



Presentazione per Sacco

Gruppo Prof. Stefano Ceri, con Anna Bernasconi e Pietro Pinoli

Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano

4 Febbraio 2022





Phenotype data representation



Phenotype / clinical data representation

<http://gmql.eu/phenotype/>



The patient phenotype definition (~150 attributes) was used as a standard to collect and harmonize data from studies in the COVID-19 Host Genetics Initiative (<https://www.covid19hg.org/>)

Focus on Patient and his/her data at:

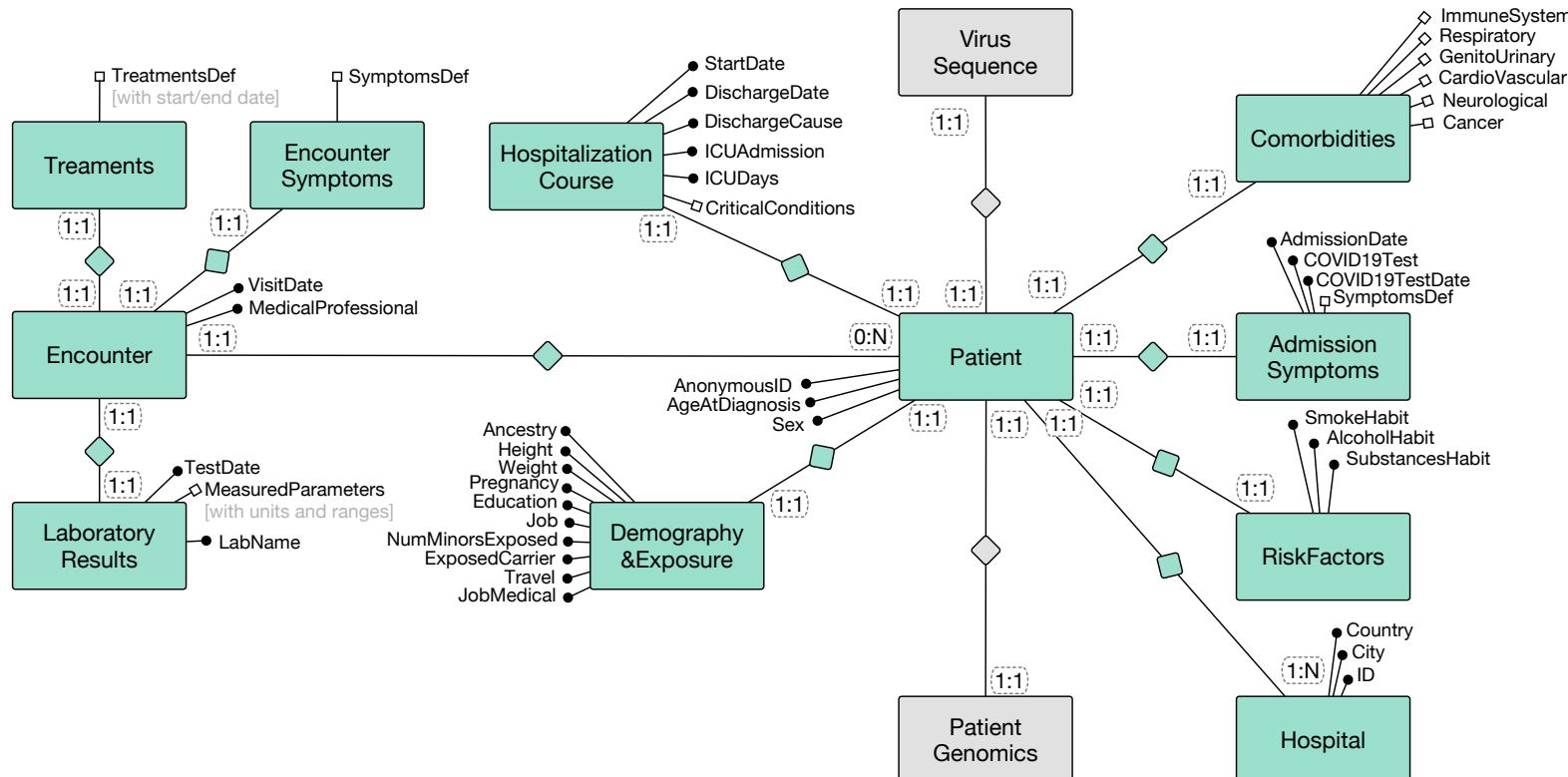
- Admission
- Course of hospitalization
- Discharge

Data collected through this are hosted by the **European Genome-phenome Archive** of EMBL-EBI

COVID-19 Host Genetics Initiative Coordination,

Data dictionary working group:

Stefano Ceri and **Anna Bernasconi** (Politecnico di Milano), Alessandra Renieri and Francesca Mari (Università degli Studi di Siena)



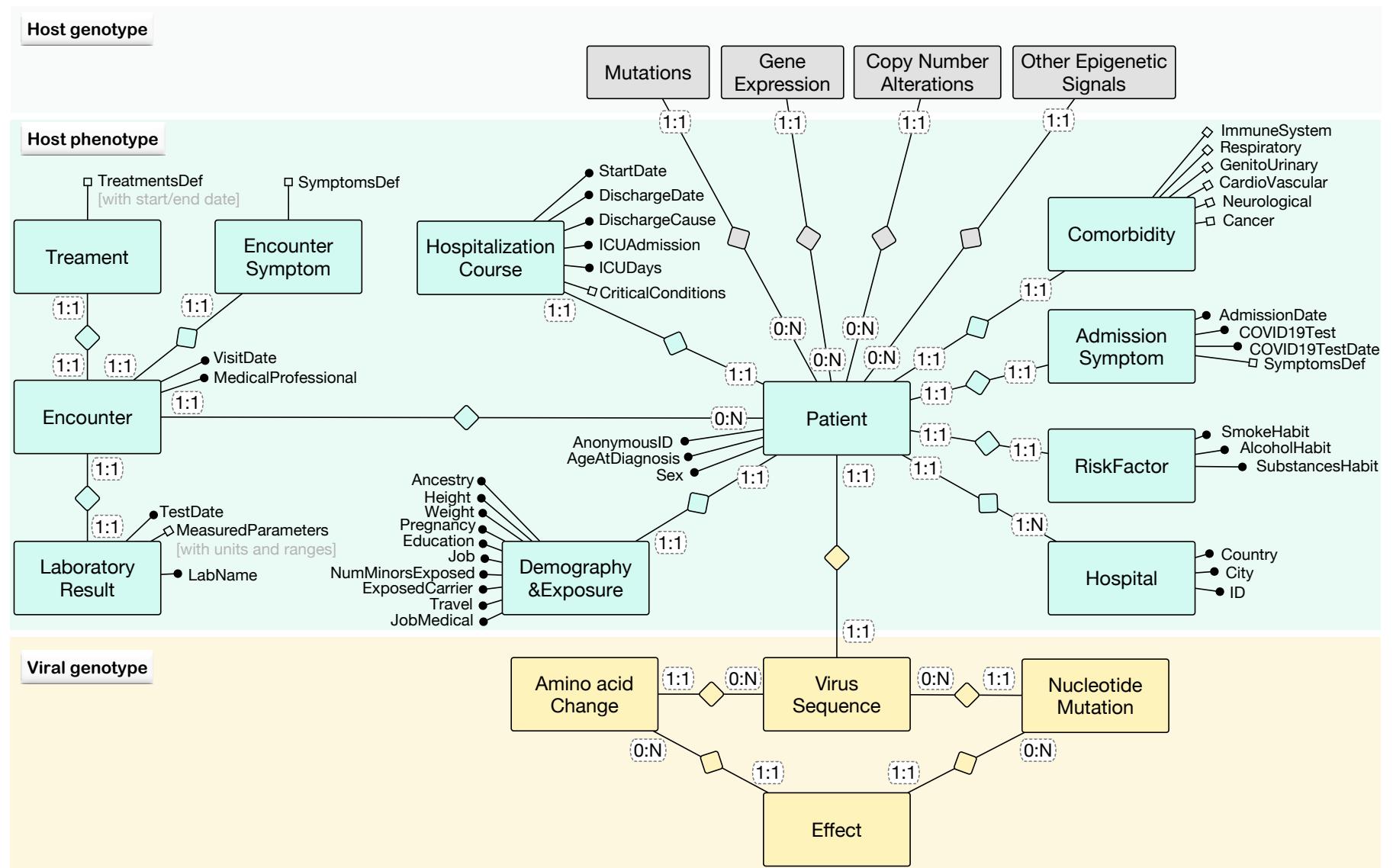
Used for studies reported in the following publications:

COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-19. Nature 600, 472–477 (2021).

<https://doi.org/10.1038/s41586-021-03767-x>

Van Blokland, I.V., ..., Lifelines COVID-19 cohort study, The COVID-19 Host Genetics Initiative, et al. (2021) Using symptom-based case predictions to identify host genetic factors that contribute to COVID-19 susceptibility. PloS one, 16(8), p.e0255402. <https://doi.org/10.1371/journal.pone.0255402>

Linking three systems



<http://gmql.eu/genosurf/>



GenoSurf

e.g., data dictionary of
<https://www.covid19hg.org/>

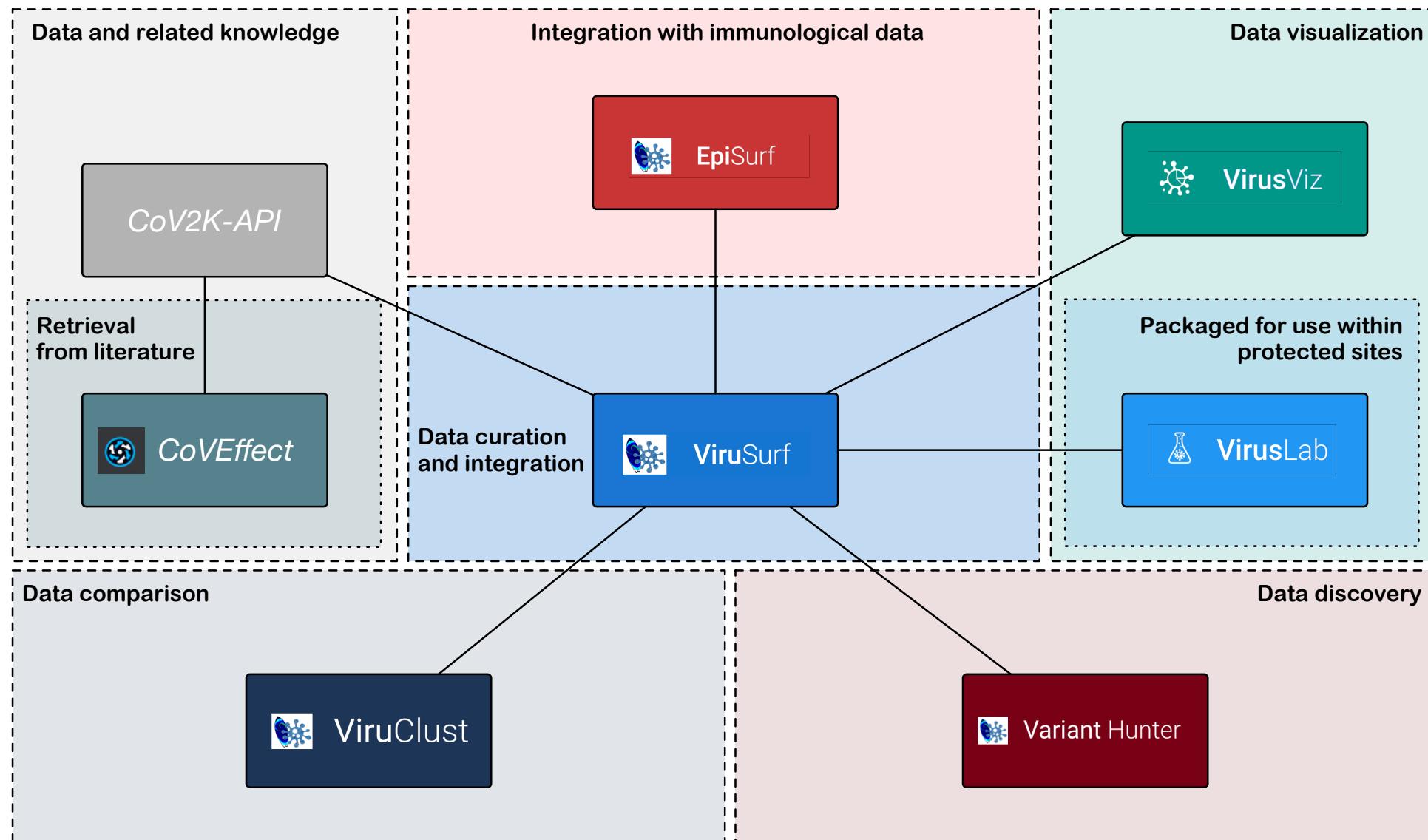
PhenotypeDB

<http://gmql.eu/virusurf/>



ViruSurf

Current focus on tools for supporting SARS-CoV-2 research



VariantHunter



Supporting two working modes:

- **Lineage independent** mutation growth
- **Lineage dependent** mutation growth
- ... to appear soon **Lineage recombination**

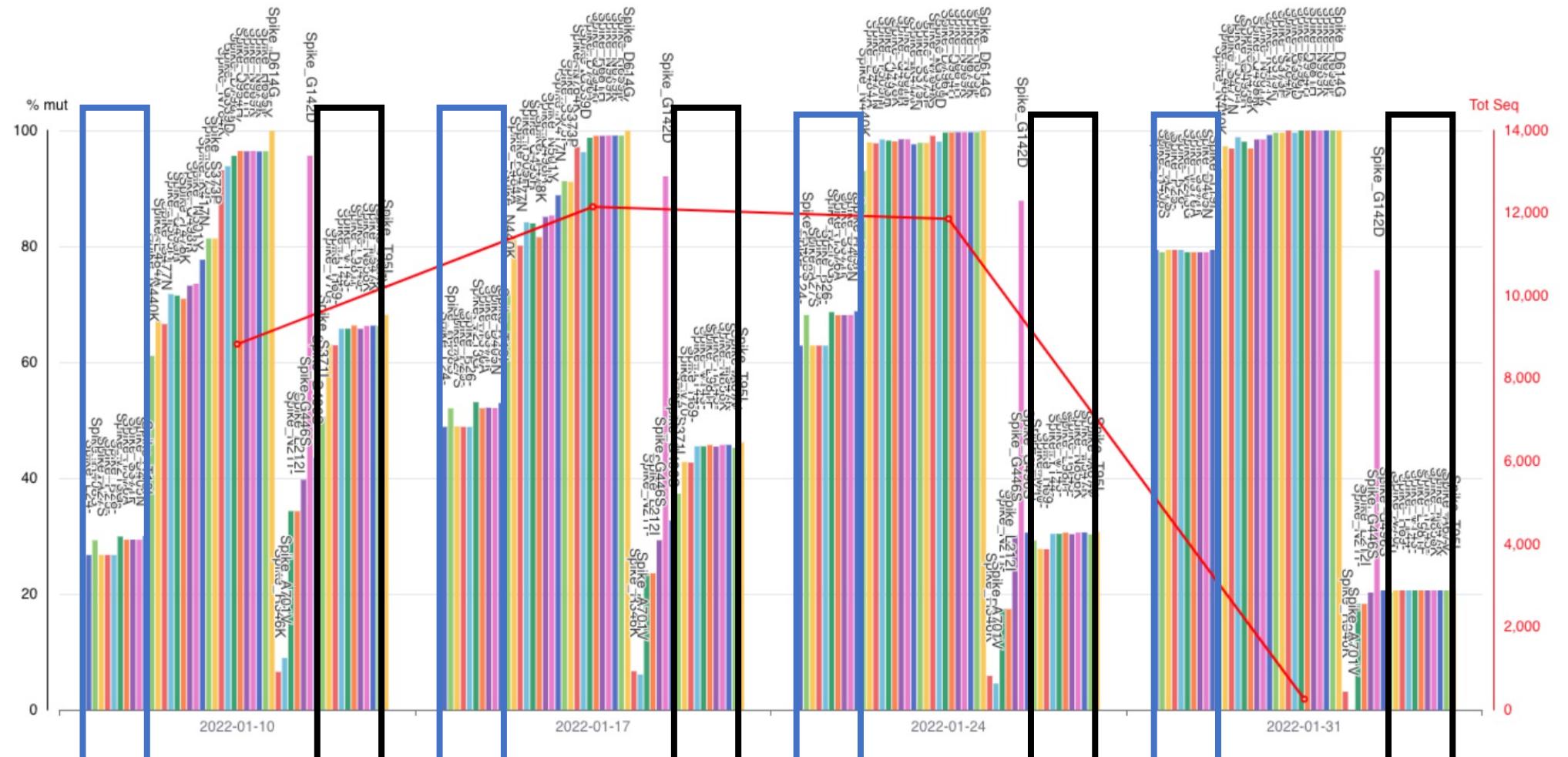
The screenshot shows the Variant Hunter web application interface. At the top, there is a dark red header bar with the title "Variant Hunter" on the left and a menu icon on the right. Below the header, there are two tabs: "LINEAGE INDEPENDENT" (which is highlighted in white) and "LINEAGE SPECIFIC". The main content area has a yellow background and features a red rectangular form titled "DEFINE ANALYSIS:". This form contains several input fields: "Granularity:" dropdown set to "world", "Location:" dropdown, "Date:" input field set to "2022-01-08" with a calendar icon, "# Week:" dropdown set to "4", and a "START ANALYSIS" button at the bottom.

Mutations that are growing fast in Denmark



Location	Protein	Mut	Slope ↓ 1	Y-intercept	Presence (%) 2022-01-10	Presence (%) 2022-01-17	Presence (%) 2022-01-24	Presence (%) 2022-01-31	P-value with mut	P-value without mut
Denmark	Spike	L24-	17	28.64	26.71 (2360)	48.83 (5936)	62.86 (7460)	79.38 (204)	0.0	0.0
Denmark	Spike	R408S	17	32.30	29.26 (2585)	52.02 (6323)	68.15 (8087)	78.99 (203)	0.0	0.0
Denmark	Spike	A27S	17	28.69	26.73 (2362)	48.91 (5945)	62.88 (7462)	79.38 (204)	0.0	0.0
Denmark	Spike	P25-	17	28.64	26.71 (2360)	48.83 (5936)	62.86 (7460)	79.38 (204)	0.0	0.0
Denmark	Spike	P26-	17	28.64	26.71 (2360)	48.83 (5936)	62.86 (7460)	79.38 (204)	0.0	0.0
Denmark	Spike	V213G	16	33.24	29.90 (2642)	53.09 (6454)	68.65 (8147)	78.99 (203)	0.0	0.0
Denmark	Spike	T376A	16	32.38	29.36 (2594)	52.04 (6326)	68.11 (8083)	78.99 (203)	0.0	0.0
Denmark	Spike	S371F	16	32.43	29.37 (2595)	52.14 (6338)	68.13 (8085)	78.99 (203)	0.0	0.0
Denmark	Spike	D405N	16	32.39	29.35 (2593)	52.08 (6331)	68.16 (8088)	78.99 (203)	0.0	0.0
Denmark	Spike	T19I	16	33.19	30.01 (2651)	52.96 (6438)	68.79 (8163)	79.38 (204)	0.0	0.0
Denmark	Spike	N440K	12	62.18	61.10 (5398)	71.96 (8748)	93.03 (11040)	93.39 (240)	3.5e-201	0.0
Denmark	Spike	E484A	11	69.89	67.16 (5934)	81.33 (9886)	97.95 (11624)	97.28 (250)	8.3e-203	0.0

BA.1 vs BA.2 (Denmark)



BA.1 vs BA.2 (Denmark)



PROTEIN	BA1	BA.2
SPIKE		S:T19I
		S:L24-
		S:P25-
		S:P26-
		S:A27S
	S:A67V	
	S:H69-	
	S:V70-	
	S:T95I	
	S:G142-	S:G142D
	S:V143-	
	S:Y144-	
	S:Y145D	
	S:N211-	
	S:L212I	
		S:V213G
	S:G339D	S:G339D
	S:S371L	S:S371F
	S:S373P	S:S373P
	S:S375F	S:S375F
		S:T376A
		S:D405N
		S:R408S
	S:K417N	S:K417N
	S:N440K	S:N440K
	S:G446S	
	S:S477N	S:S477N
	S:T478K	S:T478K
	S:E484A	S:E484A
	S:Q493R	S:Q493R
	S:G496S	
	S:Q498R	S:Q498R
	S:N501Y	S:N501Y
	S:Y505H	S:Y505H
	S:T547K	
	S:D614G	S:D614G
	S:H655Y	S:H655Y
	S:N679K	S:N679K
	S:P681H	S:P681H
	S:N764K	S:N764K
	S:D796Y	S:D796Y
	S:N856K	
	S:Q954H	S:Q954H
	S:N969K	S:N969K
	S:L981F	

Growing mutations of BA.1 in United Kingdom at the end of 2021



Sorted by descending slope

United Kingdom / BA.1 / 2021-12-31 / 4 weeks

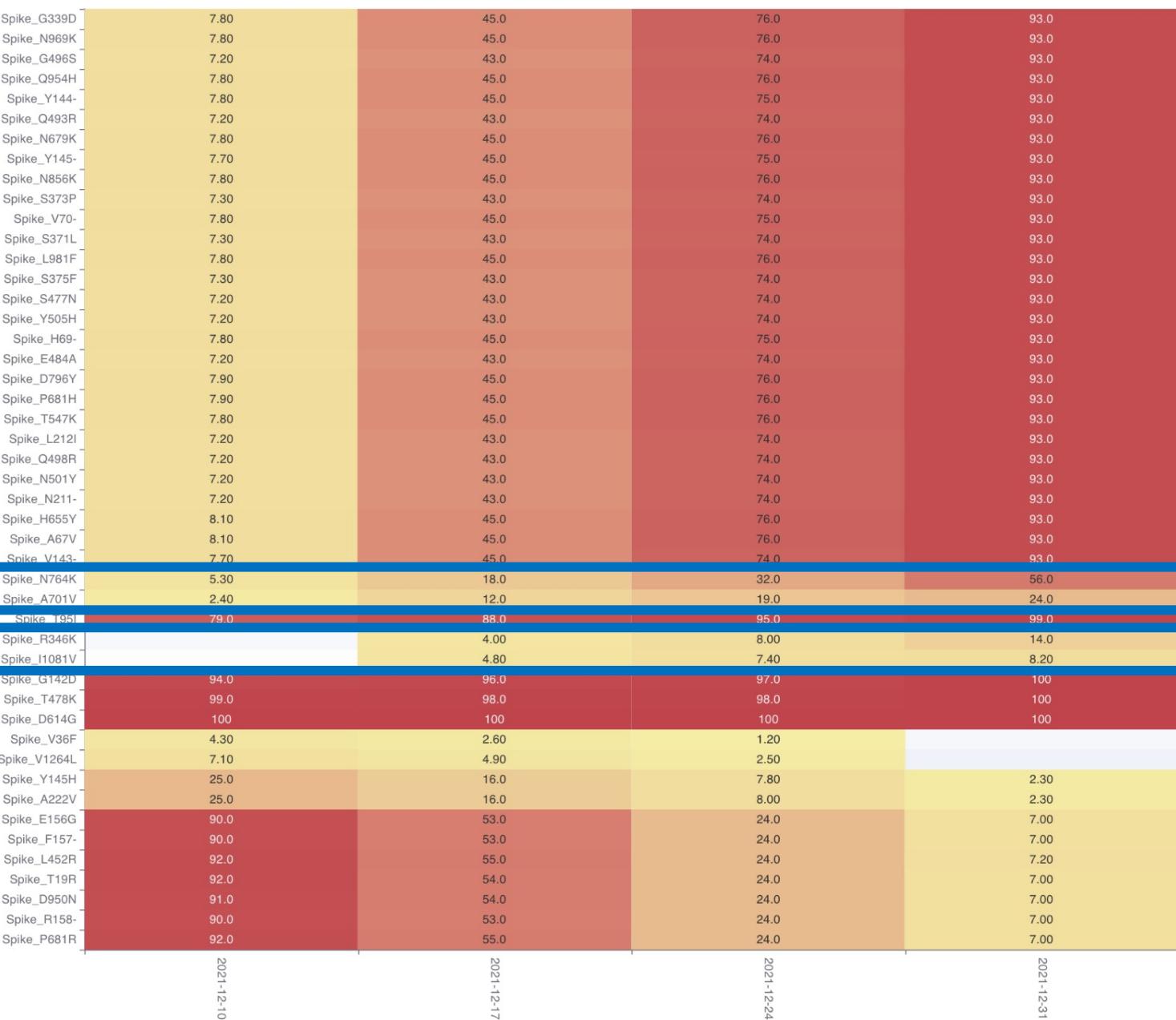
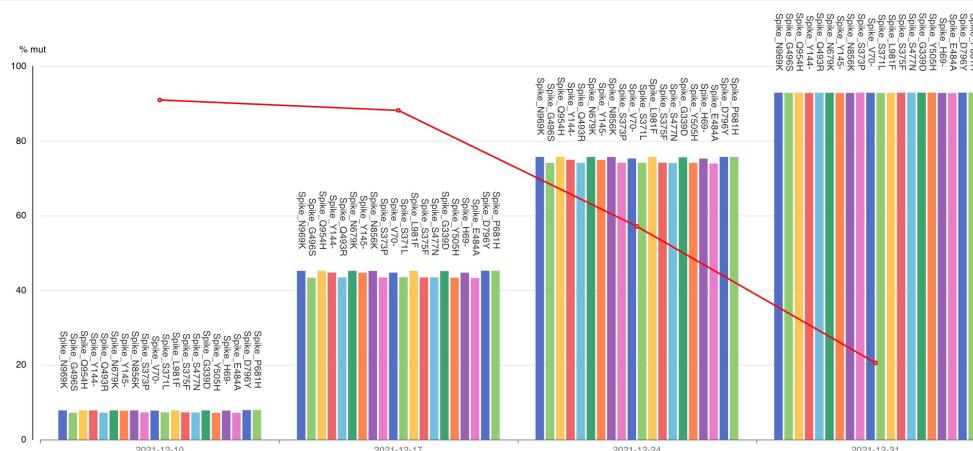
TABLE 

-	Location	Protein	Mut	Slope ↓ 1	Y-intercept	Presence (%) 2021-12-10	Presence (%) 2021-12-17	Presence (%) 2021-12-24	Presence (%) 2021-12-31
<input type="checkbox"/>	United Kingdom	Spike	G339D	29	12.50	7.82 (4976) (27843)	45.12 (27843)	75.58 (30195)	92.88 (13352)
<input type="checkbox"/>	United Kingdom	Spike	N969K	29	12.54	7.82 (4980) (27881)	45.19 (27881)	75.70 (30241)	92.89 (13354)
<input type="checkbox"/>	United Kingdom	N	R32-	29	12.42	7.78 (4953) (27736)	44.95 (27736)	75.67 (30230)	92.88 (13353)
<input type="checkbox"/>	United Kingdom	Spike	G496S	29	11.20	7.18 (4570) (26744)	43.34 (26744)	74.07 (29593)	92.84 (13347)
<input type="checkbox"/>	United Kingdom	Spike	Q954H	29	12.54	7.82 (4980) (27885)	45.19 (27885)	75.70 (30241)	92.90 (13355)
<input type="checkbox"/>	United Kingdom	Spike	Y144-	29	12.30	7.81 (4973) (27604)	44.74 (27604)	74.92 (29930)	92.80 (13341)
<input type="checkbox"/>	United Kingdom	Spike	Q493R	29	11.24	7.20 (4582) (26790)	43.42 (26790)	74.08 (29596)	92.89 (13354)
<input type="checkbox"/>	United Kingdom	NSP6	I189V	29	12.61	7.88 (5016) (27911)	45.23 (27911)	75.70 (30244)	92.85 (13348)
<input type="checkbox"/>	United Kingdom	E	T9I	29	12.34	7.77 (4946) (27651)	44.81 (27651)	75.51 (30166)	92.89 (13354)
<input type="checkbox"/>	United Kingdom	NSP6	G107-	29	12.52	7.80 (4965) (27871)	45.17 (27871)	75.66 (30228)	92.88 (13352)
<input type="checkbox"/>	United Kingdom	Spike	N679K	29	12.54	7.82 (4980) (27878)	45.18 (27878)	75.68 (30235)	92.86 (13350)
<input type="checkbox"/>	United Kingdom	NSP3	K38R	29	12.52	7.83 (4984) (27839)	45.12 (27839)	75.62 (30211)	92.85 (13348)

Relevant Spike mutations in United Kingdom at the end of 2021



United Kingdom / BA.1 / 2021-12-31 / 4 weeks



Growing mutations in United Kingdom compared to Omicron (BA.1)



United Kingdom / BA.1 / 2021-12-31 / 4 weeks

▲

TABLE ⓘ

Location	Lineage	Protein	Mut	Slope	Y-intercept	P-value with mut	P-value without mut	P-value comparative ↑ 1
United Kingdom	BA.1	Spike	N764K	-1.9	55	0.0	0.0	0.0
United Kingdom	BA.1	Spike	R346K	2.6	6.3	0.0	0.0	1.2e-92
United Kingdom	BA.1	NS3	L106F	-1.7	18	0.0	0.0	2.4e-24
United Kingdom	BA.1	N	D343G	-1.6	18	0.0	0.0	5.3e-23
United Kingdom	BA.1	Spike	I1081V	-0.47	11	0.0	0.0	5.2e-12
United Kingdom	BA.1	NSP3	S1265N	-0.89	3.9	0.0	0.0	1.4e-7
United Kingdom	BA.1	NSP3	Y129H	0.32	2.0	0.0	0.0	0.0000011
United Kingdom	BA.1	NSP12	D153Y	0.29	1.6	0.0	0.0	0.0000028
United Kingdom	BA.1	Spike	A701V	-0.67	27	0.0	0.0	0.000019
United Kingdom	BA.1	NSP3	V1069I	-0.66	34	0.0	0.0	0.000043

P value with mut: shows if the population «lineage + mutation» is growing differently compared to everything.

P value without mut: shows if the population «lineage without mutation» is growing differently compared to everything.

P value comparative: shows if the population «lineage + mutation» is growing differently compared to the population «lineage without mutation».

Focus on: Spike mutations in United Kingdom compared to Omicron (BA.1)



United Kingdom / BA.1 / 2021-12-31 / 4 weeks

TABLE ⓘ

Location	Lineage	Protein	Mut	Slope	Y-intercept	P-value with mut	P-value without mut	P-value comparative ↑
United Kingdom	BA.1	Spike	N764K	-1.9	55	0.0	0.0	0.0
United Kingdom	BA.1	Spike	R346K	2.6	6.3	0.0	0.0	1.2e-92
United Kingdom	BA.1	Spike	I1081V	-0.47	11	0.0	0.0	5.2e-12
United Kingdom	BA.1	Spike	A701V	-0.67	27	0.0	0.0	0.000019

Rows per page: 10 ▾ 1-4 of 4 < >

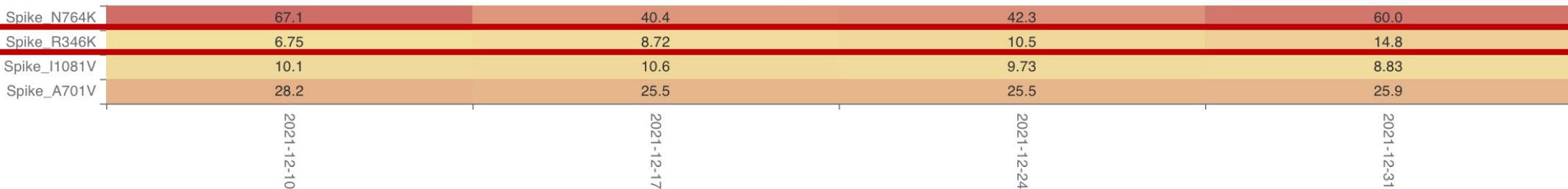
The observations of the third p-value and of the heatmap indicate that the Omicron lineage with the additional Spike mutation R346K is growing significantly faster than Omicron alone

Spike_R346K mutation is currently under scrutiny...

... but so far its addition to Omicron did not affect the neutralization susceptibility

Neutralization of SARS-CoV-2 Omicron variant by sera from BNT162b2 or Coronavac vaccine recipients

Lu Lu ¹, Bobo Wing-Yee Mok ¹, Lin-Lei Chen ¹, Jacky Man-Chun Chan ², Owen Tak-Yin Tsang ², Bosco Hoi-Shiu Lam ³, Vivien Wai-Man Chuang ⁴, Allen Wing-Ho Chu ¹, Wan-Mui Chan ¹, Jonathan Daniel Ip ¹, Brian Pui-Chun Chan ¹, Ruiqi Zhang ⁵, Cyril Chik-Yan Yip ^{1, 6}, Vincent Chi-Chung Cheng ^{1, 6}, Kwok-Hung Chan ¹, Dong-Yan Jin ⁷, Ivan Fan-Ngai Hung ⁵, Kwok-Yung Yuen ^{1, 6}, Honglin Chen ¹, Kelvin Kai-Wang To ^{1, 6}



Lineage recombination

Omicron as a result of other lineages' recombination



Hypothesis:

Omicron is a recombinant set of variants that have evolved over many months

Data extraction method:

Compute table of triplets of sequence groups
(identified by a set of amino acid changes)
that are «candidate recombinants»

Desirable properties of triplets:

- Minimal intersection
 - Cardinality of their union \sim cardinality of Omicron characterizing changes (61 changes)
 - Tentative: balanced cardinalities (i.e., three groups with 20/20/20 changes are preferable than 50/5/5)

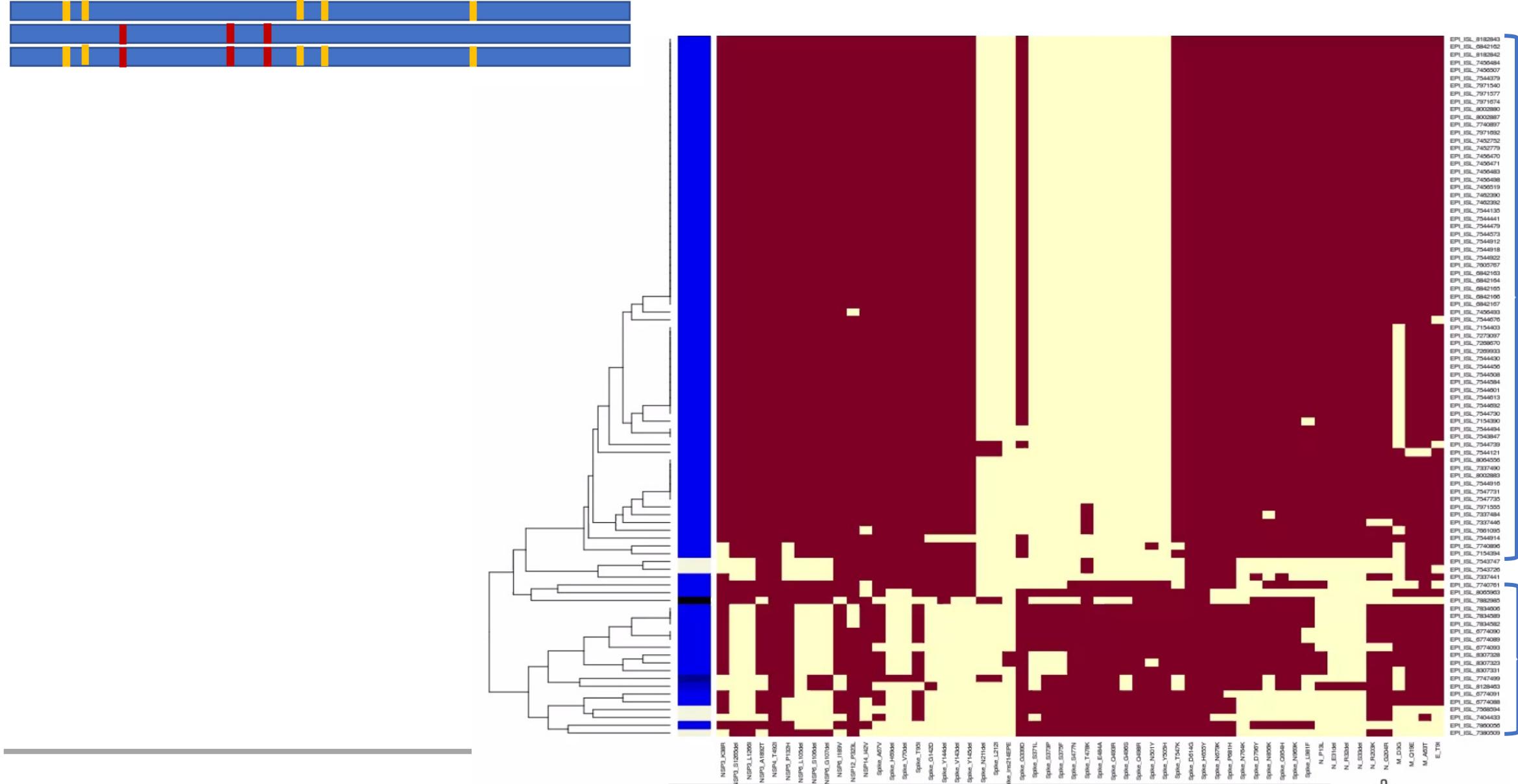
To be verified by means of:

- Phylogenetic analysis
 - Alignment checks
 - Location/Date coherency

ids_3	omicron_muts_3	num_omicro	date_3	location_3	lineage_3
	<ul style="list-style-type: none"> 'NSP14_I4ZV', 'Spike_T95I', 'M_A63T', 'Spike_N501Y', 'Spike_T478K', 'M_Q19E', 'Spike_G339D', 'NSP12_D222I' 				
ids_2	omicron_muts_2	num_omicron_muts_2	date_2	location_2	lineage_2
	<ul style="list-style-type: none"> 'NSP4_T492I' 				
ids_1	omicron_muts_1	num_omicron_muts_1	date_1	location_1	lineage_1
	<ul style="list-style-type: none"> 'NSP6_G107del', 'NSP12_P323L', 'Spike_V143del', 'Spike_G142del', 'NSP6_S106del', 'Spike_Y144del', 'Spike_H69del', 'Spike_D614G', 'Spike_P681H', 'Spike_S477N', 'EPI_ISL_3040131' 'Spike_V70del'] 				
				<ul style="list-style-type: none"> 'Republic of the Congo 	
			11	<ul style="list-style-type: none"> ['2021-07-07'] / Brazzaville] 	['B.1.620']
<ul style="list-style-type: none"> 'EPI_ISL_7661082', 'EPI_ISL_7661090' 	<ul style="list-style-type: none"> 'NSP3_S1265del', 'N_R203K', 		<ul style="list-style-type: none"> ['2021-11-28', 25 '2021-11-26'] 	<ul style="list-style-type: none"> ['Ghana / Accra', 'Ghana / Accra'] 	<ul style="list-style-type: none"> ['None', 'None']
<ul style="list-style-type: none"> 'EPI_ISL_7834586' 	<ul style="list-style-type: none"> 'Spike_P681H', 'Spike_H655Y', 	31	<ul style="list-style-type: none"> ['2021-12-06'] 	<ul style="list-style-type: none"> ['Botswana / Palapye'] 	<ul style="list-style-type: none"> ['BA.1']

Example three groups of sequences whose union reaches 57 Omicron characterizing mutations (their intersection has only 8)

Digging into results



The recombination has been tested on the «Alpha-Delta» case



[Comment on this paper](#)

Genome Recombination between Delta and Alpha Variants of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

Tsuyoshi Sekizuka, Kentaro Itokawa, Masumichi Saito, Michitsugu Shimatani, Shutoku Matsuyama, Hideki Hasegawa, Tomoya Saito, Makoto Kuroda, COVID-19 Genomic Surveillance Network in Japan (COG-JP)

doi: <https://doi.org/10.1101/2021.10.11.21264606>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract Full Text Info/History Metrics [Preview PDF](#)

Abstract

Prominent genomic recombination has been observed between the Delta and Alpha variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolated from clinical specimens in Japan. It is necessary to intensively study such marked genetic variations and characterize the emerging variants after careful verification of their lineage and clade assignment.

Target mutations are 34:

M_I82T, NS3_S26L, NS8_Q27stop, NS8_R52I, NS8_Y73C, NSP12_E254D, NSP12_G671S, NSP12_P323L, NSP13_P77L, NSP14_A394V, NSP3_A488S, NSP3_P1228L, NSP3_P1469S, NSP3_V932A, NSP4_T492I, NSP4_V167L, NSP6_T77A, N_D3L, N_G204R, N_R203K, N_S235F, Spike_D614G, Spike_D950N, Spike_E156G, Spike_F157del, Spike_G142D, Spike_L452R, Spike_L585F, Spike_P681R, Spike_Q173H, Spike_R158del, Spike_T19R, Spike_T478K, Spike_T95I

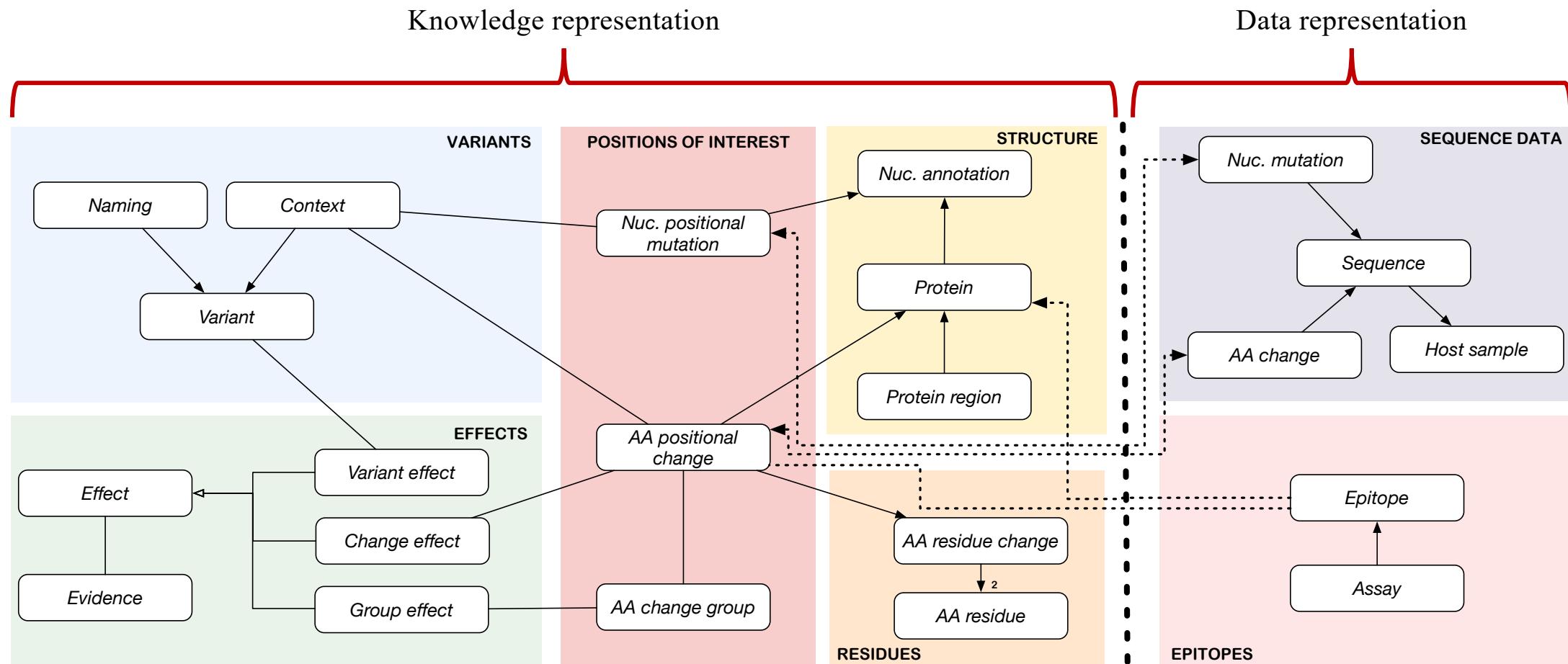
In the following 6 sequences reported by the article of COVID-19 Genomic Surveillance Network in Japan (COG-JP):
EPI_ISL_4882472, EPI_ISL_4882473, EPI_ISL_4882474, EPI_ISL_4882475, EPI_ISL_4882476, EPI_ISL_4882477

We were able to identify two groups of Japanese sequences that are candidate recombinants for these 6 target sequences.

A knowledge base

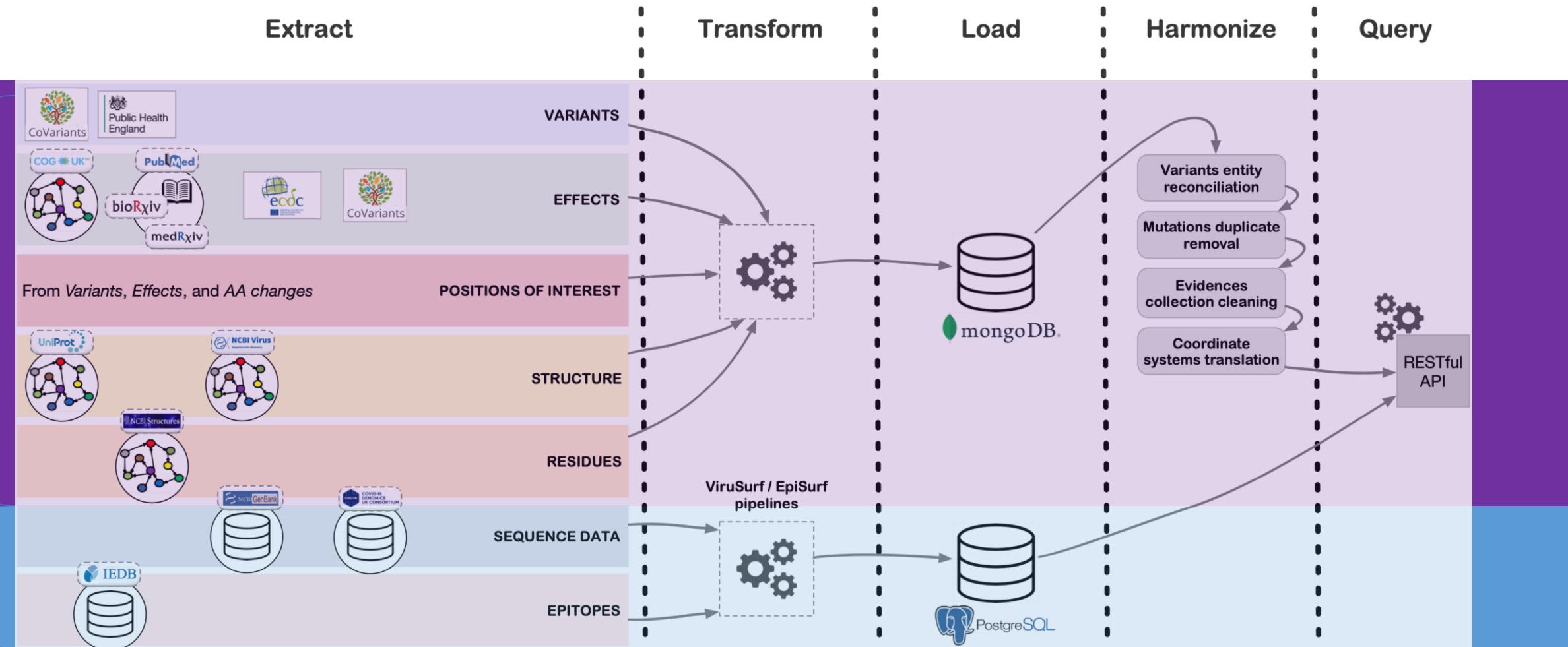
for mutation and variants effects

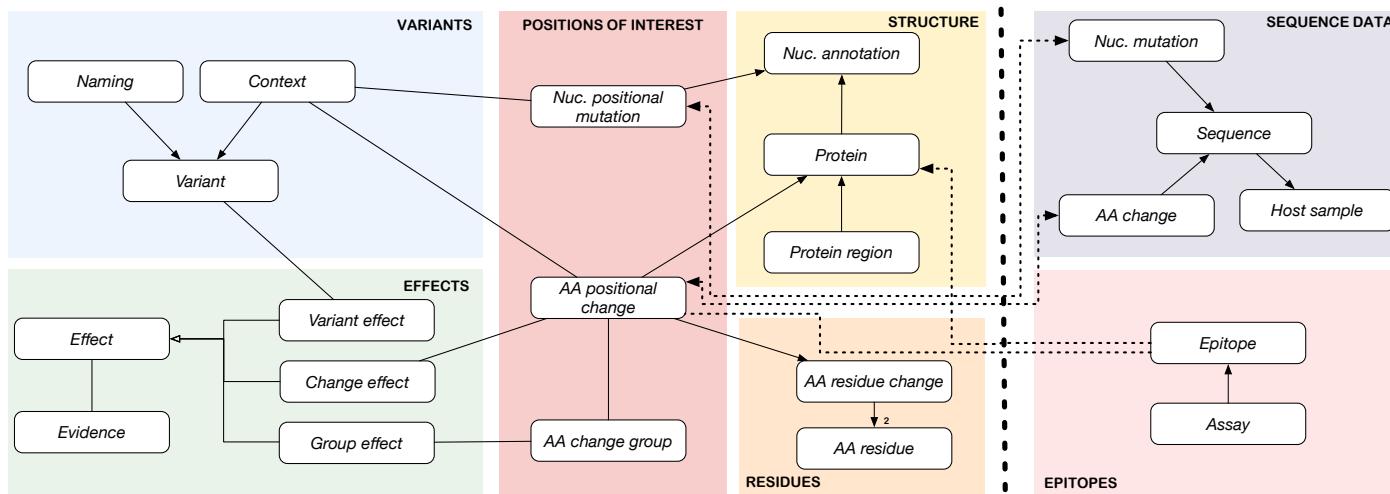
CoV2K abstract model



Alfonsi, T., Al Khalaf, R., Ceri, S., & Bernasconi, A. CoV2K model, a comprehensive representation of SARS-CoV-2 knowledge and data interplay. Submitted to Scientific Data Journal – 2-year IF 2019: 5.5, SJR: Q1

CoV2K knowledge and data integration pipeline





One endpoint for each entity of CoV2K:

- Without parameters;
- With a *same entity identifier* as a path parameter (returning only one instance);
- With an *attribute-value pair* as a query parameter (filter on the entity);
- With *another entity identifier* query parameter (returning the set of instances of the first entity that are linked to the instances of the identified second entity)

Possibility to traverse paths of the graph
(with the «combine» endpoints)

Example queries:

- What are the characteristics (Grantham distance and type) of the residue changes of the Alpha variant?
- Which amino acid changes of VOC-20DEC-02 fall within the Receptor Binding Domain (RBD)?
- Which are the effects of the variants that include the Spike amino acid change D614G?

Collection of effects of Omicron mutations



Mutation: S:Y145D

LOWER sensitivity to neutralizing mabs, ['Haslwanter et al, <https://doi.org/10.1101/2021.06.10.447999>']

Mutation: S:K417N

LOWER sensitivity to vaccine sera, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

LOWER sensitivity to vaccine sera, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>']

LOWER sensitivity to vaccine sera, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Chen et al.(2021), <https://doi.org/10.1038/s41591-021-01294-w>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Chen et al.(2021), <https://doi.org/10.1038/s41591-021-01294-w>']

HIGHER binding to host receptor, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER protein stability, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER protein flexibility, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER binding to abs, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

UNKNOWN binding to host receptor, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

Mutation: S:N440K

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Rappazzo et al. (2021), <https://doi.org/10.1126/science.abf4830>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Rappazzo et al. (2021), <https://doi.org/10.1126/science.abf4830>', 'Weisblum et al. (2020), <https://doi.org/10.7554/eLife.61312>']

LOWER protein stability, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER protein flexibility, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER binding to abs, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

Mutation: S:G446S

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

Mutation: S:S477N

LOWER sensitivity to neutralizing mabs, ['Liu et al. (2020), <https://doi.org/10.1016/j.chom.2021.01.014>']

HIGHER binding to host receptor, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER infectivity, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER viral transmission, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER protein stability, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

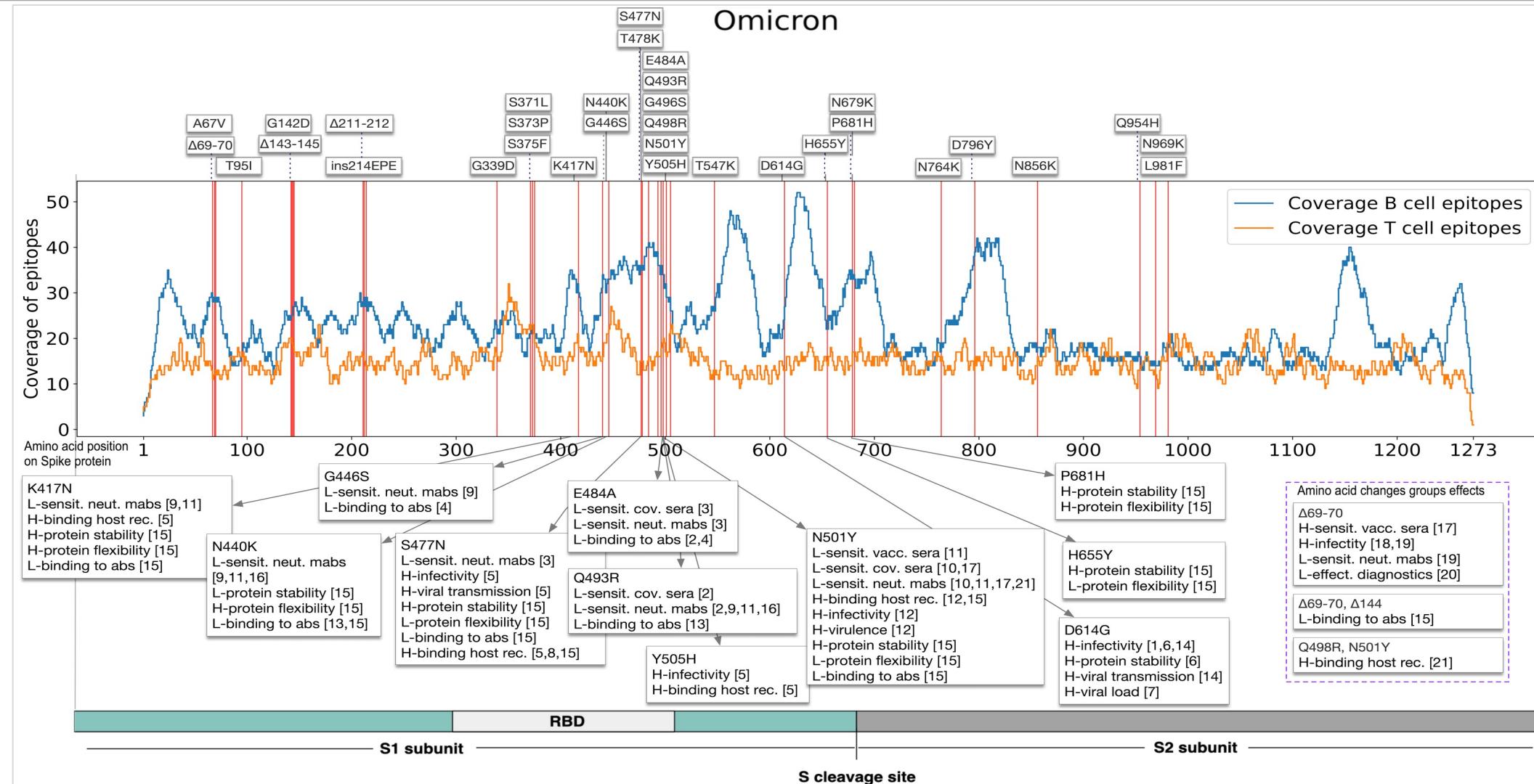
LOWER protein flexibility, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER binding to abs, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER binding to host receptor, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER binding to host receptor, ['Singh et al. (2021), <https://doi.org/10.1038/s41598-021-83761-5>']

Single and group mutations effects (according to CoV2K)



On Nov 30th, we made the 1st post on Omicron on the Virological.org web blog, reaching >14k views:

<https://virological.org/t/report-on-omicron-spike-mutations-on-epitopes-and-immunological-epidemiological-kinetics-effects-from-literature/770>