

# CoronaSV

## Group 3



# CoronaSV team

- Daniel Agostinho (Washington University in St Louis School of Medicine, MO)
- Yunxi Liu (Rice University, Houston, TX)
- Todd Treangen (Rice University, Houston, TX)
- Daniela Soto (UC Davis, Davis, CA)
- Shangzhe Zhang (Lanzhou University, Lanzhou, China)
- Max Marin (Harvard Medical School, Boston, MA)
- Arda Söylev (Konya Food and Agriculture University, Turkey)

# Background

Research | Open Access | Published: 31 August 2020

## Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1

[Francesca Benedetti](#), [Greg A. Snyder](#), [Marta Giovanetti](#), [Silvia Angeletti](#), [Robert C. Gallo](#), [Massimo Ciccozzi](#)  
✉ & [Davide Zella](#) ✉

Journal of Translational Medicine 18, Article number: 329 (2020) | Cite this article

## Method

# Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis


Adrian Viehweger,<sup>1,2,5</sup> Sebastian Krautwurst,<sup>1,2,5</sup> Kevin Lamkiewicz,<sup>1,2</sup>  
Ramakanth Madhugiri,<sup>3</sup> John Ziebuhr,<sup>2,3</sup> Martin Hölzer,<sup>1,2</sup> and Manja Marz<sup>1,2,4</sup>

<sup>1</sup>RNA Bioinformatics and High-Throughput Analysis, Friedrich Schiller University Jena, 07743 Jena, Germany; <sup>2</sup>European Virus Bioinformatics Center, Friedrich Schiller University Jena, 07743 Jena, Germany; <sup>3</sup>Institute of Medical Virology, Justus Liebig University Gießen, 35390 Gießen, Germany; <sup>4</sup>Leibniz Institute on Aging-Fritz Lipmann Institute, 07743 Jena, Germany

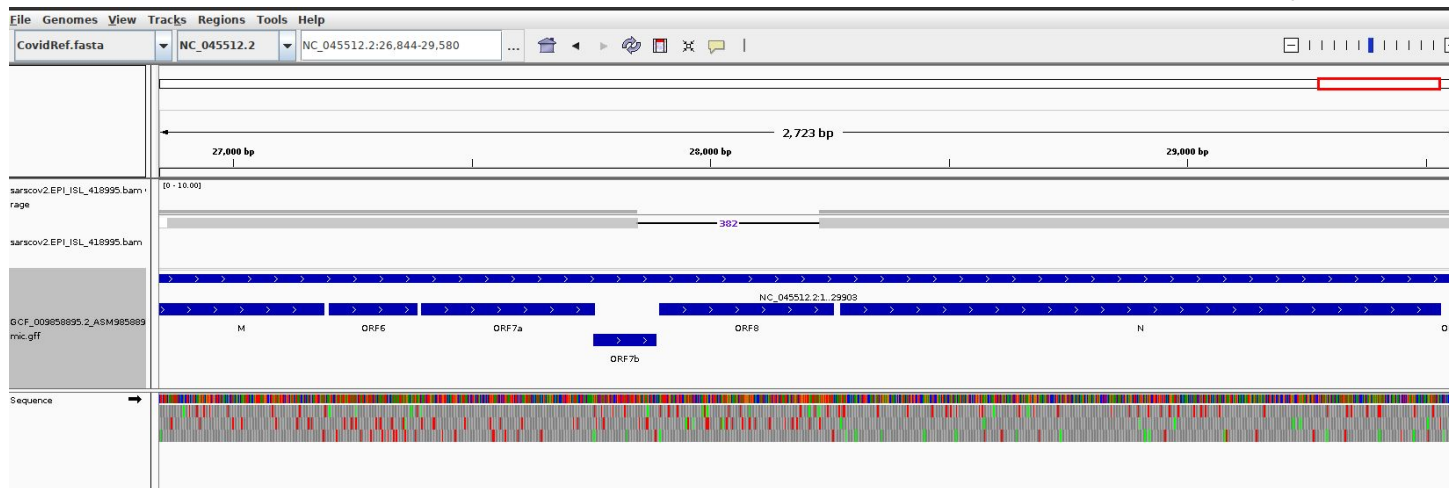
ARTICLES | VOLUME 396, ISSUE 10251, P603-611, AUGUST 29, 2020

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

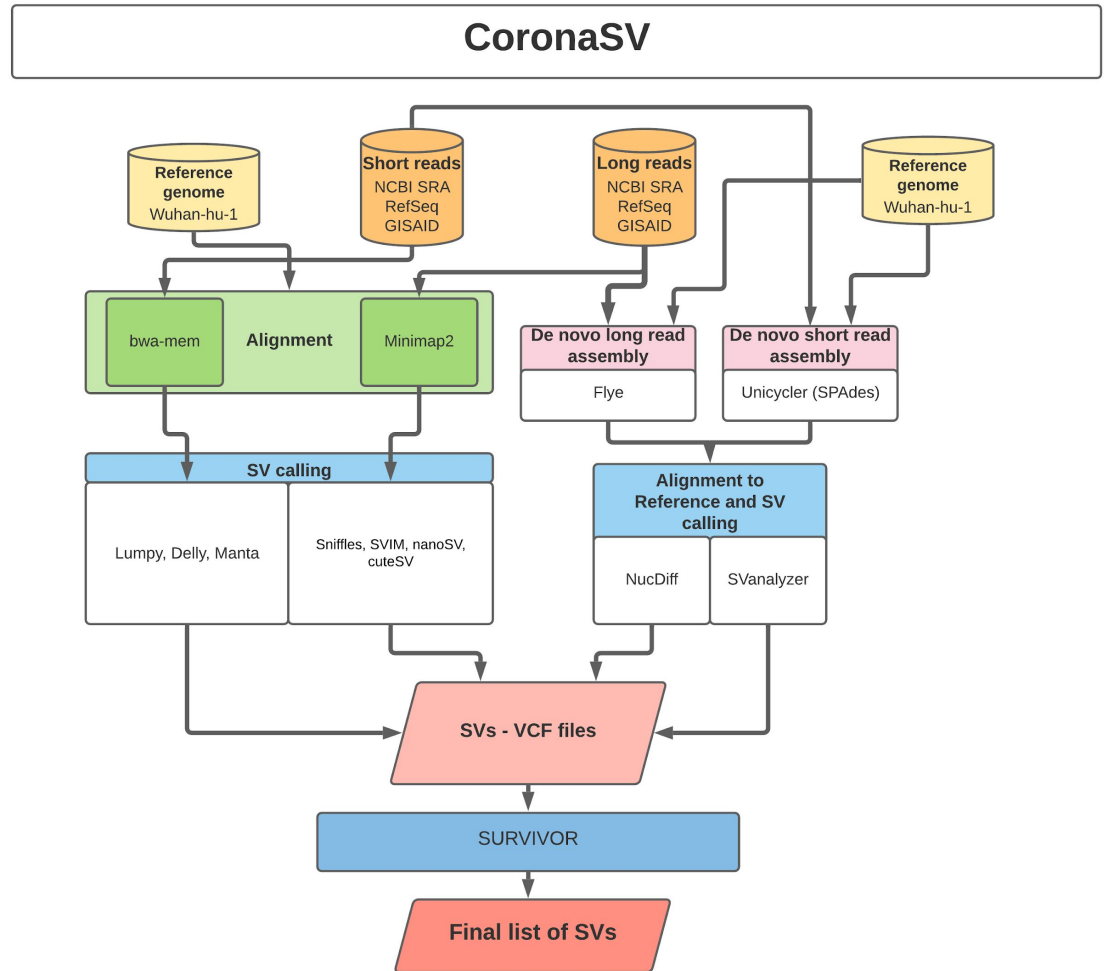
Barnaby E Young, MB BChir <sup>†</sup> • Siew-Wai Fong, PhD <sup>†</sup> • Yi-Hao Chan, PhD <sup>†</sup> • Tze-Minn Mak, PhD • Li Wei Ang, MSc  
Danielle E Anderson, PhD • et al. [Show all authors](#) • [Show footnotes](#)

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)



- We have observed discordant results across low frequency single nucleotide variant (SNV) calls for SARS-CoV-2
- Our hypothesis is that we will see similar (if not stronger) discordance for SV calls for SARS-CoV-2 given its complex transcriptional landscape and varying sequencing strategies and biases
- We will test this hypothesis by combining short read variant calling + long read variant calling + assembly based variant calling and using SURVIVOR to see which SVs “survive”



# Dataset description

- A set of 45 isolates with paired Illumina & ONT sequencing data
- 139 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...

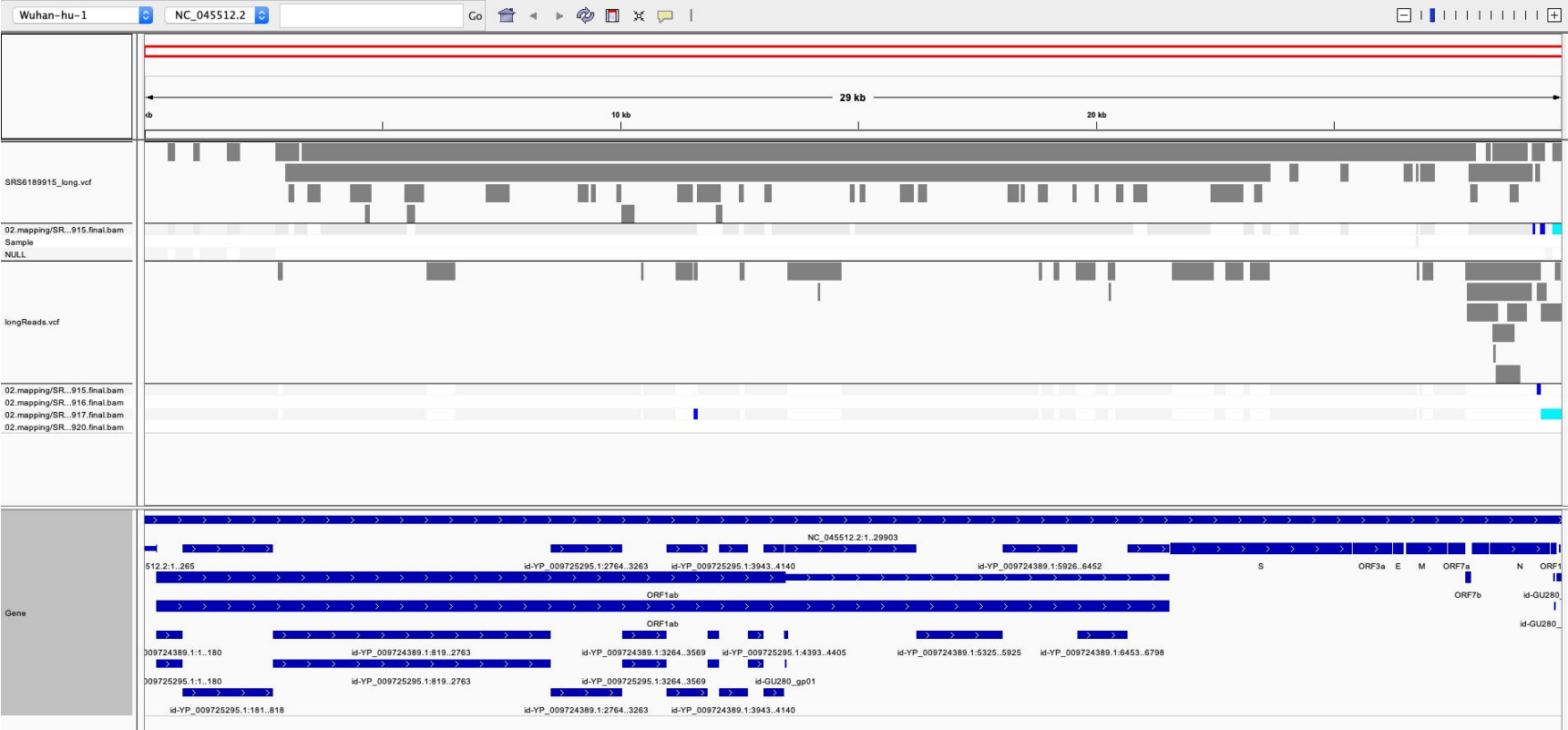
🖨️ 🔍 100% View only

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

# IGV SURVIVOR



# Next steps

- 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!