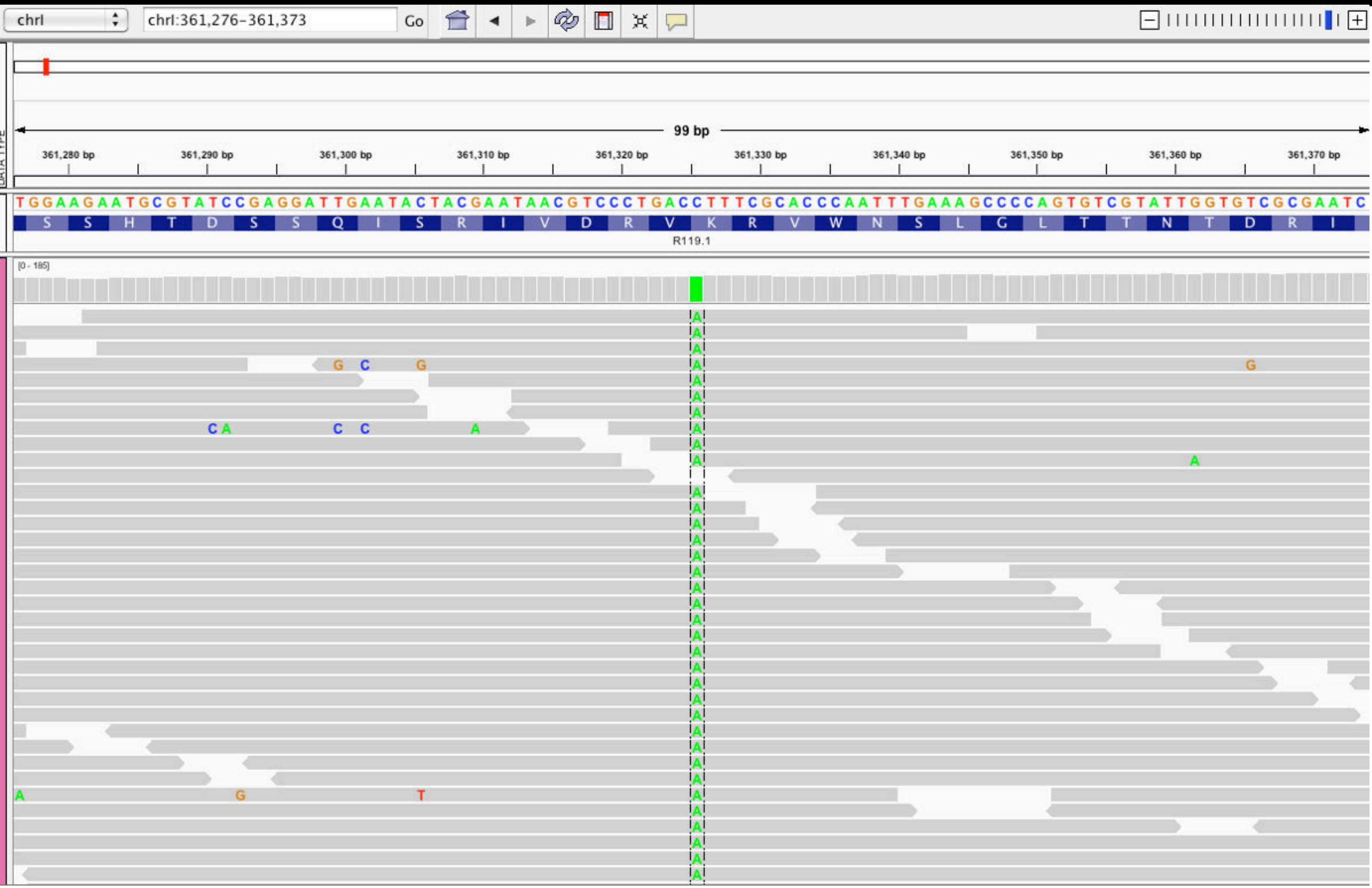
For selected datasets: Import to current history Go

FASTQ statistics (FASTQ Summary Statistics tool)



Blankenberg et al., 2010

Alignments (BWA, GATK realigner, PICARD remove dups)



Variant calling (GATK or SAMtools) & variant filtering (snpSift)

	A	B	C	D	E	F	G	H	I	J
1	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	rgSM
2	chrI	962	.	G	T	367.18	.	AC=1;AF=0.50;AN=2;BaseQRankSum=0.403;DP=23	GT:AD:DP:GQ:PL	0/1:10,13:23:99:397,0,325
3	chrI	991	.	GA	G	100.41	.	AC=1;AF=0.50;AN=2;BaseQRankSum=2.130;DP=14	GT:AD:DP:GQ:PL	0/1:8,6:14:99:139,0,246
4	chrI	1216	.	A	T	68.96	.	AC=1;AF=0.50;AN=2;BaseQRankSum=1.300;DP=7;	GT:AD:DP:GQ:PL	0/1:4,3:7:98.95:99,0,138
5	chrI	1222	.	A	C	109.76	.	AC=1;AF=0.50;AN=2;BaseQRankSum=1.754;DP=7;	GT:AD:DP:GQ:PL	0/1:3,4:7:57.20:140,0,57
6	chrI	1290	.	T	A	126.47	.	AC=1;AF=0.50;AN=2;BaseQRankSum=0.933;DP=14	GT:AD:DP:GQ:PL	0/1:9,5:14:99:156,0,306
7	chrI	1412	.	T	C	235.12	.	AC=1;AF=0.50;AN=2;BaseQRankSum=-1.203;DP=1	GT:AD:DP:GQ:PL	0/1:8,9:17:99:265,0,266
8	chrI	1414	.	G	A	205.1	.	AC=1;AF=0.50;AN=2;BaseQRankSum=-0.209;DP=1	GT:AD:DP:GQ:PL	0/1:7,8:15:99:235,0,233
9	chrI	1421	.	G	A	196.85	.	AC=1;AF=0.50;AN=2;BaseQRankSum=-1.096;DP=1	GT:AD:DP:GQ:PL	0/1:7,8:15:99:227,0,228

Variant annotation (snpEff)

# Chromo	Position	Reference	Change	Change_t	Homozyg	Quality	Coverage	Gene_ID	Gene_nar	Bio_type	Trancript_	Exon_ID	Exon_Ran	Effect	old_AA/n	Old_codo	Codon_N	Codon_D	CDS_size
I	9879698	C	T	SNP	Het	27	62	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.1			INTRON					1122
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122
I	9909039	G	A	SNP	Het	143	95	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620
I	9909895	G	T	SNP	Het	162	68	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620
I	9910274	*	#NAME?	INS	Het	4.42	102	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620
I	9910270	*	#NAME?	INS	Het	4.42	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620
I	9910274	*	#NAME?	INS	Het	217	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620
I	9994773	A	G	SNP	Het	142	73	T23D8.8	cfi-1	protein_codi	T23D8.8	exon_I_9994	7	NON_SYNON	I/T	aTc/aCc	451	0	1404
I	10139575	C	G	SNP	Het	16.1	28							INTERGENIC					
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.2	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.1	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11b			INTRON					1356
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11a			INTRON					1362
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11b	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1356
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11a	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1362
I	10207646	A	C	SNP	Het	49	480							INTERGENIC					
I	10209783	A	G	SNP	Hom	27	45							INTERGENIC					
I	10209806	G	A	SNP	Het	3.55	16							INTERGENIC					
I	10248569	G	T	SNP	Het	141	100	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218
I	10248578	C	T	SNP	Het	39	92	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218
I	10251364	T	C	SNP	Het	113	115	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218
I	10251848	C	T	SNP	Het	135	102	ZC247.3	lin-11	protein_codi	ZC247.3	exon_I_1025	5	SYNONYMOU	S/S	tcC/tcT	202	3	1218
I	10383040	G	A	SNP	Het	40	99	F45H11.4	mgl-2	protein_codi	F45H11.4.2			UTR_3_PRIME: 590 bases from CDS					
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3b			INTRON					2406
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3a			INTRON					2454
I	10477650	C	T	SNP	Het	120	83	F37D6.2	F37D6.2	protein_codi	F37D6.2a.1			INTRON					1749

List of variants + Annotation information —> List of variant effects

CloudMap variant annotation candidate checker

# Chromo	Position	Reference	Change	Change_t	Homozyg	Quality	Coverage	Gene_ID	Gene_nar	Bio_type	Trancript	Exon_ID	Exon_Ran	Effect	old_AA/n	Old_codo	Codon_N	Codon_D	CDS_size	TFs
I	9879698	C	T	SNP	Het	27	62	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122	ZF - NHR
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.1			INTRON					1122	ZF - NHR
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122	ZF - NHR
I	9909039	G	A	SNP	Het	143	95	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers
I	9909895	G	T	SNP	Het	162	68	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers
I	9910274	*	#NAME?	INS	Het	4.42	102	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers
I	9910270	*	#NAME?	INS	Het	4.42	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers
I	9910274	*	#NAME?	INS	Het	217	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers
I	9994773	A	G	SNP	Het	142	73	T23D8.8	cfi-1	protein_codi	T23D8.8	exon_I_9994	7	NON_SYNON	I/T	aTc/aCc	451	0	1404	ARID/BRIGHT
I	10139575	C	G	SNP	Het	16.1	28							INTERGENIC						WH - Fork Head, AT Hook
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.2	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585	WH - Fork Head
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.1	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585	WH - Fork Head
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11b			INTRON					1356	bHLH
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11a			INTRON					1362	bHLH
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11b	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1356	bHLH
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11a	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1362	bHLH
I	10207646	A	C	SNP	Het	49	480							INTERGENIC						WH - Fork Head, AT Hook
I	10209783	A	G	SNP	Hom	27	45							INTERGENIC						WH - Fork Head, AT Hook
I	10209806	G	A	SNP	Het	3.55	16							INTERGENIC						WH - Fork Head, AT Hook
I	10248569	G	T	SNP	Het	141	100	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM
I	10248578	C	T	SNP	Het	39	92	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM
I	10251364	T	C	SNP	Het	113	115	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM
I	10251848	C	T	SNP	Het	135	102	ZC247.3	lin-11	protein_codi	ZC247.3	exon_I_1025	5	SYNONYMOU	S/S	tcC/tcT	202	3	1218	HD - LIM
I	10383040	G	A	SNP	Het	40	99	F45H11.4	mgl-2	protein_codi	F45H11.4.2			UTR_3_PRIME: 590 bases from CDS						bZIP
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3b			INTRON					2406	ZF - C2H2 - 4 fingers
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3a			INTRON					2454	ZF - C2H2 - 4 fingers
I	10477650	C	T	SNP	Het	120	83	F37D6.2	F37D6.2	protein_codi	F37D6.2a.1			INTRON					1749	ZF - C2H2 - 5 fingers

1) Transcription factors

2) Transgene silencers

3) Genes expressed in the nervous system

4) Anything you want. . .

Variant annotation (snpEff)

Number of effects by impact

Type (alphabetical order)	Count	Percent
HIGH	245	0.774%
LOW	16,895	53.371%
MODERATE	501	1.583%
MODIFIER	14,015	44.273%

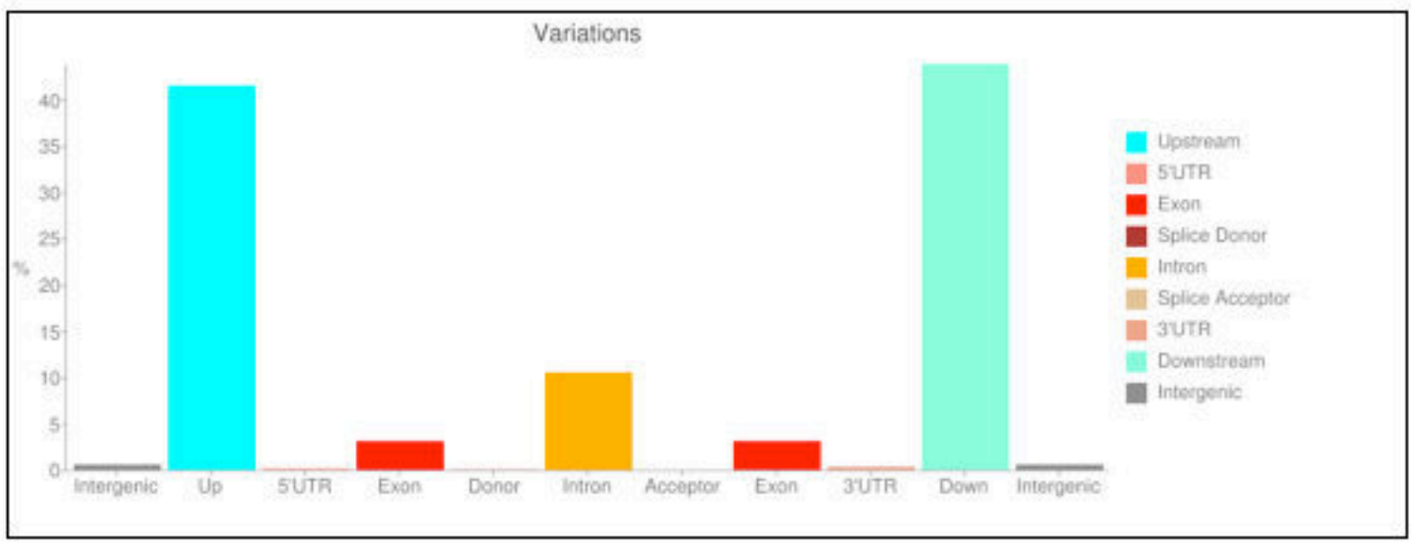
Number of effects by type and region

Type

Type (alphabetical order)	Count	Percent
CODON_CHANGE_PLUS_CODON_DELETION	1	0.003%
CODON_CHANGE_PLUS_CODON_INSERTION	1	0.003%
CODON_DELETION	2	0.006%
CODON_INSERTION	4	0.013%
DOWNSTREAM	13,868	43.808%
FRAME_SHIFT	211	0.667%
INTERGENIC	177	0.559%
INTRON	3,332	10.526%
NON_SYNONYMOUS_CODING	493	1.557%
SPLICE_SITE_ACCEPTOR	6	0.019%
SPLICE_SITE_DONOR	9	0.028%
STOP_GAINED	15	0.047%
STOP_LOST	4	0.013%
SYNONYMOUS_CODING	250	0.79%
SYNONYMOUS_START	1	0.003%
SYNONYMOUS_STOP	1	0.003%
UPSTREAM	13,134	41.49%
UTR_3_PRIME	104	0.329%
UTR_5_PRIME	43	0.136%

Region

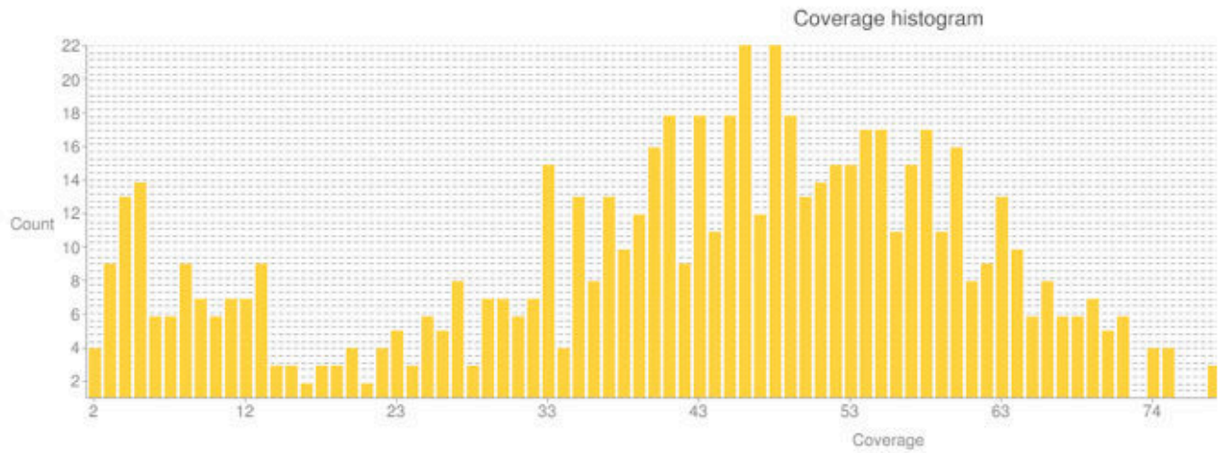
Type (alphabetical order)	Count	Percent
DOWNSTREAM	13,868	43.808%
EXON	983	3.105%
INTERGENIC	177	0.559%
INTRON	3,332	10.526%
SPLICE_SITE_ACCEPTOR	6	0.019%
SPLICE_SITE_DONOR	9	0.028%
UPSTREAM	13,134	41.49%
UTR_3_PRIME	104	0.329%
UTR_5_PRIME	43	0.136%



Variant annotation (snpEff)

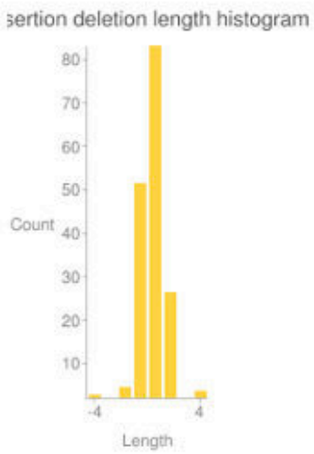
Coverage:

Min	2
Max	1,370
Mean	54.308
Median	46
Standard deviation	91.976
Values	2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 5
Count	4 9 13 14 6 6 9 7 6 7 7 9 3 3 2 3 3 4 2 4 5 3 6 5 8 3 7 7 6 7 15 4 13 8 13 10 12 16 18 9 18 11 18 22 12 22 18 13 14 15 15 17 17 11 15 17 11 16 8 9 1



Insertions and deletions length:

Min	-4
Max	4
Mean	0.444
Median	1
Standard deviation	1.457
Values	-4 -3 -2 -1 1 2 3 4
Count	3 2 5 52 83 27 2 4



Variant annotation (snpEff)

Base changes (SNPs)

	A	C	G	T
A	0	17	17	28
C	13	0	11	183
G	172	14	0	20
T	29	19	10	0

Ts/Tv (transitions / transversions)

Note: Only SNPs are used for this statistic.
Note: This Ts/Tv ratio is a 'raw' ratio. Some people prefer to use a ratio of rates, not observed events. In that case, you need to multiply by 2.0 (since there are twice as many possible transitions than transversions, E[Ts/Tv] ratio is twice the ratio of events).

Transitions	391
Transversions	142
Ts/Tv ratio	2.7535

All variants:

```
    Sample          : Total
Transitions    : 391      391
Transversions  : 142      142
Ts/Tv          : 2.754    2.754
```

Only known variants (i.e. the ones having a non-empty ID field):

```
No results available (empty input?)
```

Variant annotation (snpEff)

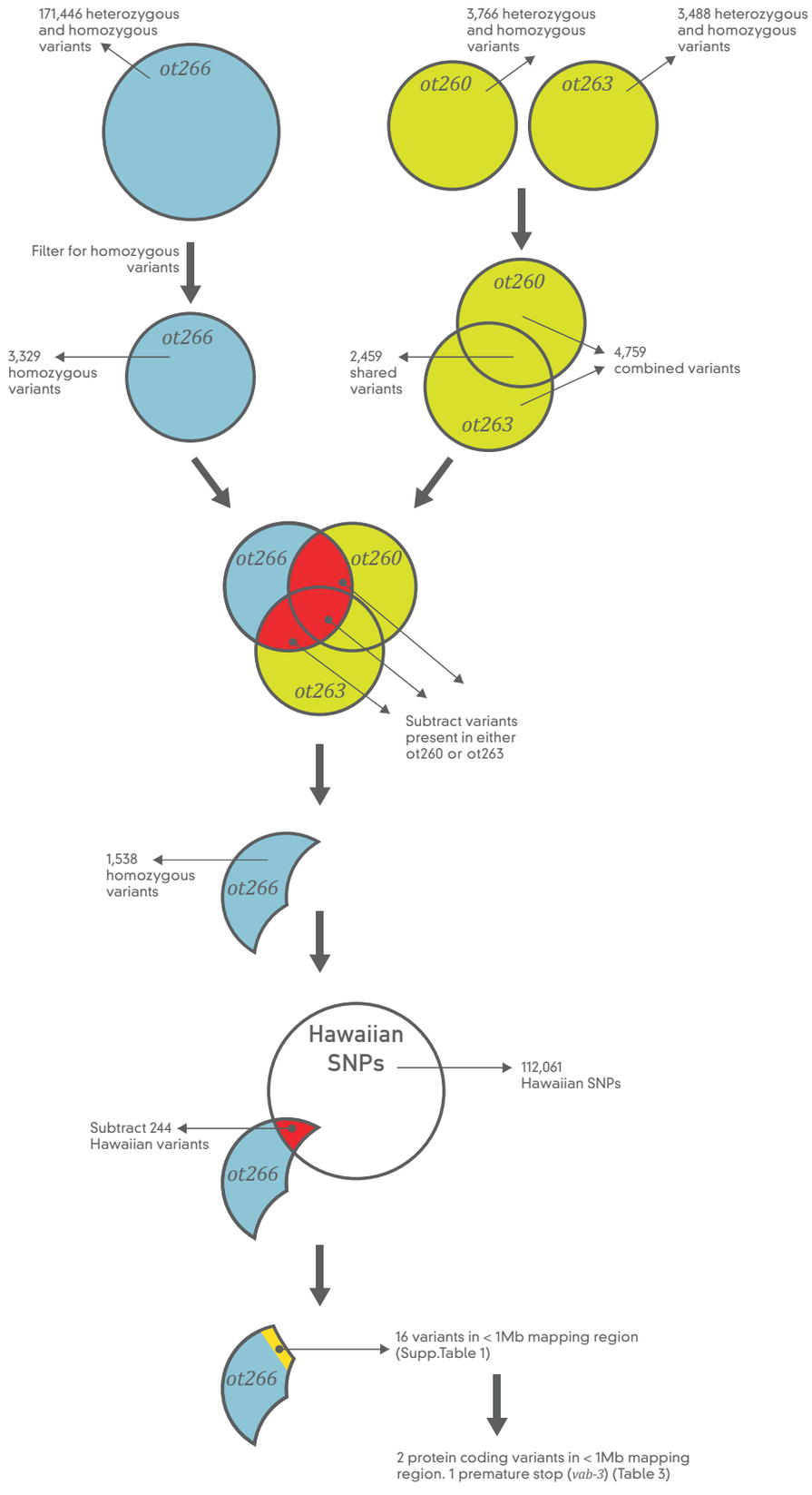
Amino acid changes

How to read this table:
- Rows are reference amino acids and columns are changed amino acids. E.g. Row 'A' column 'E' indicates how many 'A' amino acids have been replaced by 'E' amino acids.
- Red background colors indicate that more changes happened (heat-map).
- Diagonals are indicated using grey background color
- WARNING: This table may include different translation codon tables (e.g. mamalian DNA and mitochondrial DNA).

	*	-	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
*																						
-								1														
A			2												1			1	4	8		
C				2			1															2
D					1	1		2												3		
E		1				7					11											
F							3			1												
G					6	20		2									3			1		
H																						3
I										1												
K											5			1								
L							4					6										
M										6		1										
N														2								
P												9			2			13				
Q									1							12	2					
R				1						1	11						2					
S												2		1	1			2				
T										2					1				1			
V								1		1			2							2		
W	2																					
Y																						

Variant subtraction and filtration (GATK)

Fig.8



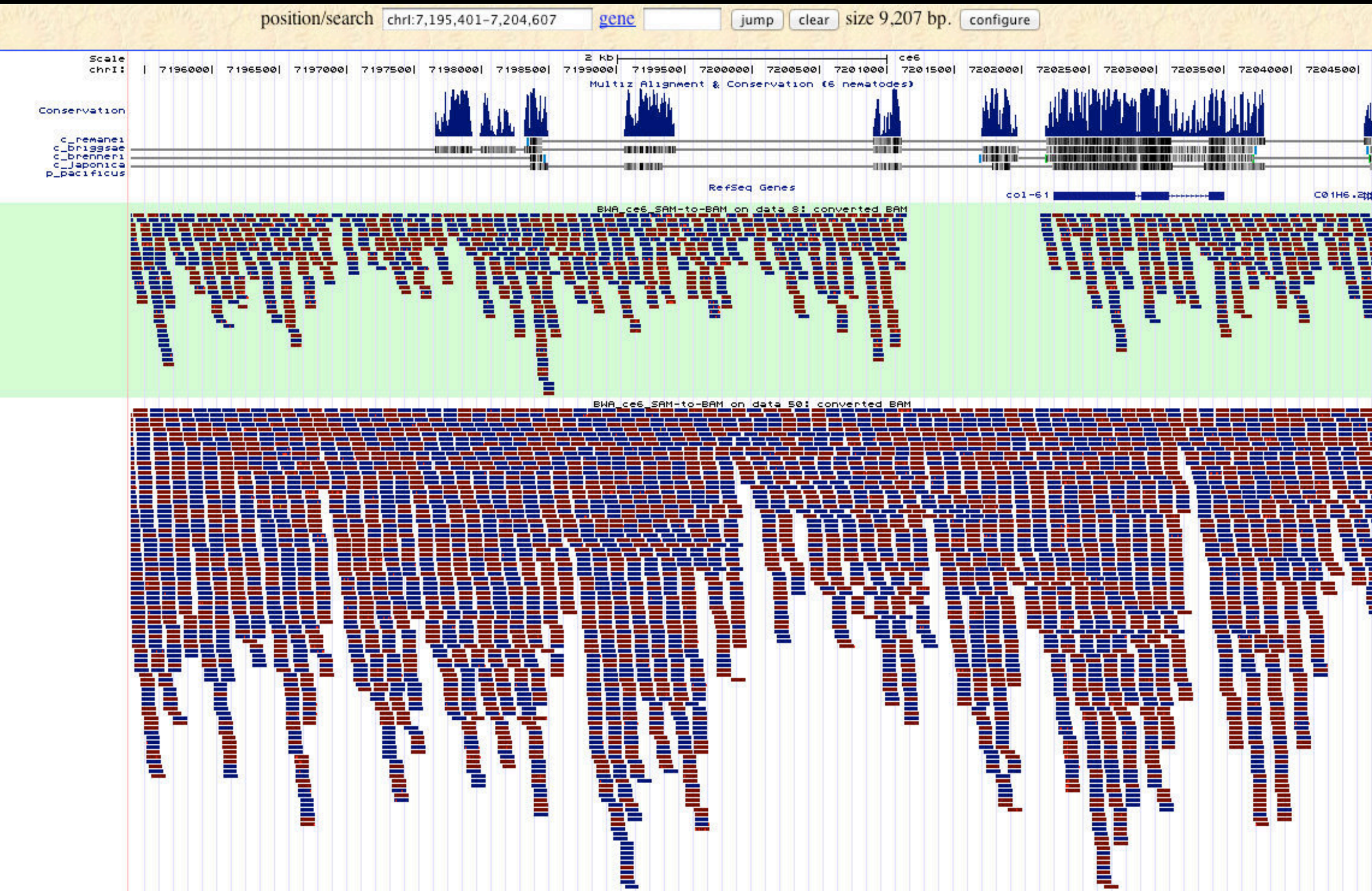
In silico complementation testing (compare two datasets tool)

A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T	U	Y	Z
# Chromo	Position	Reference	Change	Change_t	Homozyg	Quality	Coverage	Gene_ID	Gene_nar	Bio_type	Trancript_	Exon_ID	Exon_Ran	Effect	old_AA/n	Old_codo	Codon_N	Codon_D	CDS_size	TFs	
I	9879698	C	T	SNP	Het	27	62	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122	ZF - NHR	
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.1			INTRON					1122	ZF - NHR	
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122	ZF - NHR	
I	9909039	G	A	SNP	Het	143	95	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9909895	G	T	SNP	Het	162	68	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9910274	*	#NAME?	INS	Het	4.42	102	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9910270	*	#NAME?	INS	Het	4.42	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9910274	*	#NAME?	INS	Het	217	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9994773	A	G	SNP	Het	142	73	T23D8.8	cfi-1	protein_codi	T23D8.8	exon_I_9994	7	NON_SYNON	I/T	aTc/aCc	451	0	1404	ARID/BRIGHT	
I	10139575	C	G	SNP	Het	16.1	28							INTERGENIC						WH - Fork Head, AT Hook	
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.2	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585	WH - Fork Head	
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.1	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585	WH - Fork Head	
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11b			INTRON					1356	bHLH	
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11a			INTRON					1362	bHLH	
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11b	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1356	bHLH	
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11a	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1362	bHLH	
I	10207646	A	C	SNP	Het	49	480							INTERGENIC						WH - Fork Head, AT Hook	
I	10209783	A	G	SNP	Hom	27	45							INTERGENIC						WH - Fork Head, AT Hook	
I	10209806	G	A	SNP	Het	3.55	16							INTERGENIC						WH - Fork Head, AT Hook	
I	10248569	G	T	SNP	Het	141	100	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM	
I	10248578	C	T	SNP	Het	39	92	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM	
I	10251364	T	C	SNP	Het	113	115	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM	
I	10251848	C	T	SNP	Het	135	102	ZC247.3	lin-11	protein_codi	ZC247.3	exon_I_1025	5	SYNONYMOU	S/S	tcC/tcT	202	3	1218	HD - LIM	
I	10383040	G	A	SNP	Het	40	99	F45H11.4	mgl-2	protein_codi	F45H11.4.2			UTR_3_PRIME: 590 bases from CDS						bZIP	
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3b			INTRON					2406	ZF - C2H2 - 4 fingers	
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3a			INTRON					2454	ZF - C2H2 - 4 fingers	
I	10477650	C	T	SNP	Het	120	83	F37D6.2	F37D6.2	protein_codi	F37D6.2a.1			INTRON					1749	ZF - C2H2 - 5 fingers	

What genes have been hit in both members of a complementation group yet have different mutations?

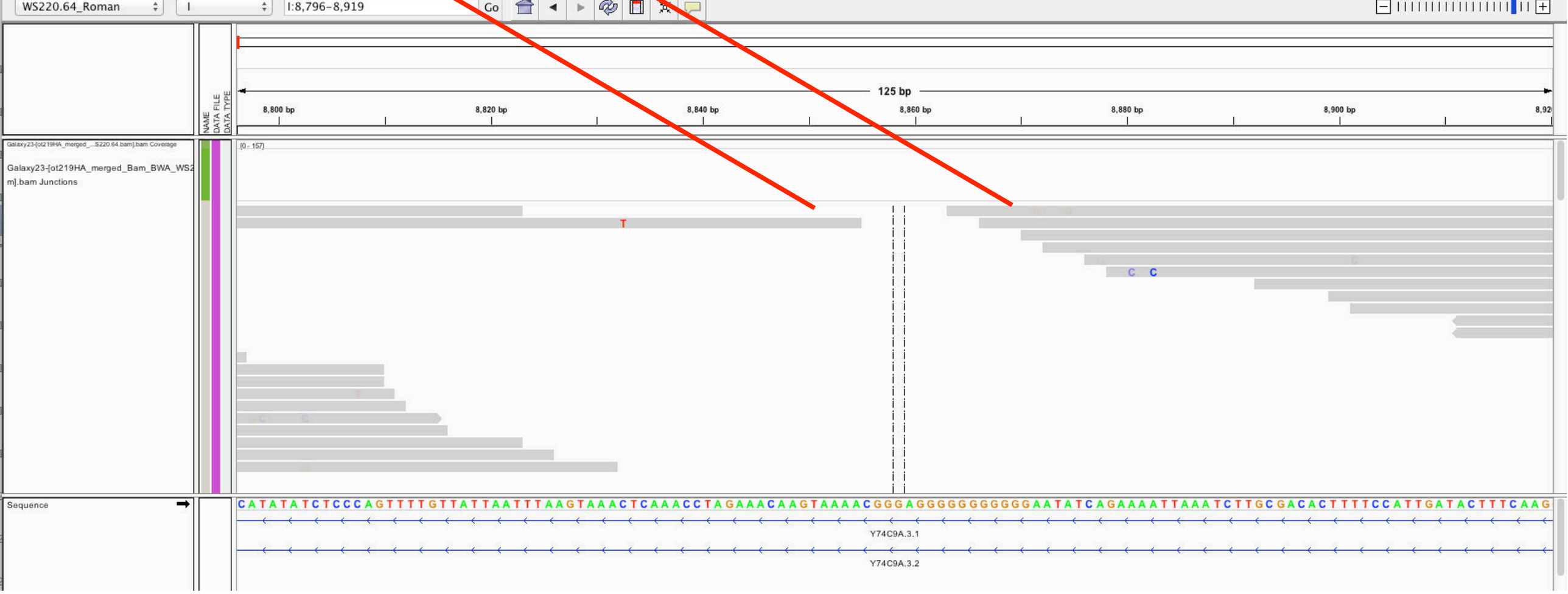
A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T	U	Y	Z
# Chromo	Position	Reference	Change	Change_t	Homozyg	Quality	Coverage	Gene_ID	Gene_nar	Bio_type	Trancript_	Exon_ID	Exon_Ran	Effect	old_AA/n	Old_codo	Codon_N	Codon_D	CDS_size	TFs	
I	9879698	C	T	SNP	Het	27	62	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122	ZF - NHR	
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.1			INTRON					1122	ZF - NHR	
I	9880489	C	T	SNP	Het	192	109	T23H4.2	nhr-69	protein_codi	T23H4.2.2			INTRON					1122	ZF - NHR	
I	9909039	G	A	SNP	Het	143	95	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9909895	G	T	SNP	Het	162	68	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9910274	*	#NAME?	INS	Het	4.42	102	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9910270	*	#NAME?	INS	Het	4.42	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9910274	*	#NAME?	INS	Het	217	103	F52F12.6	ztf-11	protein_codi	F52F12.6			INTRON					1620	ZF - C2HC 2 fingers	
I	9994773	A	G	SNP	Het	142	73	T23D8.8	cfi-1	protein_codi	T23D8.8	exon_I_9994	7	NON_SYNON	I/T	aTc/aCc	451	0	1404	ARID/BRIGHT	
I	10139575	C	G	SNP	Het	16.1	28							INTERGENIC						WH - Fork Head, AT Hook	
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.2	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585	WH - Fork Head	
I	10163982	A	G	SNP	Het	182	125	C25A1.2	fkh-10	protein_codi	C25A1.2.1	exon_I_1016	3	SYNONYMOU	H/H	caT/caC	112	1	585	WH - Fork Head	
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11b			INTRON					1356	bHLH	
I	10194812	G	A	SNP	Het	133	79	C25A1.11	aha-1	protein_codi	C25A1.11a			INTRON					1362	bHLH	
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11b	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1356	bHLH	
I	10195403	A	T	SNP	Het	222	147	C25A1.11	aha-1	protein_codi	C25A1.11a	exon_I_1019	5	SYNONYMOU	R/R	cgT/cgA	212	3	1362	bHLH	
I	10207646	A	C	SNP	Het	49	480							INTERGENIC						WH - Fork Head, AT Hook	
I	10209783	A	G	SNP	Hom	27	45							INTERGENIC						WH - Fork Head, AT Hook	
I	10209806	G	A	SNP	Het	3.55	16							INTERGENIC						WH - Fork Head, AT Hook	
I	10248569	G	T	SNP	Het	141	100	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM	
I	10248578	C	T	SNP	Het	39	92	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM	
I	10251364	T	C	SNP	Het	113	115	ZC247.3	lin-11	protein_codi	ZC247.3			INTRON					1218	HD - LIM	
I	10251848	C	T	SNP	Het	135	102	ZC247.3	lin-11	protein_codi	ZC247.3	exon_I_1025	5	SYNONYMOU	S/S	tcC/tcT	202	3	1218	HD - LIM	
I	10383040	G	A	SNP	Het	40	99	F45H11.4	mgl-2	protein_codi	F45H11.4.2			UTR_3_PRIME: 590 bases from CDS						bZIP	
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3b			INTRON					2406	ZF - C2H2 - 4 fingers	
I	10411503	A	G	SNP	Het	218	77	F25D7.3	blmp-1	protein_codi	F25D7.3a			INTRON					2454	ZF - C2H2 - 4 fingers	
I	10477650	C	T	SNP	Het	120	83	F37D6.2	F37D6.2	protein_codi	F37D6.2a.1			INTRON					1749	ZF - C2H2 - 5 fingers	

Which uncovered regions are deletions?

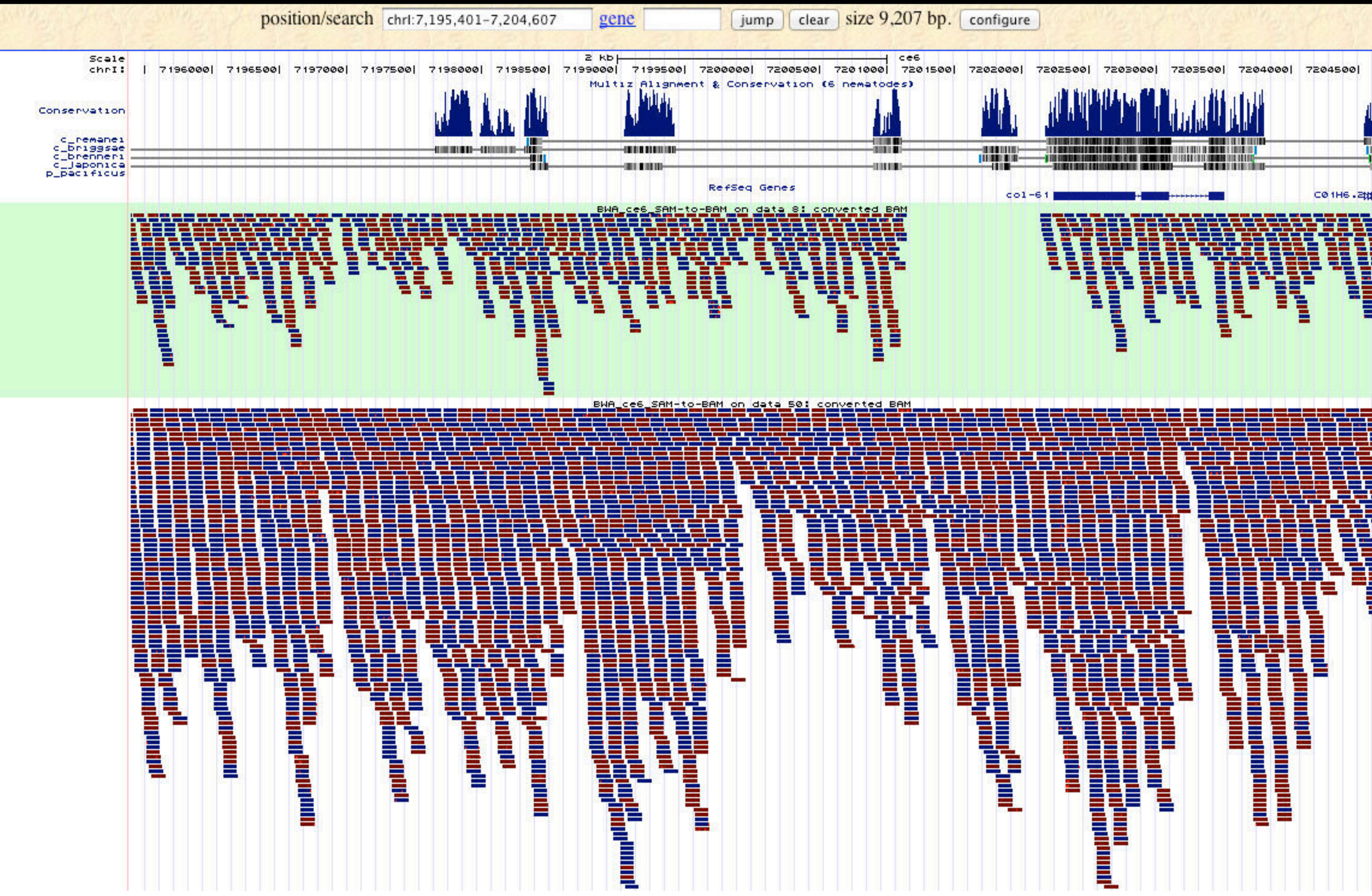


Identifying unique uncovered regions

	# Chromo	Position	Reference	Homozygous	Coverage	Warnings	Gene_ID	Gene_name	Bio_type	Trancript_ID	Exon_ID	Exon_Rank	Effect	old_AA/new	Old_codon/N	Codon_Num	Codon_Dege	CDS_size	Codons_arou	AAs_around	Custom_interval_ID
2	I	8855	8862	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.6			UPSTREAM: 2763 bases							0
3	I	8855	8862	Interval	0	0		Y74C9A.6	Y74C9A.6	snoRNA	Y74C9A.6			UPSTREAM: 4946 bases							0
4	I	8855	8862	Interval	0	0		Y74C9A.3	Y74C9A.3	protein_codi	Y74C9A.3.1			INTRON					705		0
5	I	8855	8862	Interval	0	0		Y74C9A.3	Y74C9A.3	protein_codi	Y74C9A.3.2			INTRON					705		0
6	I	8855	8862	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.3			UPSTREAM: 1558 bases							0
7	I	8855	8862	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.2			UPSTREAM: 2640 bases							0
8	I	8855	8862	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.1			UPSTREAM: 2644 bases							0
9	I	8855	8862	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.5			UPSTREAM: 2644 bases							0
10	I	8855	8862	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.4			UPSTREAM: 2650 bases							0
11	I	11248	11266	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.6			UPSTREAM: 370 bases							0
12	I	11248	11266	Interval	0	0		Y74C9A.3	Y74C9A.3	protein_codi	Y74C9A.3.2			UPSTREAM: 1018 bases							0
13	I	11248	11266	Interval	0	0		Y74C9A.3	Y74C9A.3	protein_codi	Y74C9A.3.1			UPSTREAM: 1016 bases							0
14	I	11248	11266	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.3			INTRON					372		0
15	I	11248	11266	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.2			UPSTREAM: 247 bases							0
16	I	11248	11266	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.1			UPSTREAM: 251 bases							0
17	I	11248	11266	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.5			UPSTREAM: 251 bases							0
18	I	11248	11266	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.4			UPSTREAM: 257 bases							0
19	I	15671	15712	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.2			INTRON					372		0
20	I	15671	15712	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.1			INTRON					372		0
21	I	15671	15712	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.4			INTRON					372		0
22	I	15671	15712	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.3			INTRON					372		0
23	I	15671	15712	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.6			INTRON					372		0
24	I	15671	15712	Interval	0	0		Y74C9A.2	nlp-40	protein_codi	Y74C9A.2.5			INTRON					372		0

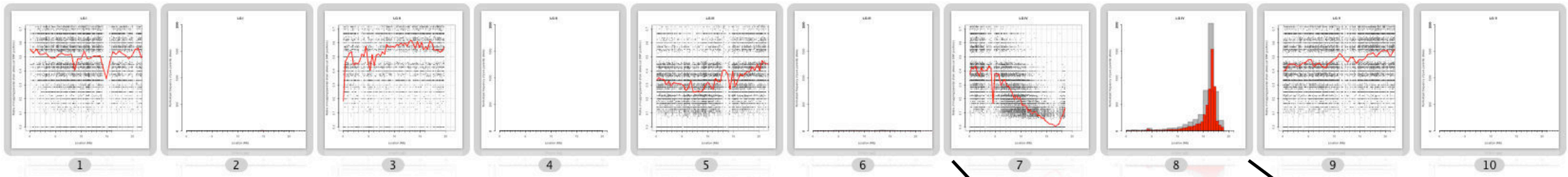


Putative deletion analysis

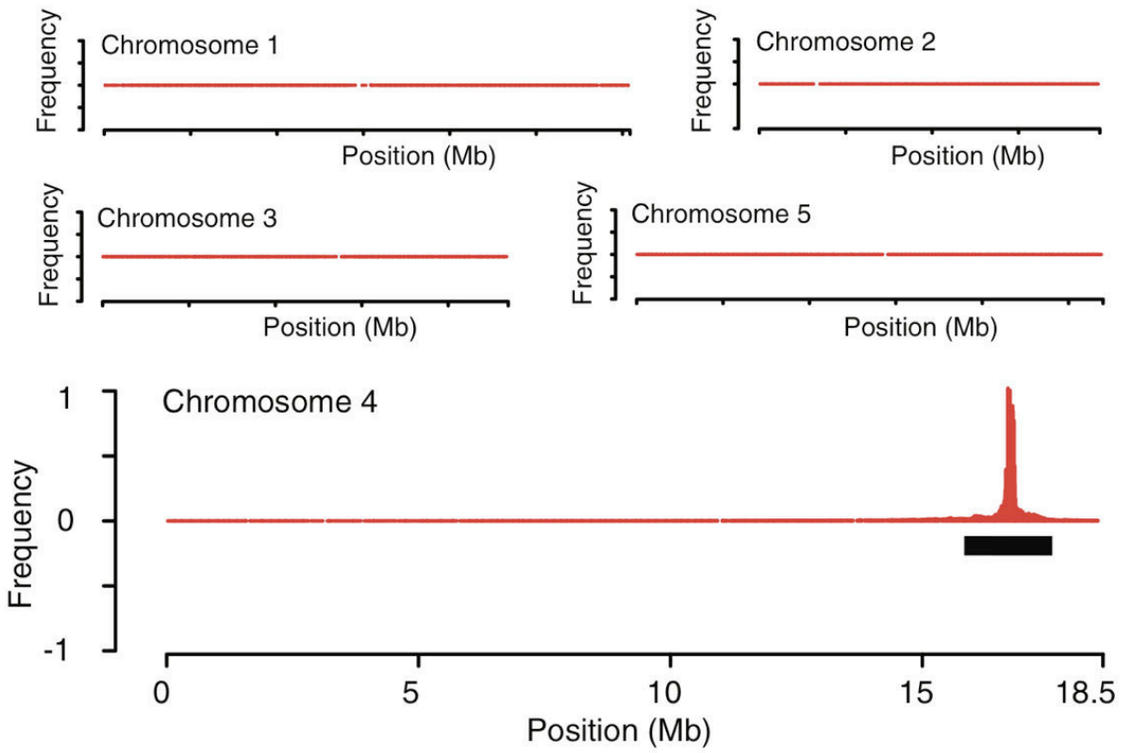


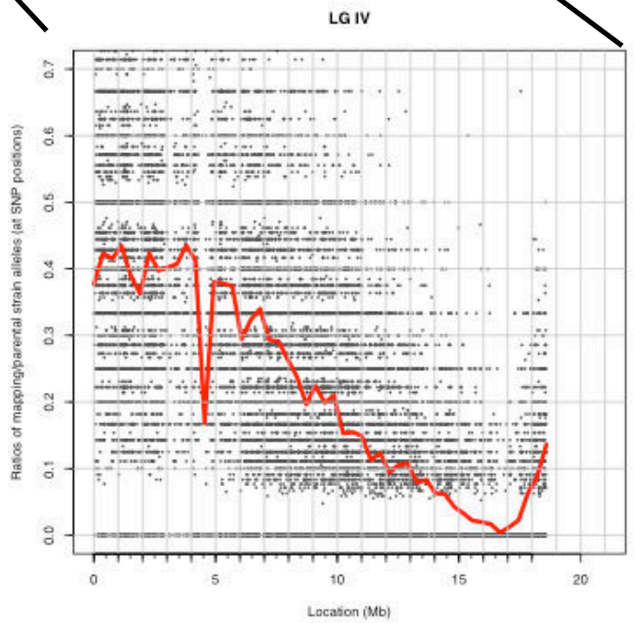
Supports Arabidopsis thaliana (and other species)

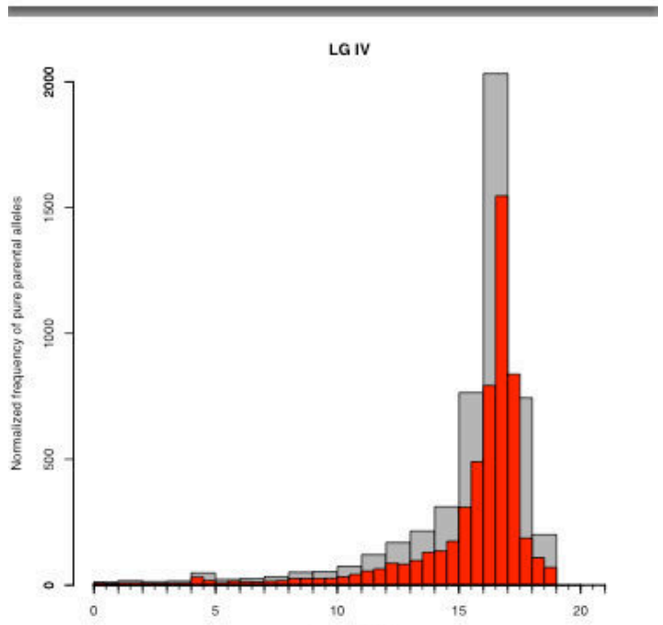
Arabadopsis_AT4G35090gene_Galaxy39-[CloudMap_SNP_mapping_with_WGS].pdf



y-axis: relative parental allele frequency





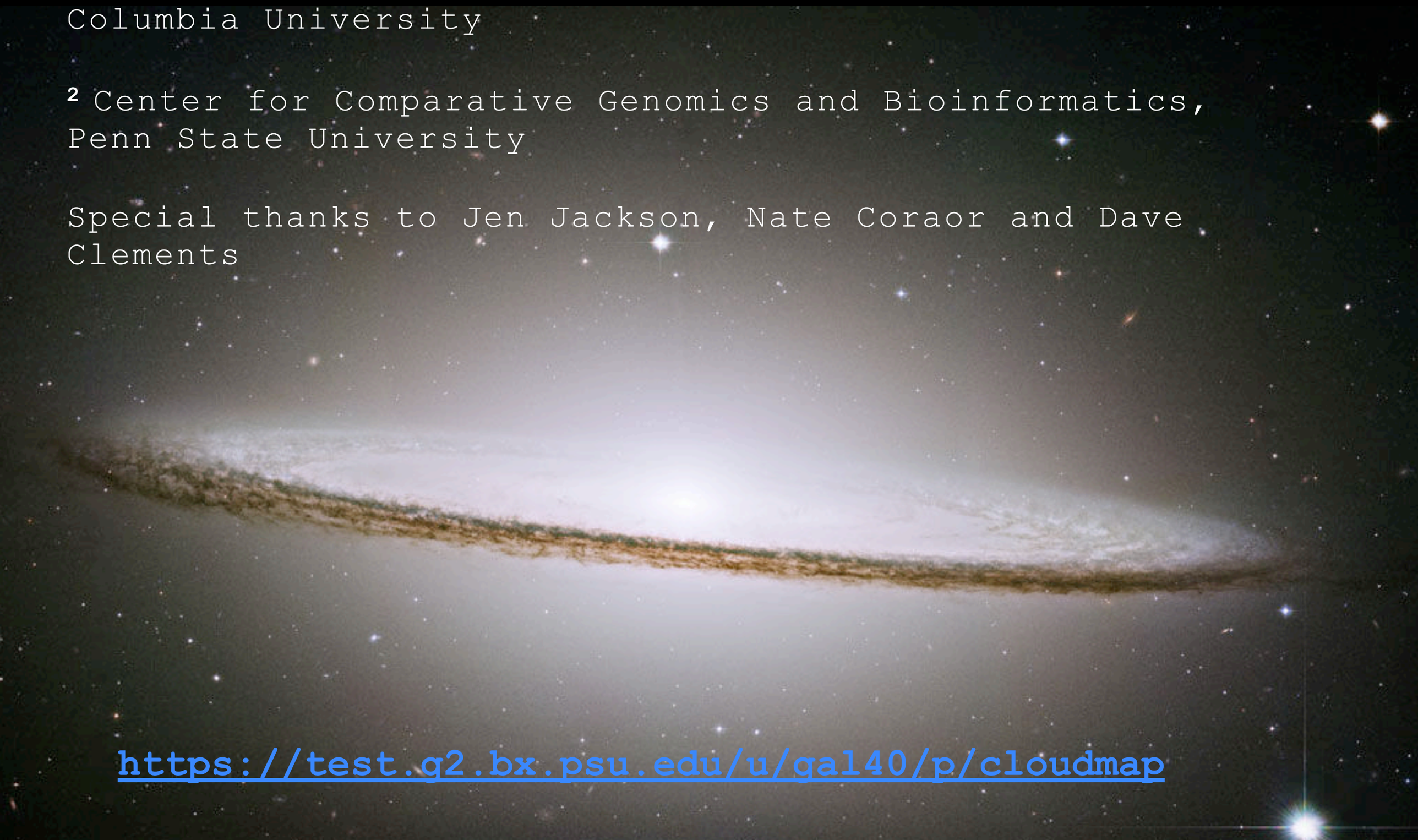


Gregory Minevich[1], Danny S. Park[1], Richard Poole[1], Daniel Blankenberg[2], Anton Nekrutenko[2], Oliver Hobert[1]

[1] Department of Biochemistry and Molecular Biophysics, HHMI Columbia University

[2] Center for Comparative Genomics and Bioinformatics, Penn State University

Special thanks to Jen Jackson, Nate Coraor and Dave Clements



https://test.g2.bx.psu.edu/u/ga140/p/cloudmap