

CoronaSV

Awesome Group 3

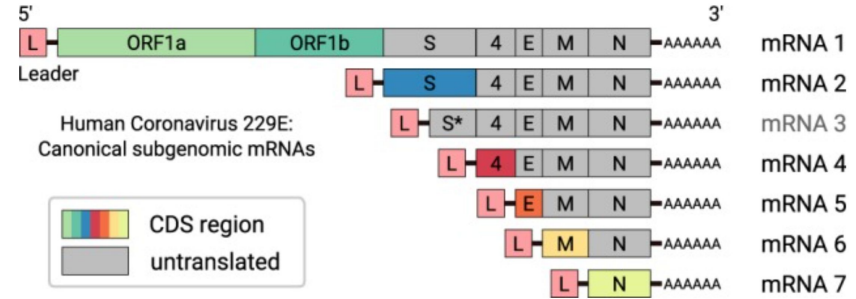
Background, or why SARS-CoV-2 + SVs? (Todd)

ARTICLES | VOLUME 396, ISSUE 10251, P603-611, AUGUST 29, 2020

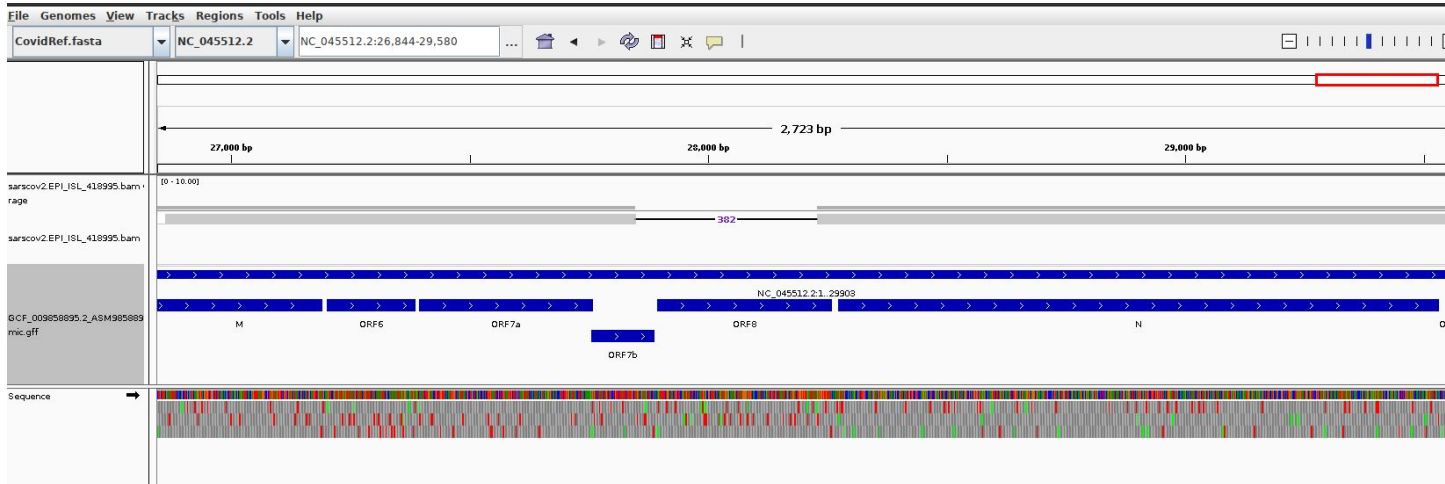
Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study

Barnaby E Young, MB BChir [†] • Siew-Wai Fong, PhD [†] • Yi-Hao Chan, PhD [†] • Tze-Minn Mak, PhD • Li Wei Ang, MSc • Danielle E Anderson, PhD • et al. [Show all authors](#) • [Show footnotes](#)

Published: August 18, 2020 • DOI: [https://doi.org/10.1016/S0140-6736\(20\)31757-8](https://doi.org/10.1016/S0140-6736(20)31757-8) • [Check for updates](#)



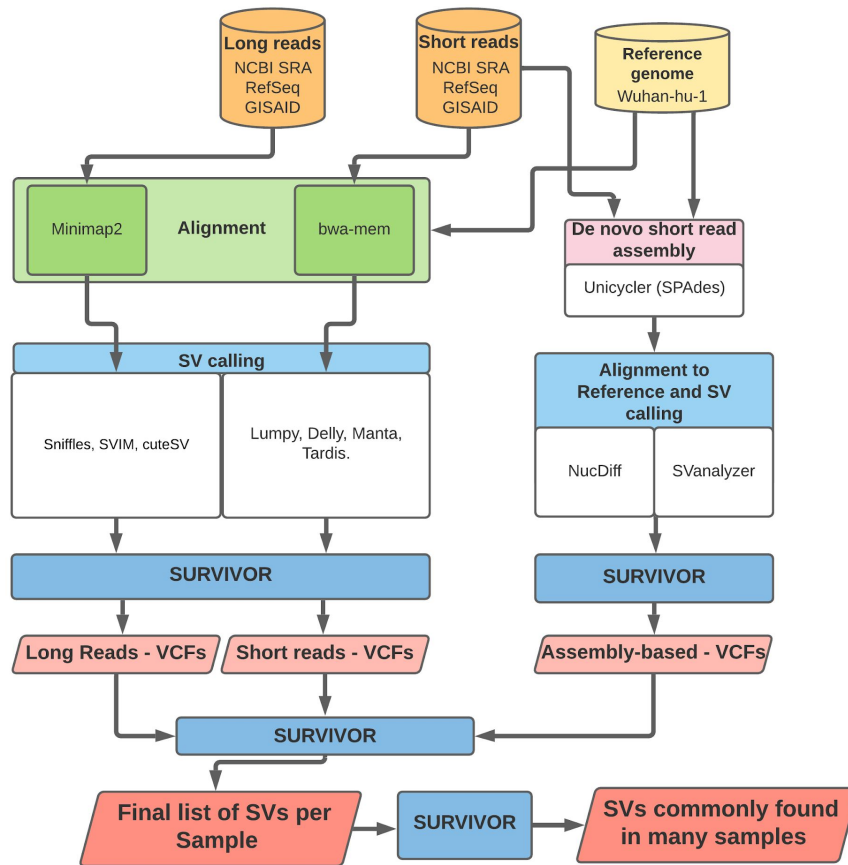
Verified the 382 bp deletion of ORF8 (Based on available public assembly)



CoronaSV Overview (Arda)

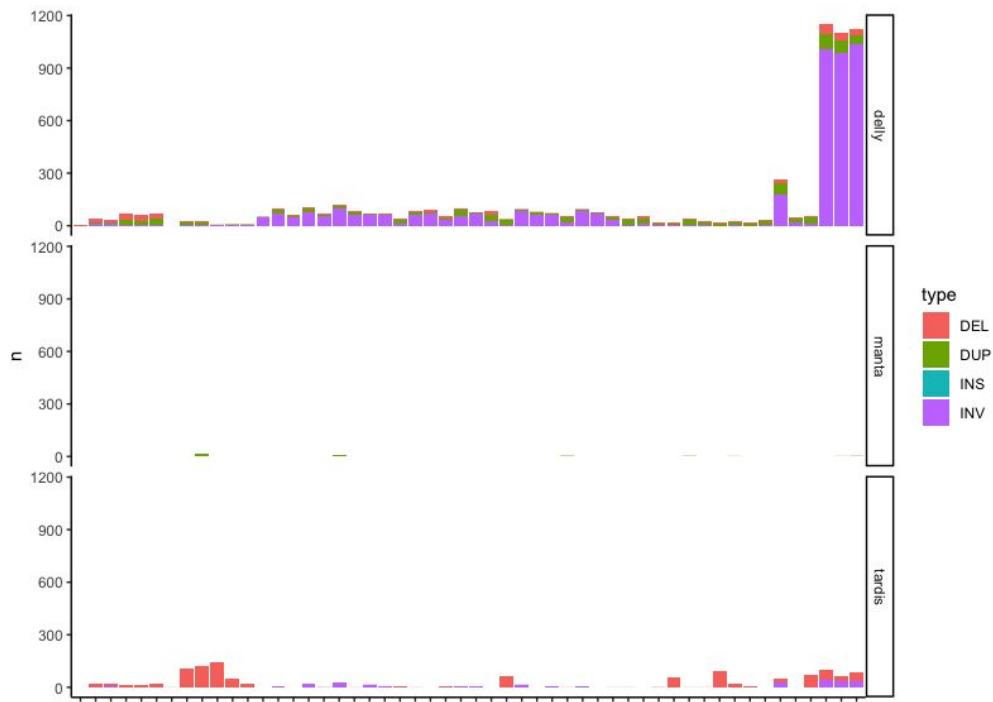
Highlights

- Performs read mapping, genome assembly, and genome alignment
- 9 total SV callers included
- Supports short reads, long reads, and genomes
- SURVIVOR used to merge and calculate support for SVs



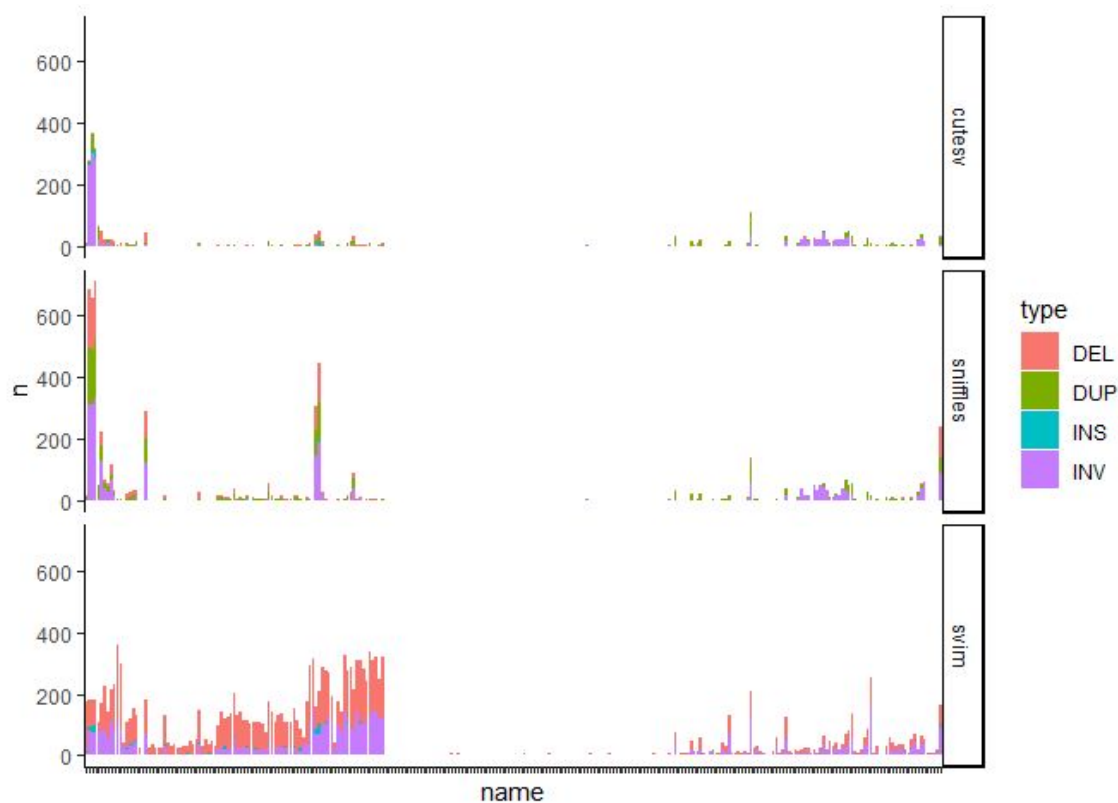
Short read SV calling (Daniela)

- 56 Illumina PE reads
- Four SV callers
 - Delly
 - Lumpy
 - Manta
 - Tardis



Long read SV calling (Shangzhe)

- 298 Nanopore accessions
 - Nanoflt
 - Minimap2
 - Sniffles, cuteSV, SVIM
-
- Overview of the long-read results



Assembly based SV calling (Max)

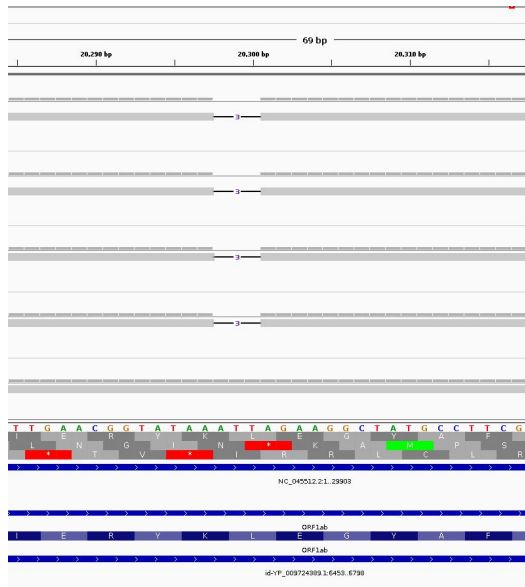
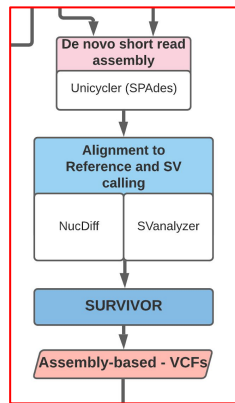
Results:

- A) Short read assemblies (SPAdes) were of varying quality (some fragmented ...)
- B) No significant SV were detected using an assembly based method
- C) Laughable deletions were in fact detected (3 base pairs)

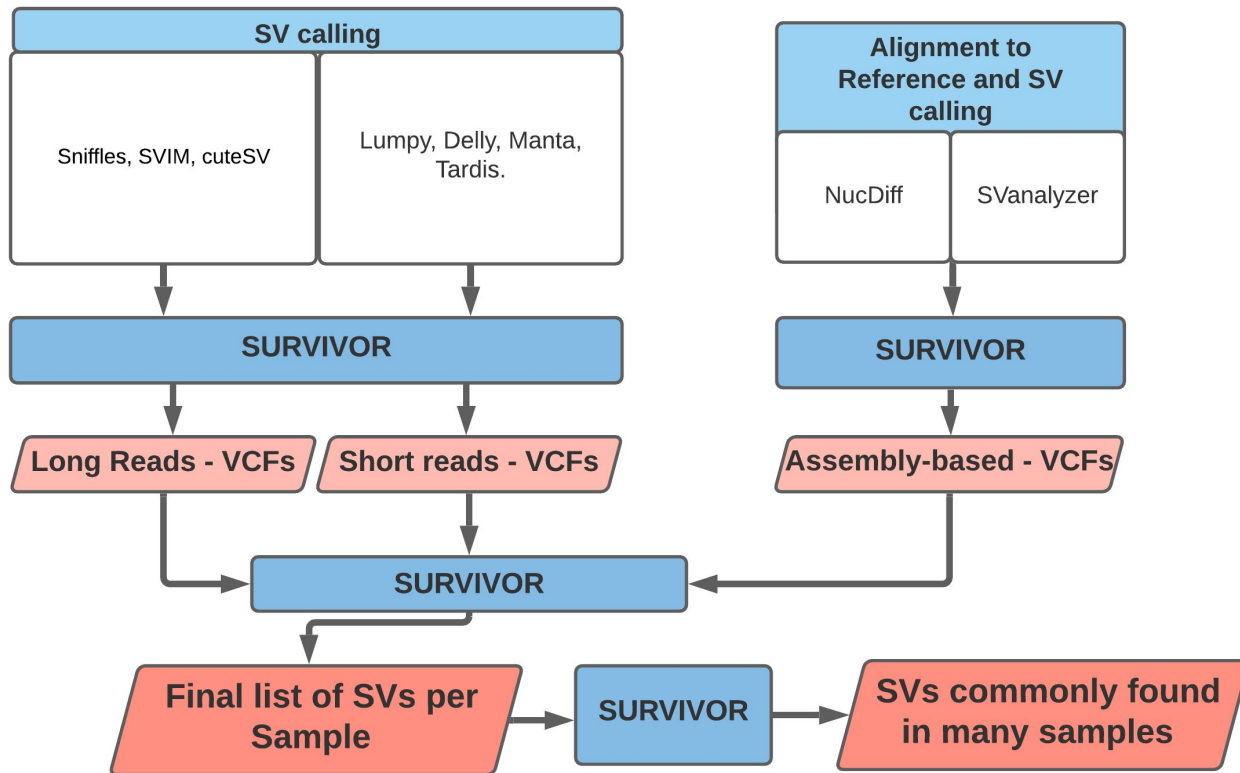
1)



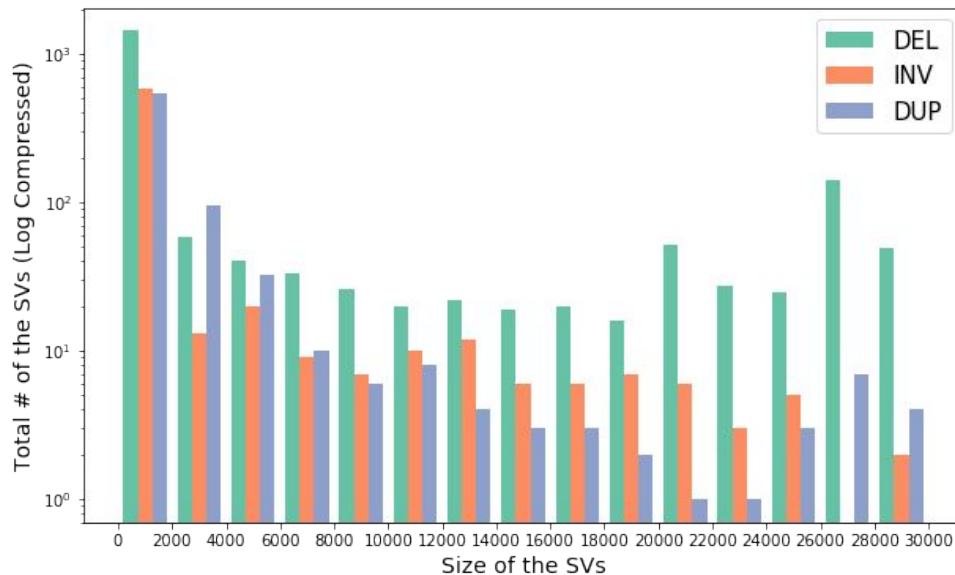
2)



SURVIVOR description (Dan)



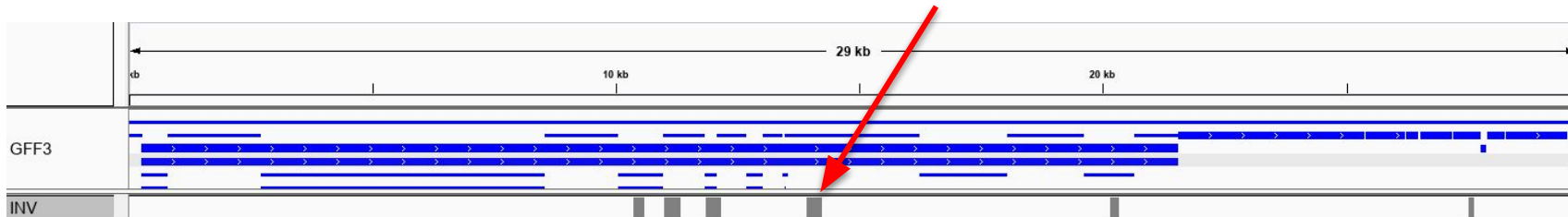
Results (Louis)



More than 400 SRA runs, 200GB of the data has been processed.

Histogram showing high confidence SVs generated with SURVIVOR by taking a majority vote across multiple SV callers.

SVs shared across multiple samples, few inversions survived. Those inversion may related to transcriptional landscape.





Yunxi Liu



Todd Treangen



Maximilian Marin



Daniela Soto



shangzhe zhang



Daniel Agostinho



Arda

- Yunxi Liu
- Todd Treangen
- Daniela Soto
- Shangzhe Zhang

- Arda Söylev
- Daniel Agostinho
- Daniel's Cat
- Max Marin

Extra slides

CoronaSV hacking

October 11, 2020 – October 14, 2020

Period: 3 days ▾

Overview

0 Active Pull Requests

0 Active Issues

0

Merged Pull Requests

0

Open Pull Requests

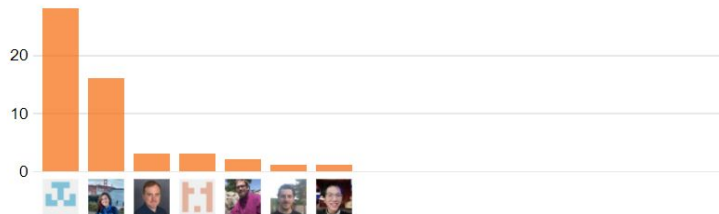
0

Closed Issues

0

New Issues

Excluding merges, **7 authors** have pushed **54 commits** to main and **54 commits** to all branches. On main, **97 files** have changed and there have been **13,287 additions** and **67 deletions**.






















CoronaSV

- bam_files
- long_reads_sv
- preliminaryVCFs
- raw_reads_hcov_229e
- raw_reads_sars_cov_2
- reference_genome
- results
- test_vcf
- trimmed_reads
- vcf_files
- vcf_split

All Projects > CoronaSV

Current Folder Only Any Name Any ID Any Type Any Class

<input type="checkbox"/>	Name ^	Type / Class	File Size	Created	Sponsored Until	Status
<input type="checkbox"/>	 bam_files	Folder	--	--	--	--
<input type="checkbox"/>	 long_reads_sv	Folder	--	--	--	--
<input type="checkbox"/>	 preliminaryVCFs	Folder	--	--	--	--
<input type="checkbox"/>	 raw_reads_hcov_229e	Folder	--	--	--	--
<input type="checkbox"/>	 raw_reads_sars_cov_2	Folder	--	--	--	--
<input type="checkbox"/>	 reference_genome	Folder	--	--	--	--
<input type="checkbox"/>	 results	Folder	--	--	--	--
<input type="checkbox"/>	 test_vcf	Folder	--	--	--	--
<input type="checkbox"/>	 trimmed_reads	Folder	--	--	--	--
<input type="checkbox"/>	 vcf_files	Folder	--	--	--	--
<input type="checkbox"/>	 vcf_split	Folder	--	--	--	--
<input type="checkbox"/>	 data_fetching	Workflow	--	10/11/2020 6:35 PM	--	Ready
<input type="checkbox"/>	 ERR4173381.vcf	File	5.84 KiB	10/13/2020 4:56 PM	--	Live
<input type="checkbox"/>	 ERR4173381.vcf	File	5.84 KiB	10/13/2020 4:58 PM	--	Live
<input type="checkbox"/>	 HCoV_229E_meta.csv	File	1.52 KiB	10/11/2020 7:27 PM	--	Live
<input type="checkbox"/>	 results.tar.gz	File	392.4 KiB	10/14/2020 11:46 AM	--	Live
<input type="checkbox"/>	 SARS_CoV_2_meta.csv	File	2.84 KiB	10/11/2020 7:27 PM	--	Live
<input type="checkbox"/>	 Untitled Workflow - Oct 13th 2020 9:07pm	Workflow	--	10/13/2020 1:07 PM	--	Ready
<input type="checkbox"/>	 vcf_split_with_empty.tar.gz	File	3.02 MiB	10/13/2020 9:07 PM	--	Live

Conclusions (Todd, add group photo)

- Finalize individual pipelines (short read, long read, assembly based)
- Finalize SURVIVOR integration
- Explore adding Trinity/transcript assembler for tracking canonical and noncanonical junctions
- Integrate everything into a single snakemake workflow
- Run CoronaSV on 100s of samples and analyze results

and...

- Win awesome prize!

Dataset description

- A set of 45 isolates with paired Illumina & ONT sequencing data
- 138 SRA sequencing runs (86 Illumina, 52 ONT)
- Raw reads were trimmed and filtered in the pre-processing step
 - Trimmomatic for short reads
with universal adapters
 - NanoFilt for long reads

Links from BioSample

Items: 2

☐ [Illumina NovaSeq 6000 paired end sequencing](#)

1. 1 ILLUMINA (Illumina NovaSeq 6000) run: 29.4M spots, 8.9G bases, 2.6Gb downloads
Accession: ERX4136993

☐ [GridION sequencing](#)

2. 1 OXFORD_NANOPORE (GridION) run: 1.2M spots, 592.3M bases, 536.8Mb downloads
Accession: ERX4136994

Metadata

CoronaSV_metadata

File Edit View Insert Format Data Tools Add-ons Help Last edit was made 21 mi...

100% View only

fx	BioProject							
	A	B	C	D	E	F	G	
1	BioProject	SRA_Study	Run	Experiment	BioSample	SRA_Sample	Assay_Type	P
2	PRJEB33797	ERP116617	ERR3460958	ERX3482709	SAMEA5841278	ERS3629828	RNA-Seq	IL
3	PRJEB33797	ERP116617	ERR3460959	ERX3482710	SAMEA5841279	ERS3629829	RNA-Seq	IL
4	PRJEB33797	ERP116617	ERR3460960	ERX3482711	SAMEA5841280	ERS3629830	RNA-Seq	IL
5	PRJEB33797	ERP116617	ERR3460961	ERX3482712	SAMEA5841281	-	RNA-Seq	C
6	PRJEB33797	ERP116617	ERR3460962	ERX3482713	SAMEA5841282	-	RNA-Seq	C
7	PRJEB38459	ERP121890	ERR4173381	ERX4136993	SAMEA6853602	ERS4581192	RNA-Seq	IL
8	PRJEB38459	ERP121890	ERR4173382	ERX4136994	SAMEA6853602	ERS4581192	AMPLICON	C
9	PRJNA607948	SRP250294	SRR11140744	SRX7777166	SAMN14154205	SRS6189924	WGS	IL
10	PRJNA607948	SRP250294	SRR11140745	SRX7777165	SAMN14154204	SRS6189920	WGS	C
11	PRJNA607948	SRP250294	SRR11140746	SRX7777164	SAMN14154203	SRS6189919	WGS	IL
12	PRJNA607948	SRP250294	SRR11140747	SRX7777163	SAMN14154202	SRS6189917	WGS	C
13	PRJNA607948	SRP250294	SRR11140748	SRX7777162	SAMN14154201	SRS6189918	WGS	IL
14	PRJNA607948	SRP250294	SRR11140749	SRX7777161	SAMN14154200	SRS6189916	WGS	C
15	PRJNA607948	SRP250294	SRR11140750	SRX7777160	SAMN14154199	SRS6189914	WGS	IL
16	PRJNA607948	SRP250294	SRR11140751	SRX7777159	SAMN14154198	SRS6189915	WGS	C
17	PRJNA616147	SRP254488	SRR11445485	SRX8023307	SAMN14483190	SRS6395996	AMPLICON	C
18	PRJNA616147	SRP254488	SRR11445486	SRX8023306	SAMN14483189	SRS6395995	AMPLICON	C
19	PRJNA616147	SRP254488	SRR11514749	SRX8086361	SAMN14483190	SRS6395996	AMPLICON	IL
20	PRJNA616147	SRP254488	SRR11514750	SRX8086360	SAMN14483189	SRS6395995	AMPLICON	IL
21	PRJNA616147	SRP254488	SRR11542243	SRX8112342	SAMN14596409	SRS6477428	AMPLICON	IL

Reference

Metadata

Paired data

Assemblies