GRAPHEN
An Artificial Intelligence Company

EECS 6893 Big Data Analytics Lecture 13

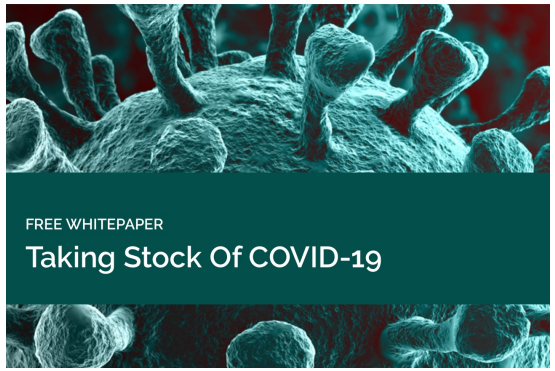
Big Data and AI for Medical Applications

Prof. Ching-Yung Lin
Depts. of EECS, Columbia University

12/08/2023

AI for New Drug Design

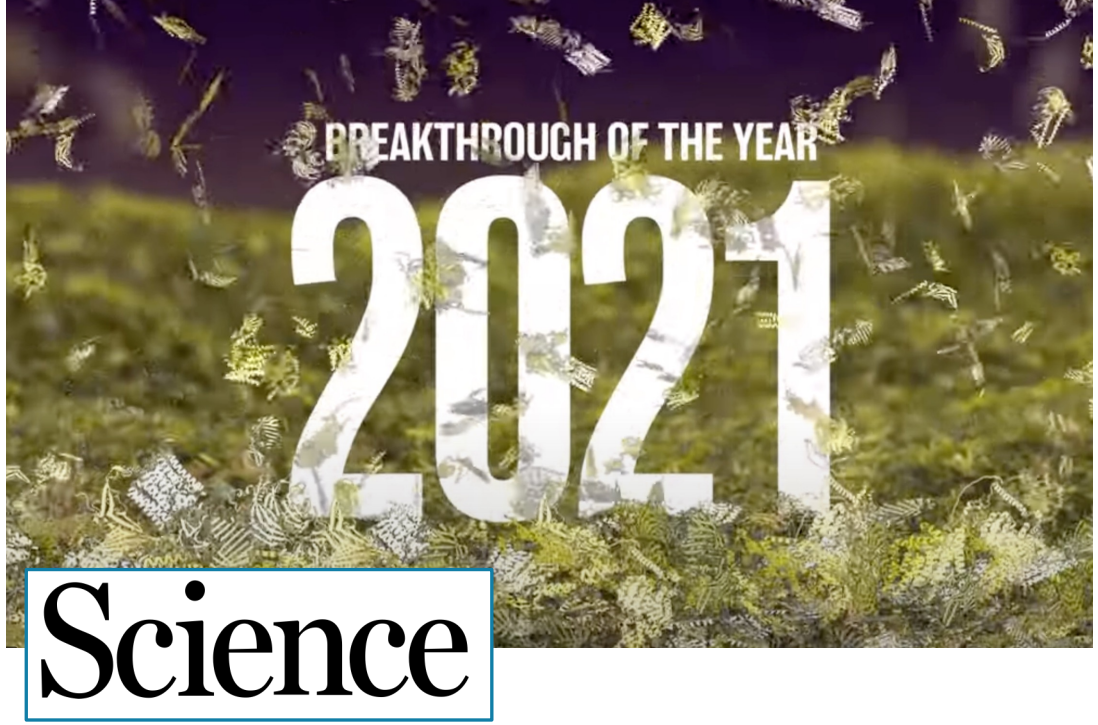
ABIresearch
for visionaries

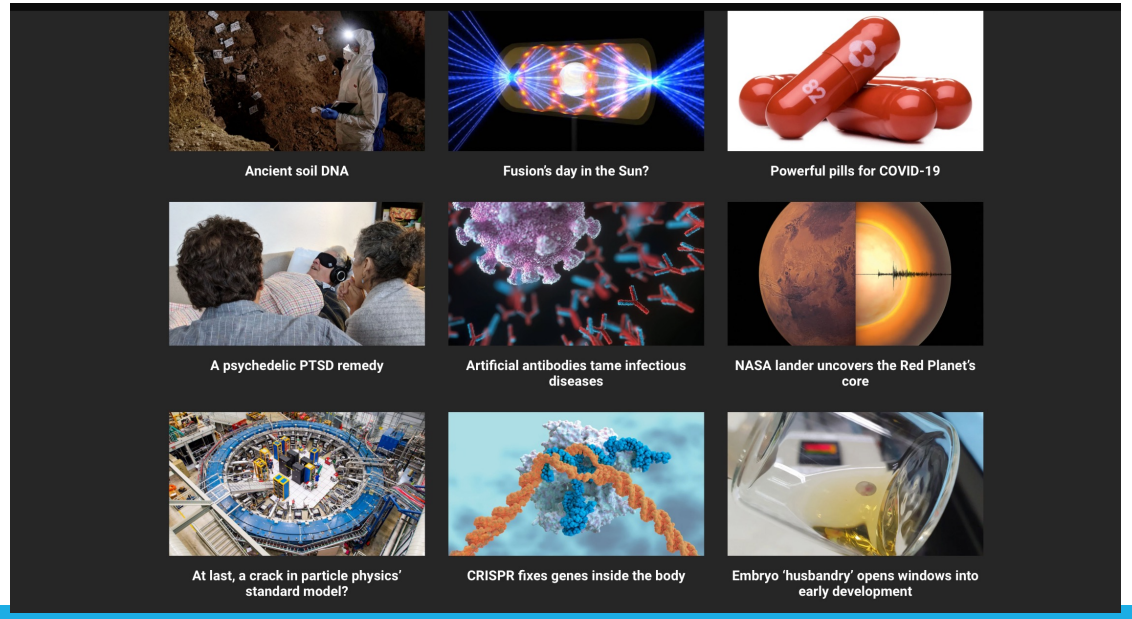


GRAPHEN
An Artificial Intelligence Company

*"Tools from established companies like **Google** DeepMind, startups like **Graphen**, and AI chipsets from vendors like **NVIDIA** and **Intel** will help accelerate the speed of drug discovery, development, and testing, allowing pharmaceutical companies and healthcare authorities to combat the pandemic." – ABI research, May 2020*

Protein == Biological Machine





I never thought I'd see this in my lifetime.

JOHN MOULT | UNIVERSITY OF MARYLAND, SHADY GROVE

https://youtu.be/iUMpm3tYsVE?t=218

Graphen Medical

GRAPHEN
An Artificial Intelligence Company

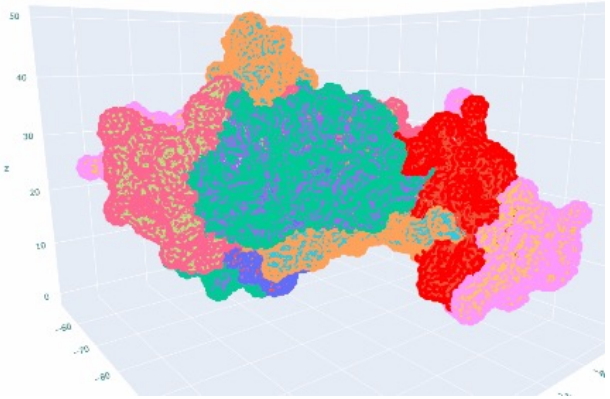
Each Person knows Biological Digitized-Self, with Personalized Precision Treatment and Drugs

-- Graphen's Vision to Help Making Future of Medicine

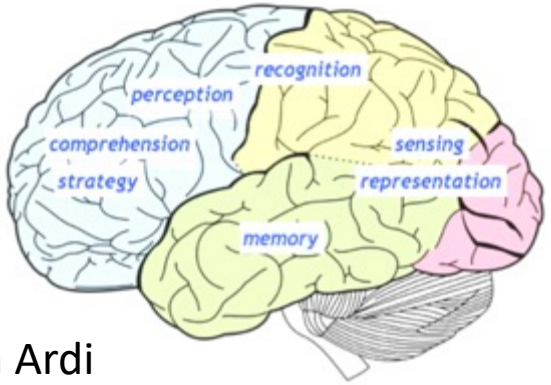
Graphen Drugomics

Graphen Genomics

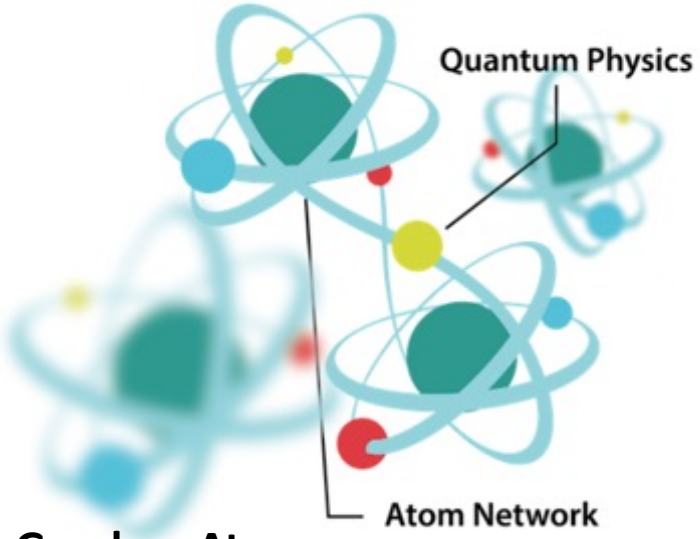
Life is composed of graphs of atoms



https://www.graphen.ai/products/medical.html

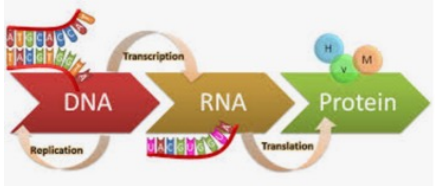


Graphen Ardi

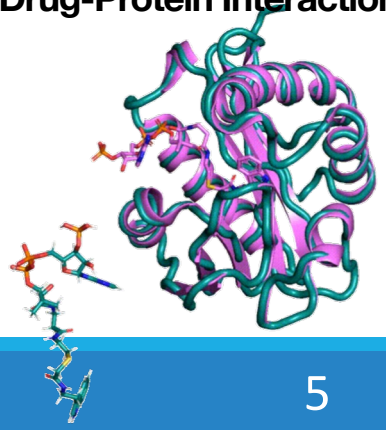


Graphen Atom

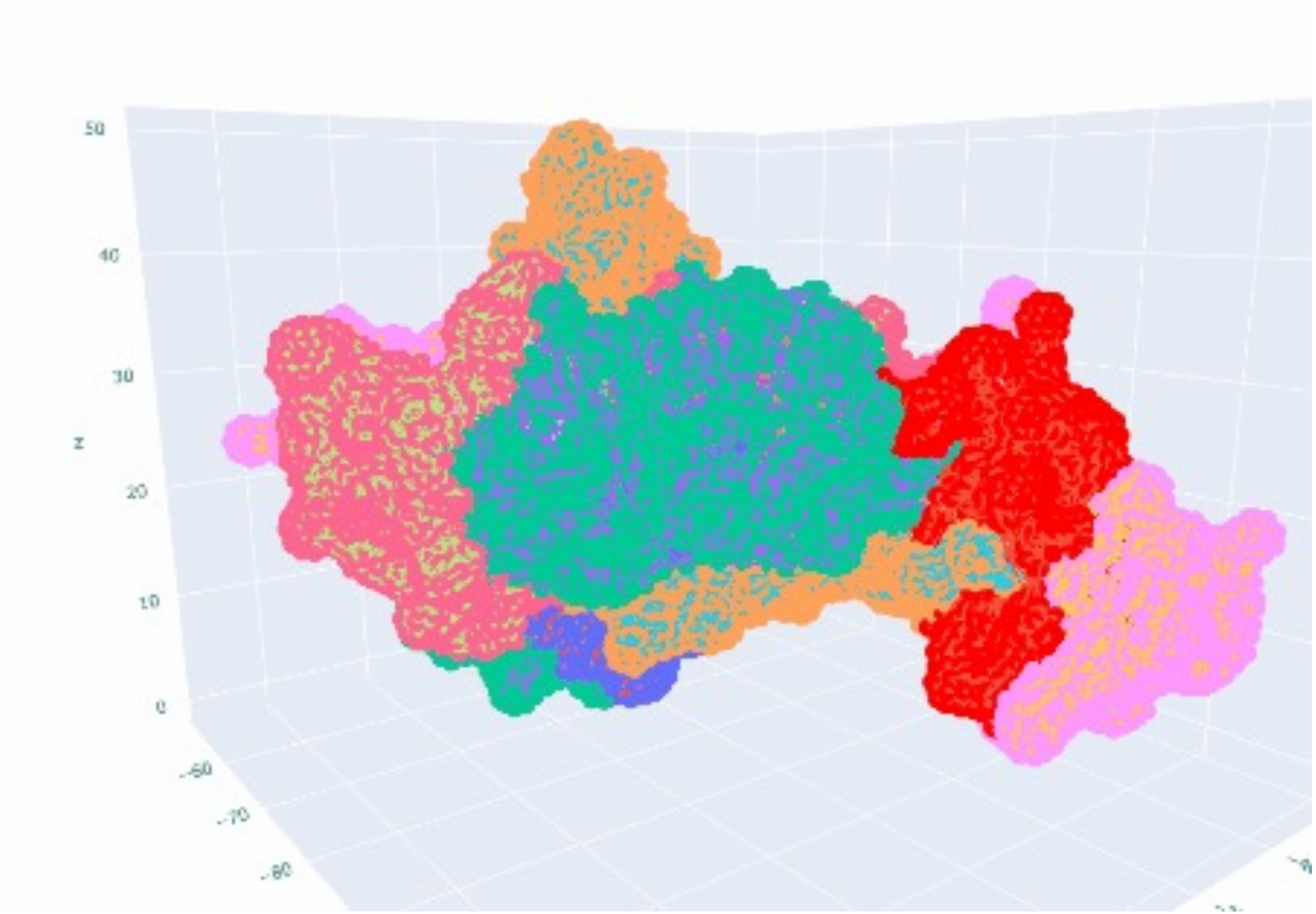
Central Dogma of Biology



Drug-Protein Interaction



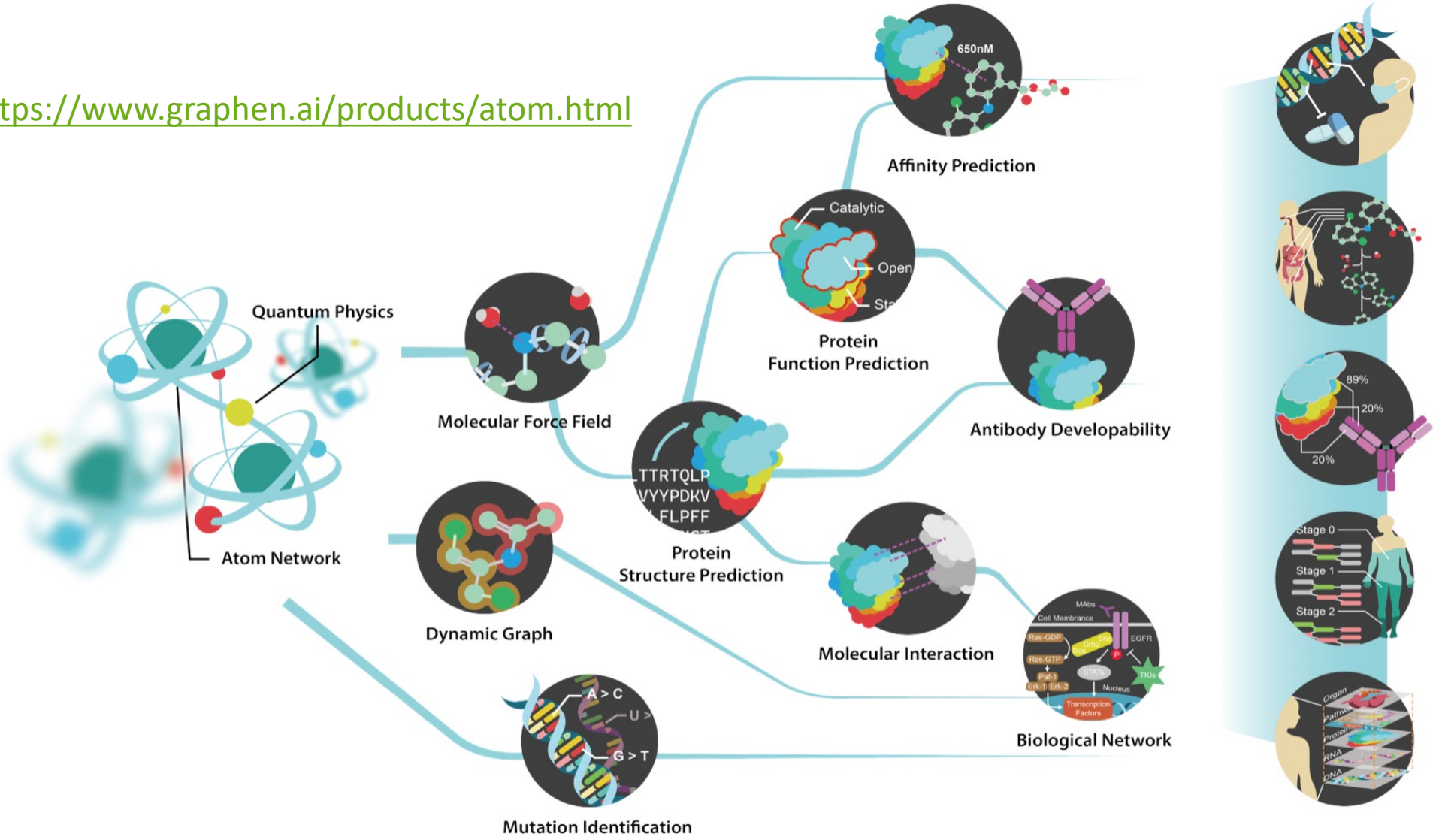
Life is composed of graphs of atoms



Graphen Atom as a Foundation for Comprehensive Drug Development and Precision Medicine

GRAPHEN
An Artificial Intelligence Company

https://www.graphen.ai/products/atom.html



Mutation Intelligence

- Disease Risk Evaulation
- Drug Resistance
- Genome Variant Reasoning

Drug Development

- New Drug Generation
- Drug Synthesis Reaction Simulation
- ADME Prediction

Immunal formation

- Antibody-Antigen contact Prediction
- Antibody-Antigen Affinity Prediction

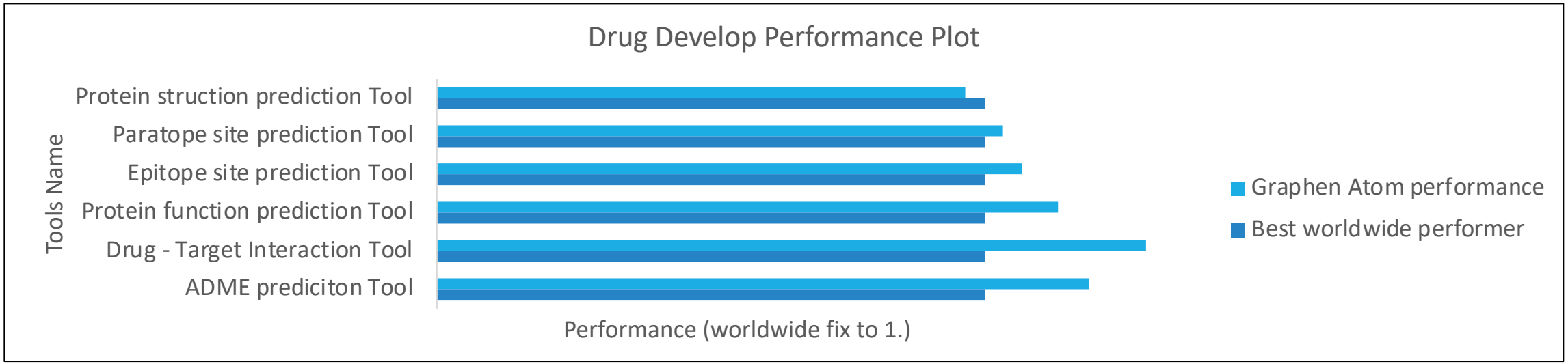
Progress Prediction

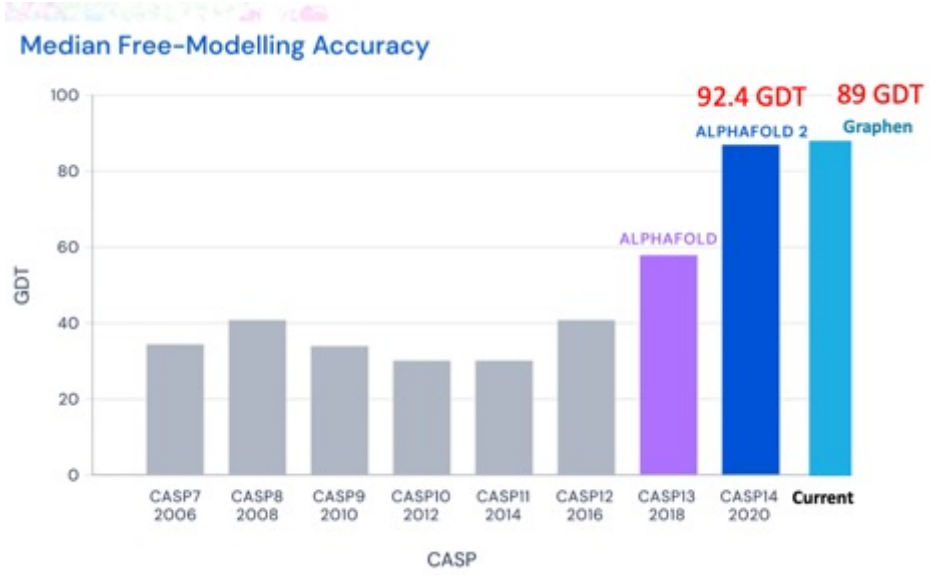
- Disease Progression
- Race Differentiation

Multi-Omics Analysis

- Multi-Omics Network
- Single-Cell Analysis

Graphen Atom: Best Performance on most key tools for AI Drug Development





- Graphen Atom uses **much fewer computational hardware** with similar performance to Alphafold2: ~1/50
- Graphen Atom has **much more tools for drug development** than Alphafold2, which is limited to Protein Structure Prediction.

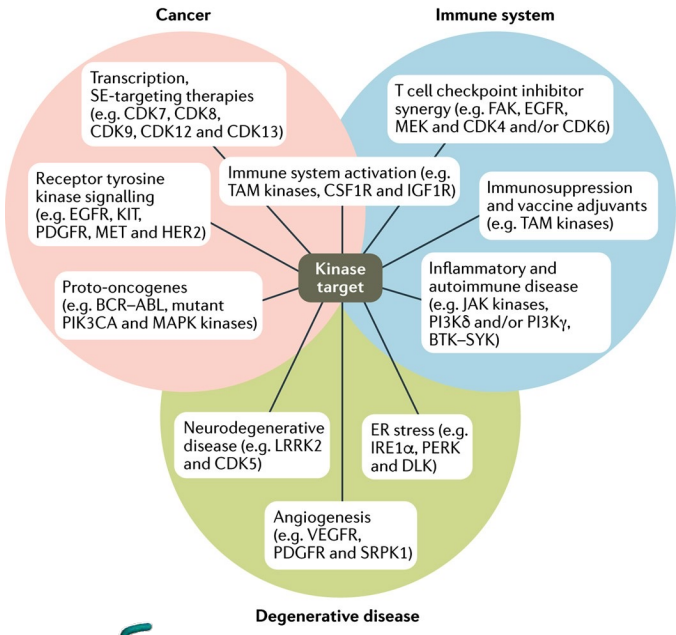
Graphen Small Mole Drug Dev ➔ 1/27 of the Time; 1/9000 of the Cost, comparing to traditional methods

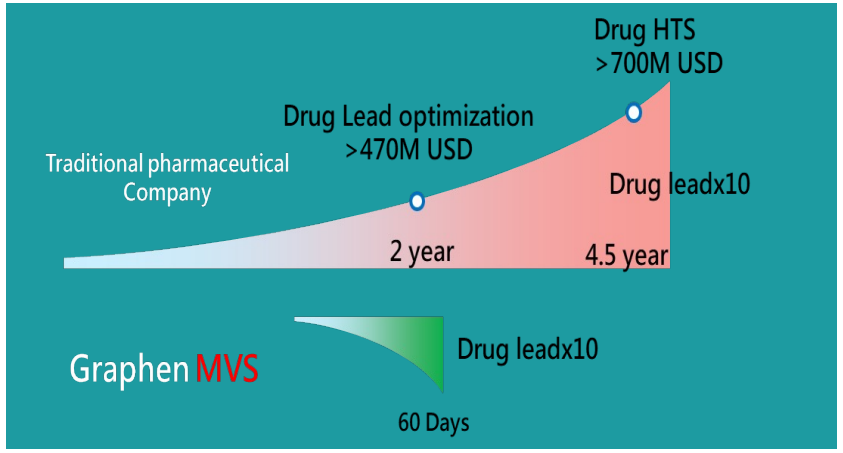
GRAPHEN
An Artificial Intelligence Company

1. Generate Synthesis and low side effects Drug
2. Filtering drug by ADMET and Solubility
3. Filtering Drug by Graphen QF Energy model and Kinase model
4. Disease-Clinical/Cell Target mapping (Spectrum Mutant Prediction)
5. Synthesis and Wet Lab Evaluation

	Samples	Required Time
Generate de novo new compound	30K	7 Days
Manufacturable and Safety Screening	2.5K	3 Days
Dynamic Target-Lead interaction simulation	0.1K	5 Days
Target Cell and Disease Finding	30	5 Days
	10	30+30 Days Synthesis, Kinase Assay, and Cell Assay

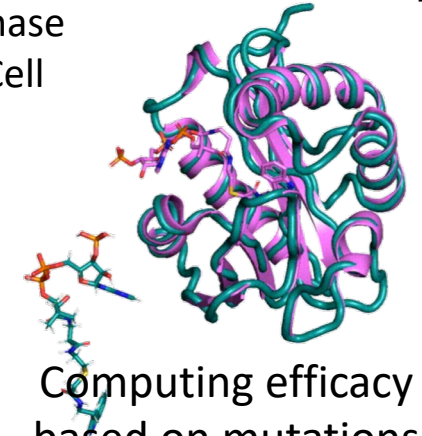
Targeting on Kinase Drugs



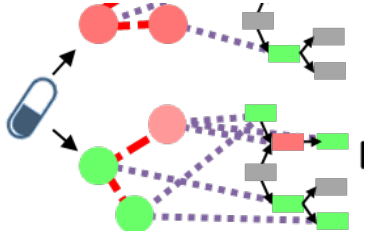


Graphen's Cost to create a Drug Candidate (cell proven) : ~USD $15K

Graphen is creating 2-3 effective drug candidates per month



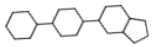
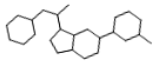
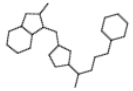
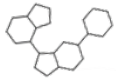
Computing efficacy based on mutations



Kinase functional grouping

Graphen Drug Development Timeline (~20 drugs)

GRAPHEN
An Artificial Intelligence Company

Product	Indication	Drug Discovery	Target & Lead Validation	Proof of Concept*	IND-Enabling
GD-2579 	Fatty Liver, NASH	Jun. 2023	Aug. 2023	Jul. 2024	Q1 2025
	Bladder Cancer	Jun. 2023	Sep. 2023	Jul. 2024	
GD-0414 	Pancreatic Cancer, PAAD	Aug. 2023	Dec. 2023	Apr. 2024	Q4 2024
	Breast Cancer, TNBC	Aug. 2023	Dec. 2023		
GD-1291 	Gastrointestinal Cancer	Jun. 2023	Sep. 2023	Jul. 2024	Q1 2025
GD-1080 	Colorectal Cancer	Aug. 2023	Dec. 2023	Apr. 2024	Q4 2024
16 leads**	Cardiovascular: 1 Metabolism: 2 Cancers: 8 Central Neurological: 5 Auto-Immune: 1	Dec. 2023	Mar. 2024	Apr. 2025	Q4 2025

*1: All structures from the model will be pushed to POC
*2: Leads ready to validate indication

Graphen's upcoming 16 more drugs

GRAPHEN
An Artificial Intelligence Company

16 leads

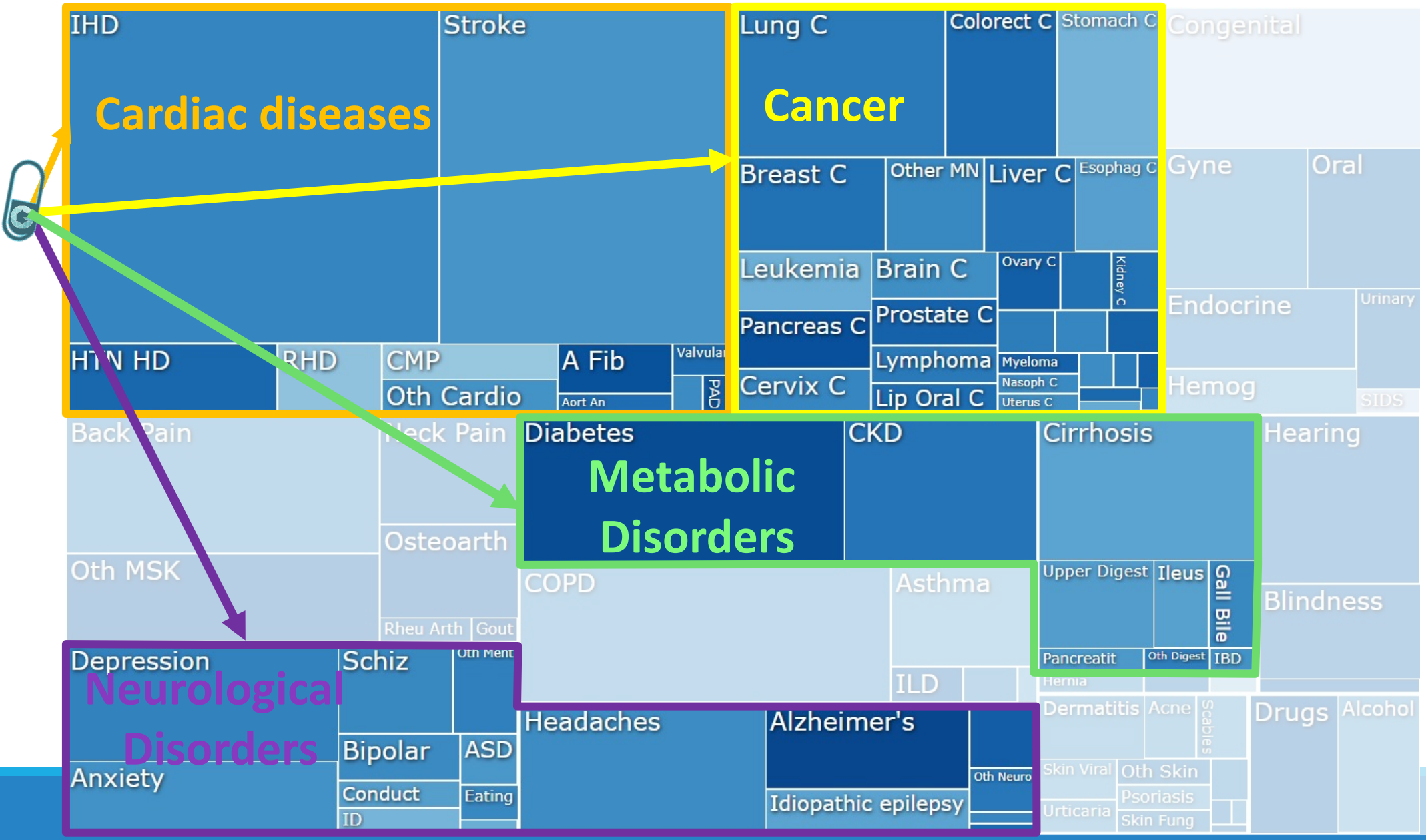
Cardiovascular: 1
Metabolism: 2
Cancers: 8
Central Neurological: 5
Auto-Immune: 1

Potential Indications	Number of Candidate Leads*	Current First-line drugs	Under-Clinical drugs
Coronary Artery Disease (CAD)	1	Propranolol, Beta blockers	None
Non-alcoholic Fatty Liver Disease (NAFLD)	2	None	Resmetirom Phase III
Obesity	3	Semaglutide	Mazdutide Phase III
Rheumatoid Arthritis (RA)	1	Sulfasalazine, Hydroxychloroquine	Peresolimab Phase II
Depressive Disorder	1	SSRIs	Navacaprant Phase III
Schizophrenia	1	SGAs	Brilaroxazine Phase III
Hepatocellular carcinoma	2	Sorafenib, Tislelizumab	Lenvatinib Phase III
Prostate cancer	1	Talazoparib + Enzalutamide	177Lu-PSMA-617 Phase II

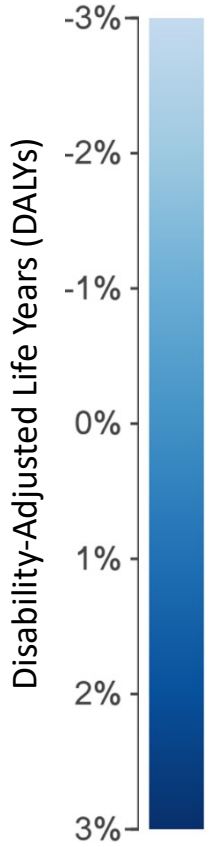
- Cardiovascular
- Cancers
- Central Neurological
- Metabolic
- Immune & Inflammation

Graphen's drugs are covering around 70% of disease fields

GRAPHEN
An Artificial Intelligence Company



Annual % change
(2010 to 2019)
DALYs/100,000

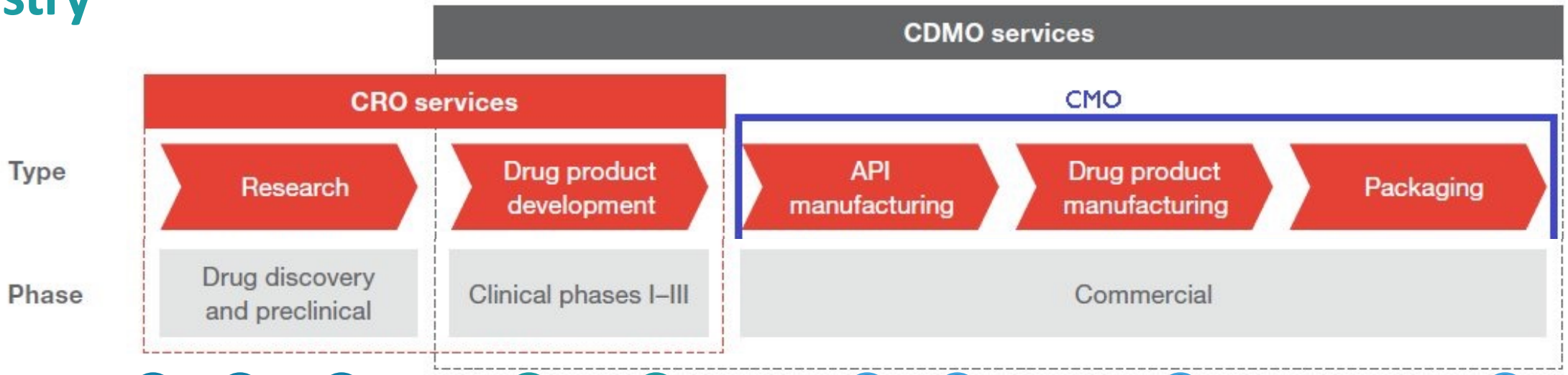


Data from IHME
(1990 to 2019)

IHME
THE INSTITUTE FOR HEALTH METRICS AND EVALUATION

AI drives Future Drug Design, Development and Manufacturing Industry

GRAPHEN
An Artificial Intelligence Company



Drug

- Therapeutic Biomarker & Druggable Candidate
- De novo Drug Design Drug Repurposing
- Drug ADMET Drug Developability
- Drug Mechanism Analysis
- Biomarker-based patient subgroups screening
- Retrosynthesis & Optimize manufacturing
- Drug Synergy
- Clinical use suggestion
- Companion Diagnostic IVD Development

Vaccine:

- Targeting Domain
- Immunogenicity
- Generalization
- Percentage yield
- Conformational Stability

AI is Solving Big Challenges of Traditional Pharmaceutical Drug Discovery

GRAPHEN
An Artificial Intelligence Company

AI is making a paradigm shift to reduce the risk of drug development via:

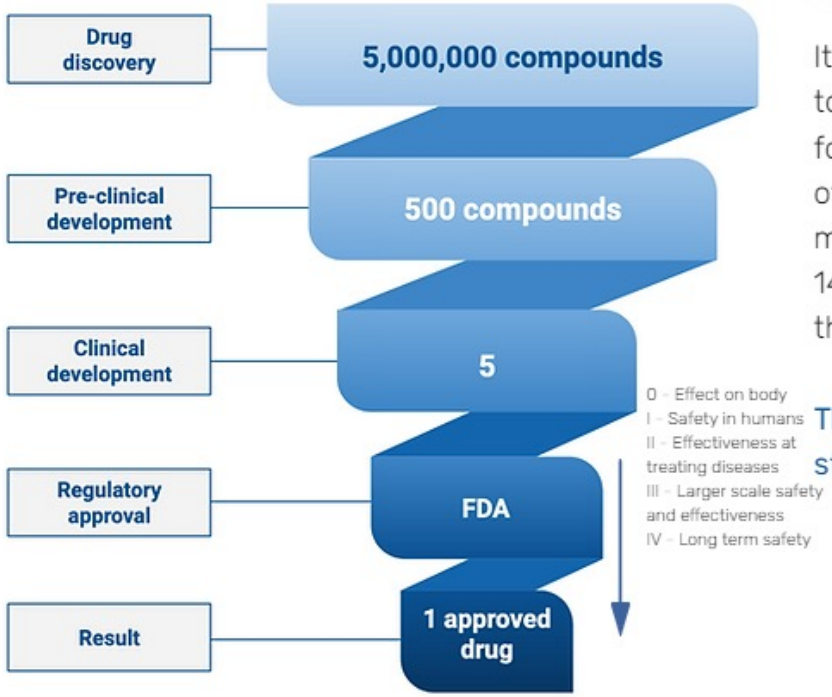
- **Time reduction (Graphen 1/27 vs other AI companies 1/2)**
- **Enrich pipeline**
- **Increase Probability of Success**
- **Cost Reduction (Graphen 1/9000 vs other AI companies 1/3)**
- **More Precise, Fewer Side Effects**
- **Accelerating innovation**

With the acceleration, drug discovery and development of (pharmacological) therapies of the above become possible:

- **Rare diseases or new antibiotics.**
- **Previously undruggable targets** due to protein structural properties**.**
- **Predict clinical trial results** to address the complex issues of clinical trial failure.

Pharma Efficiency: Challenges

10 years + $2.6 bln = 1 new drug

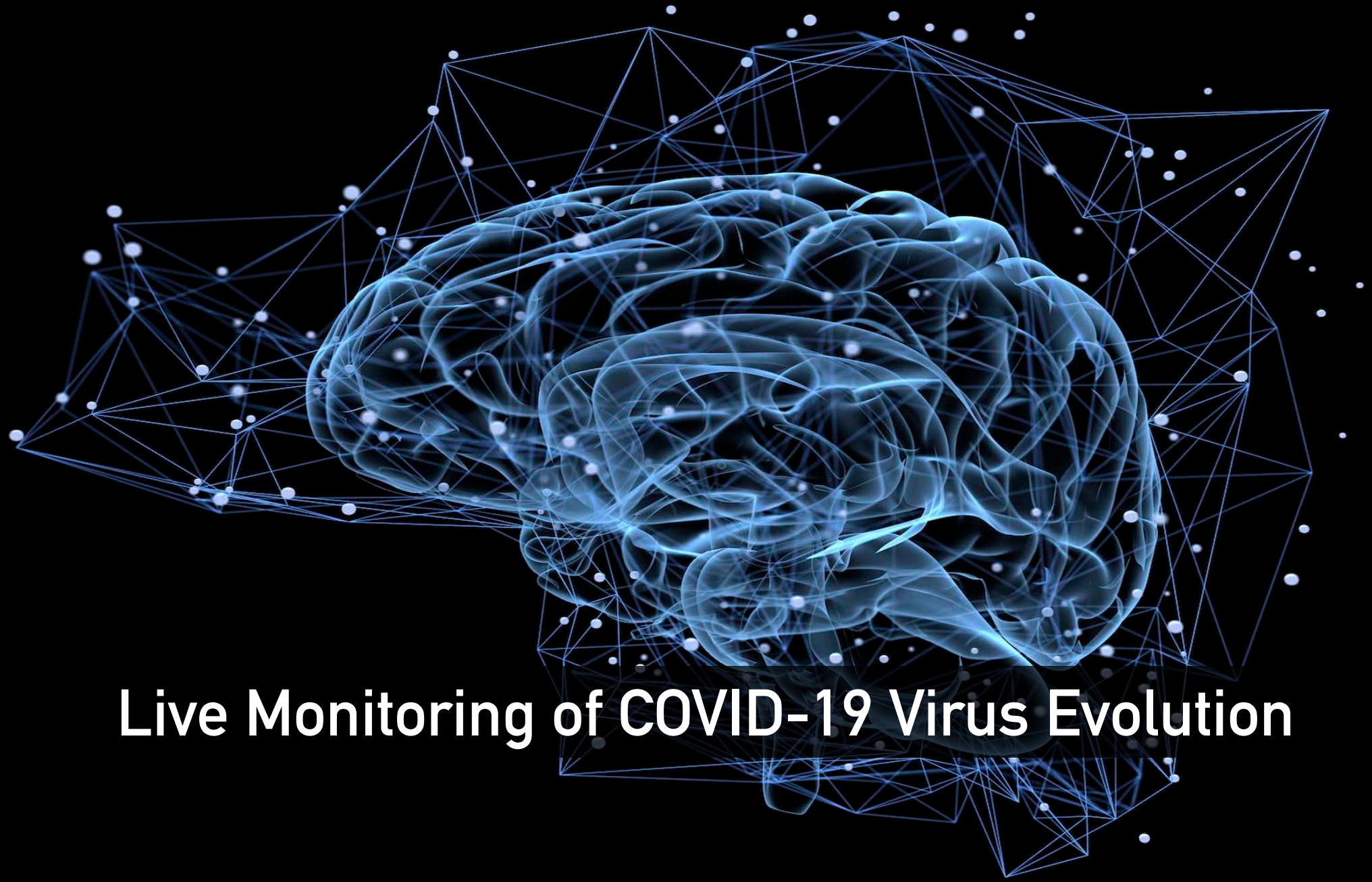


It takes, on average, over 10 years to bring a new drug to market. As of 2014, according to the Tufts Center for the Study of Drug Development (CSDD), the cost of developing a new prescription drug that gains market approval is approximately $2.6 billion. This is a 145% increase, correcting for inflation, compared to the same report made in 2003.

The solution to this problem comes from three key strategies:

- evolution of business models towards more collaboration and early pipeline diversification
- implementation of AI as a universal shift towards data-centric drug discovery
- discovery of new therapeutic modalities (biologics, therapies, etc.)

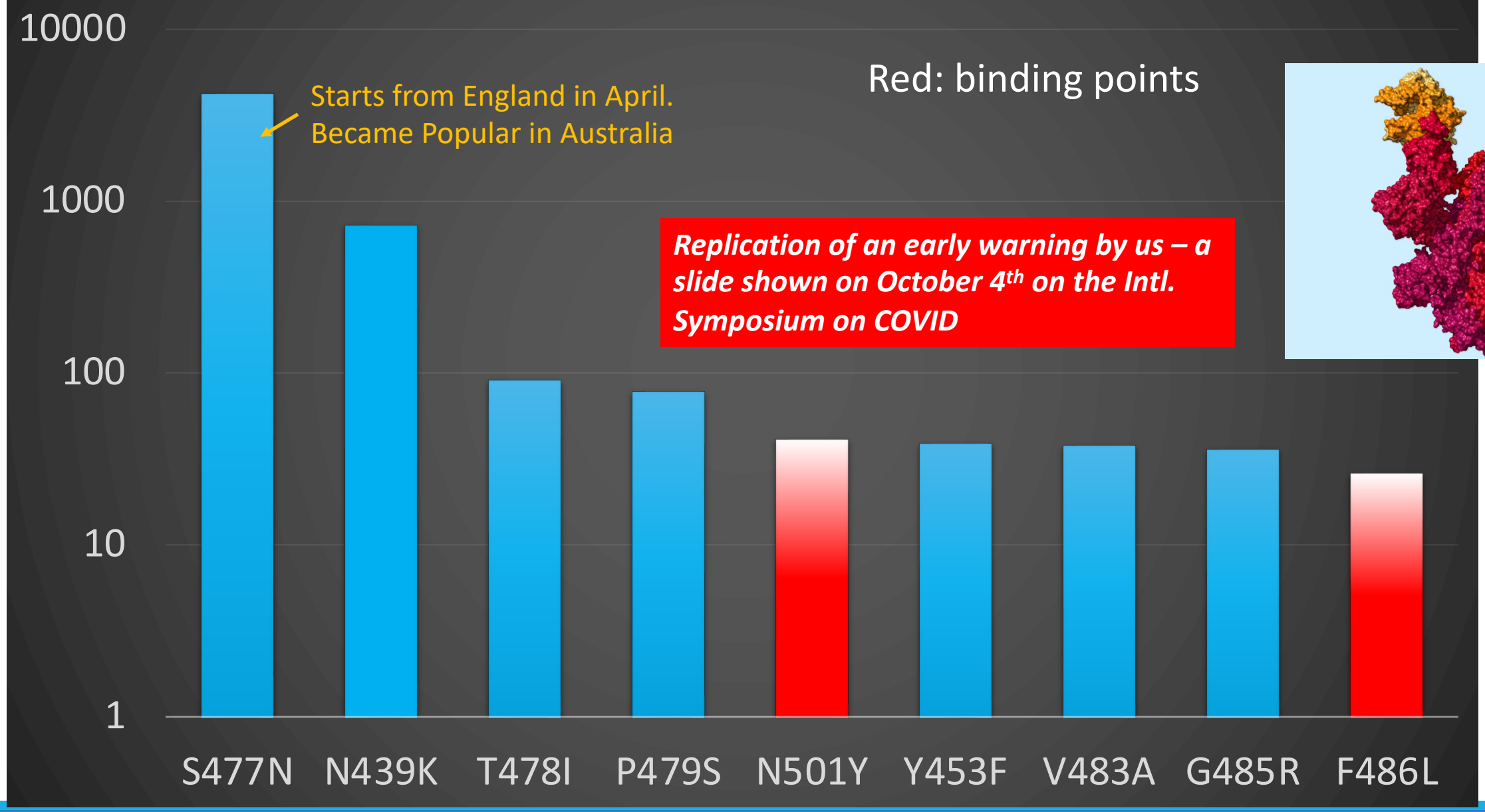
Ref: https://www.deep-pharma.tech/ai-in-dd-q1-2023-subscribe



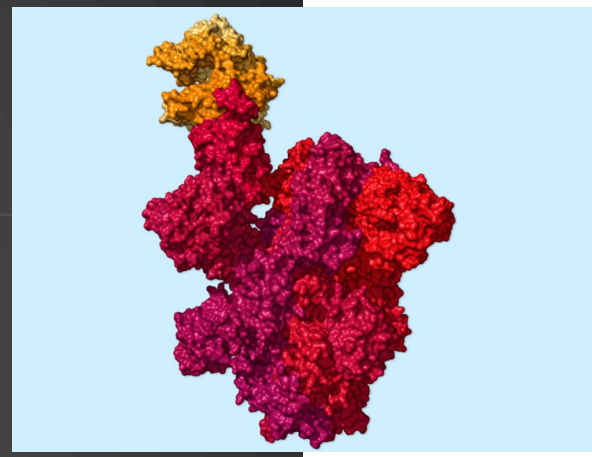
Live Monitoring of COVID-19 Virus Evolution

of Mutated Strains in RBD area

GRAPHEN
An Artificial Intelligence Company



Red: binding points

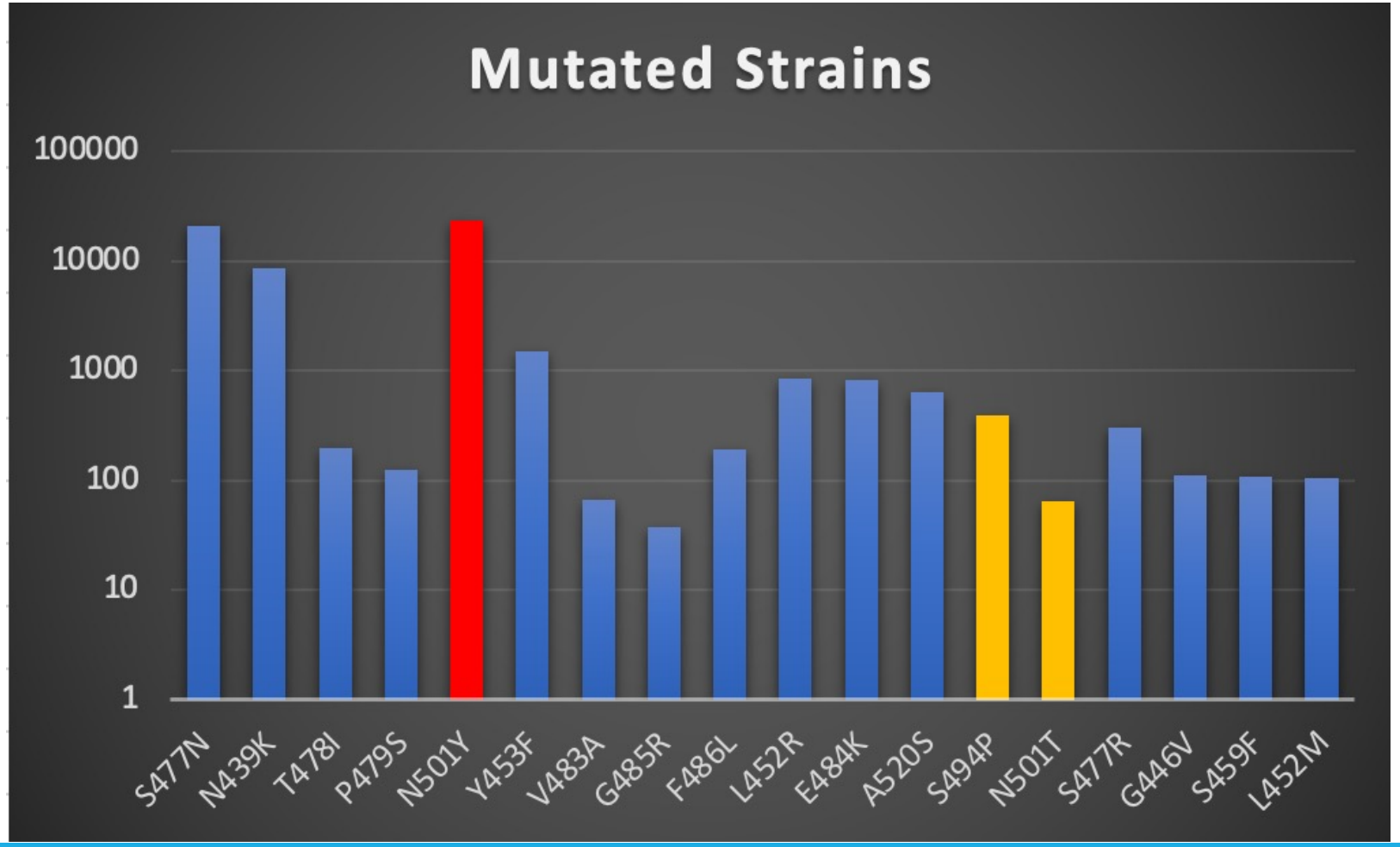


Replication of an early warning by us – a slide shown on October 4th on the Intl. Symposium on COVID

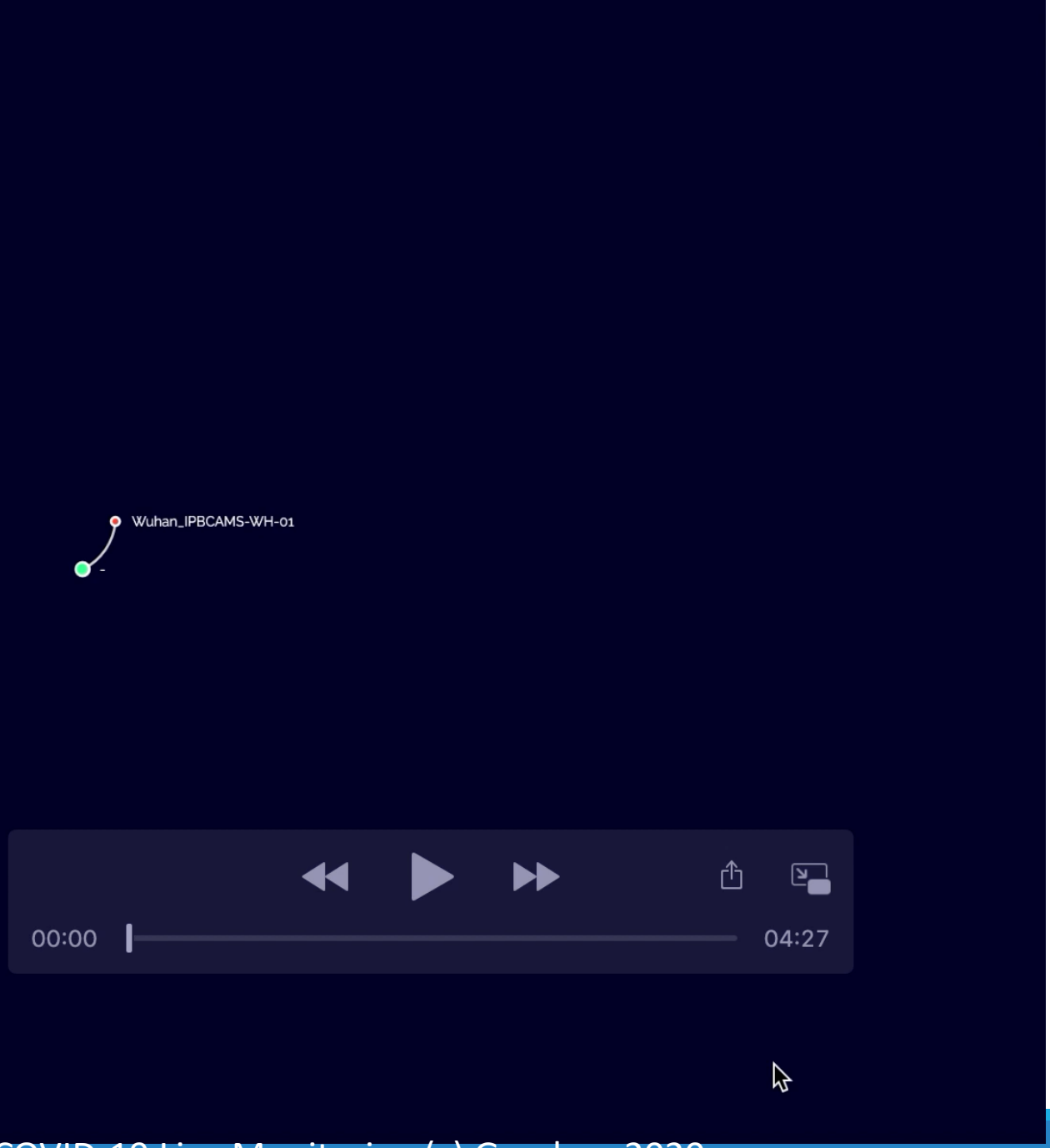
Summary

GRAPHEN
An Artificial Intelligence Company

(January 25, 2021) Monitored 395,018 sequenced SARS-CoV-2 viruses



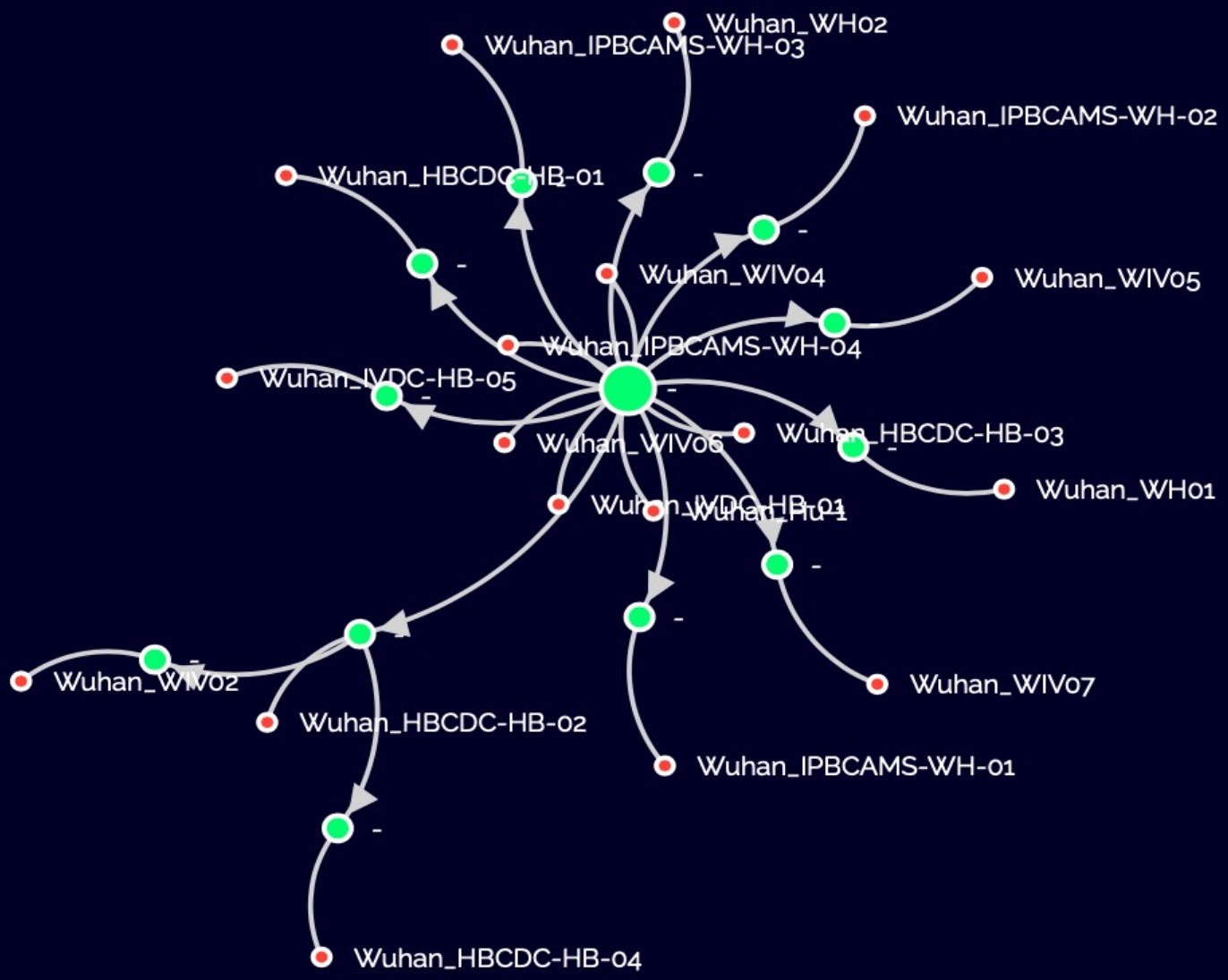
GRAPHEN
An Artificial Intelligence Company



Virus #: 18

Set #: 13

12-31-2019

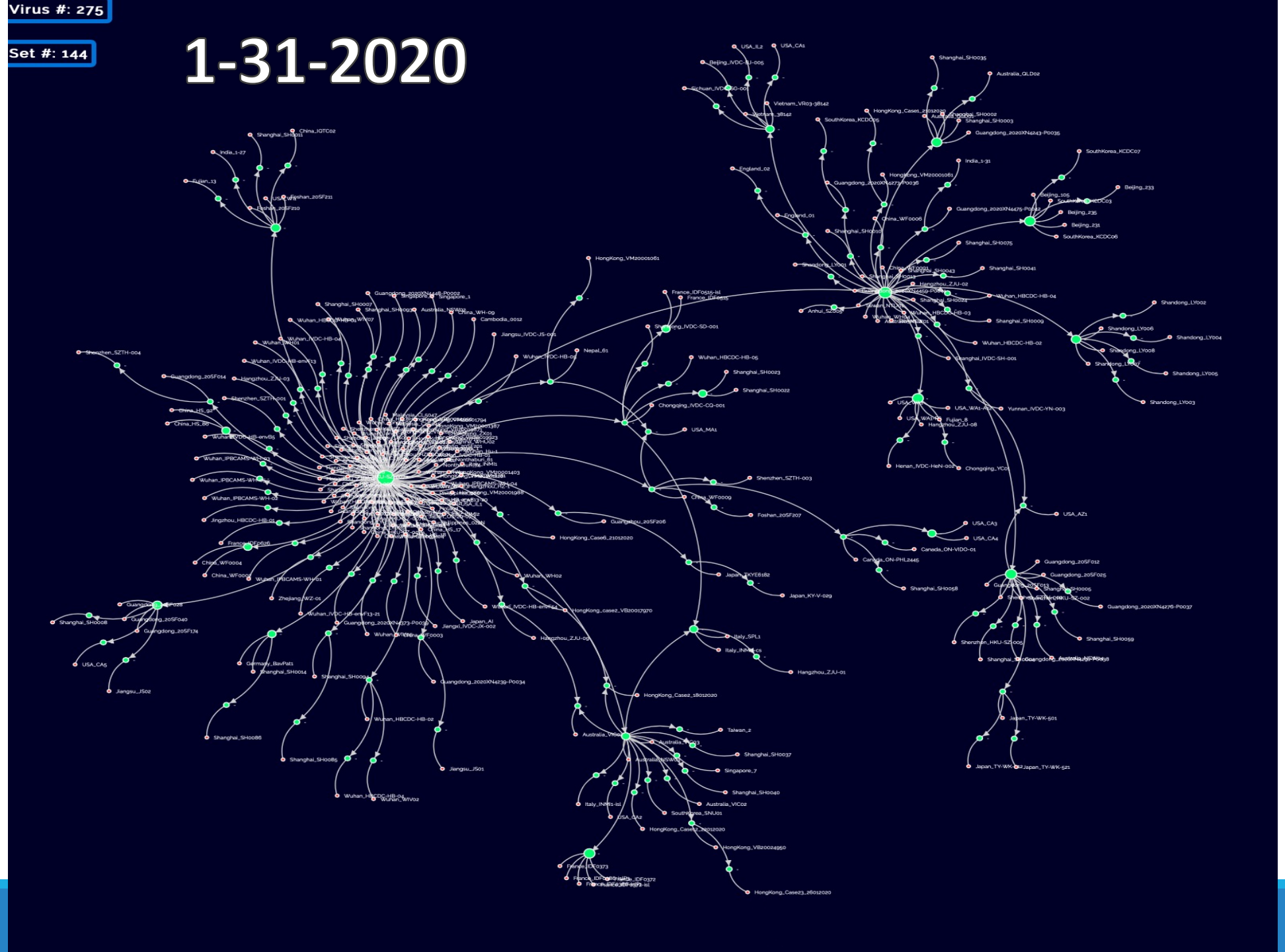


GRAPHEN
An Artificial Intelligence Company

Virus #: 275

Set #: 144

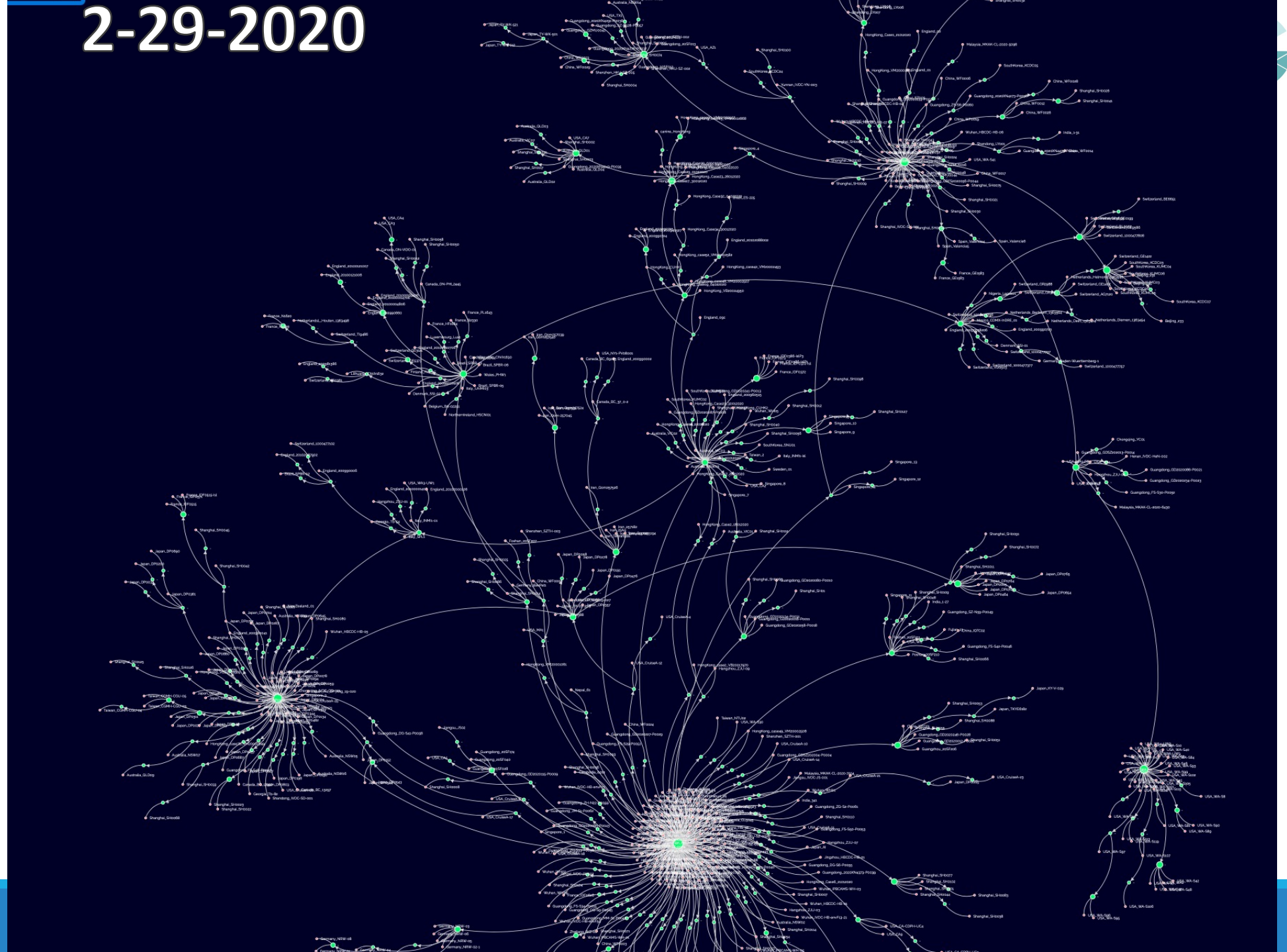
1-31-2020

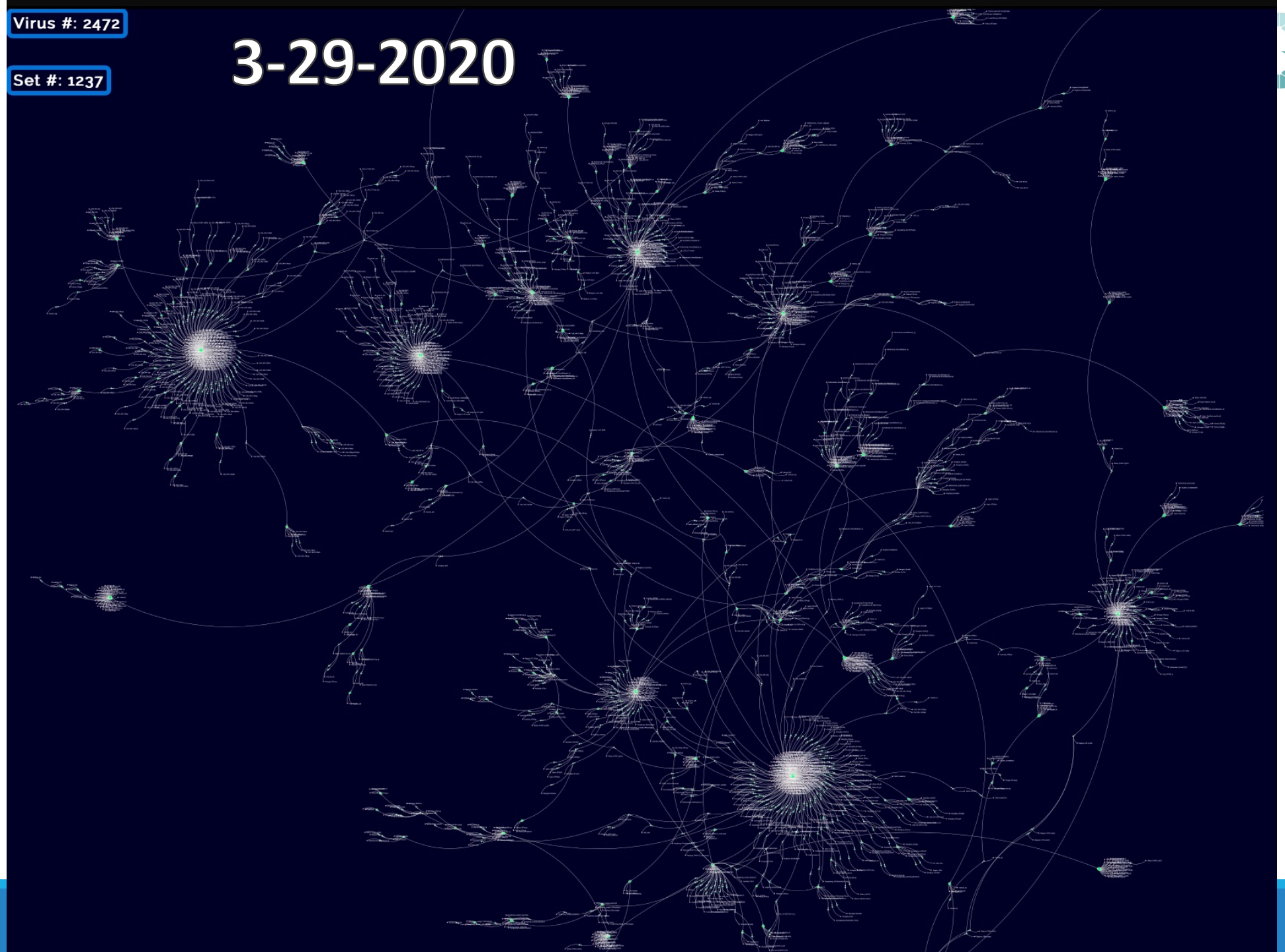


GRAPHEN
An Artificial Intelligence Company

2-29-2020

GRAPHEN
An Artificial Intelligence Company

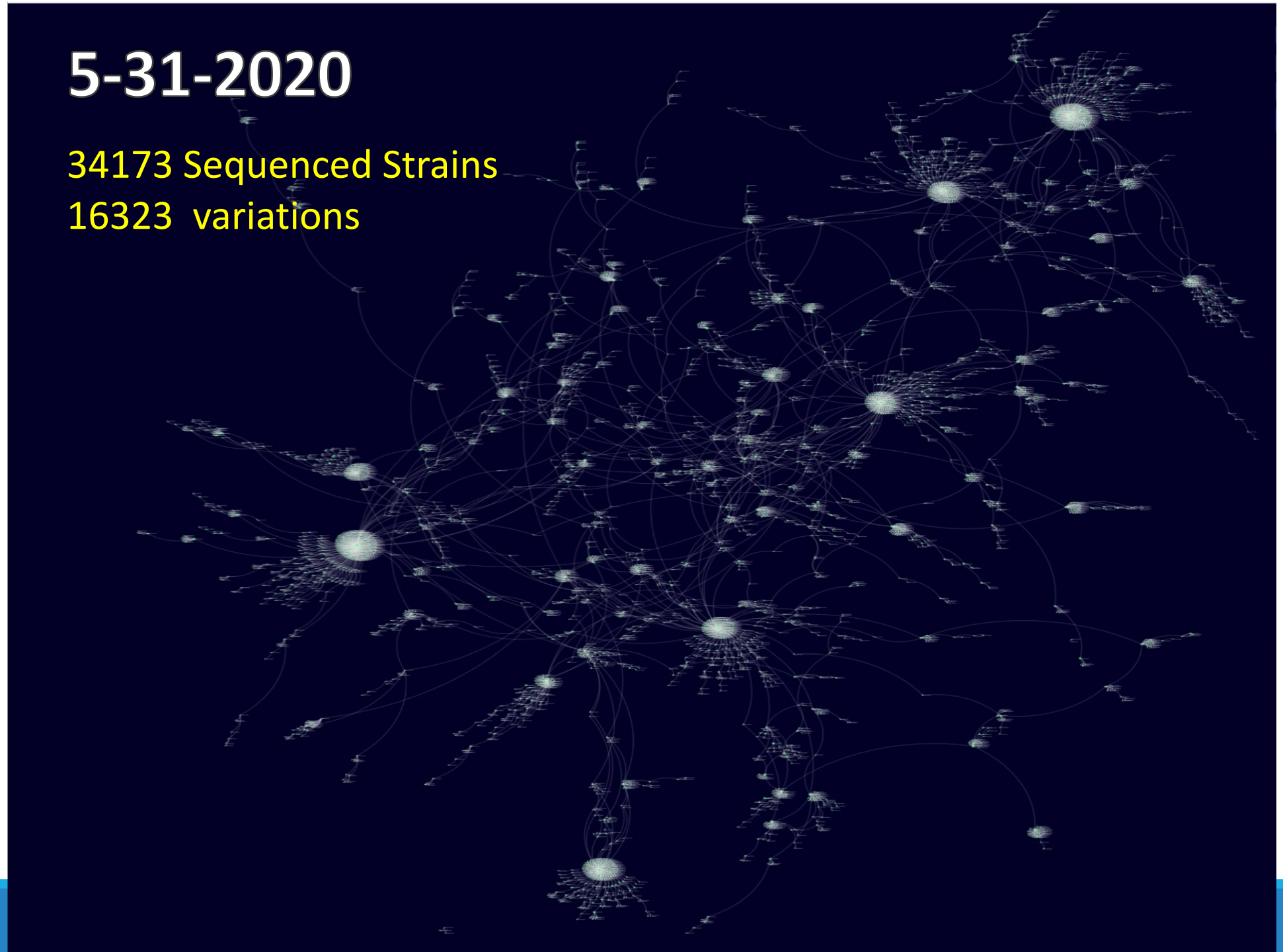




GRAPHEN
An Artificial Intelligence Company

5-31-2020

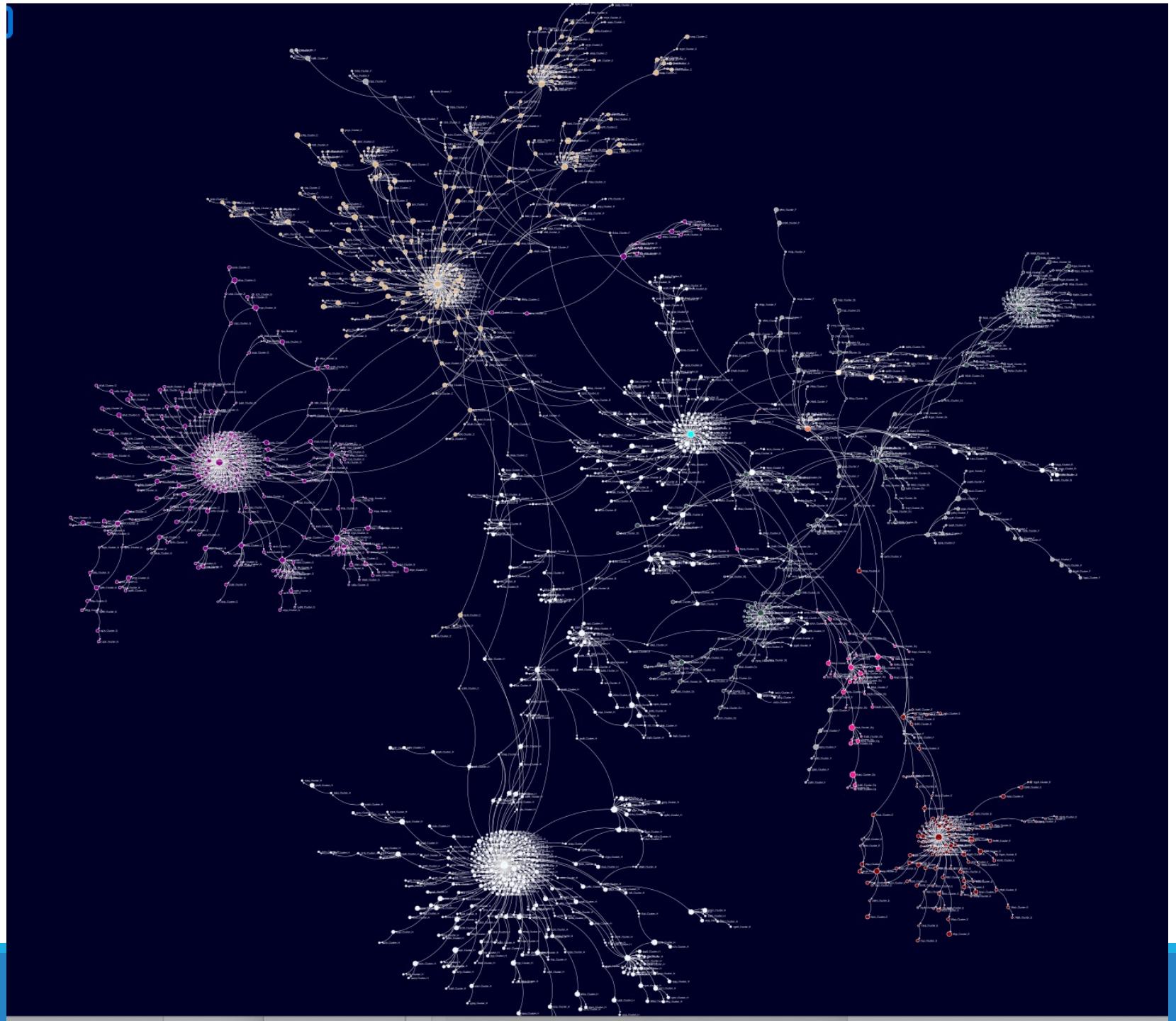
34173 Sequenced Strains
16323 variations



GRAPHEN
An Artificial Intelligence Company

October 21, 2020

Sequenced SARS-COV-2 gene: 147,379



GRAPHEN
An Artificial Intelligence Company

Graphen's Main Objectives

GRAPHEN
An Artificial Intelligence Company

Ardi AI Platform

Contextual Analysis| Autonomous Learning

Anita AI Finance

Fimmuno | Customer Intelligence | Trader

Adam AI Medical

Explorer | Genome | View

Graphen was established in 2017

Graphen Core -- Full-Brain AI Platform

GRAPHEN
An Artificial Intelligence Company

Developing advanced Graph Computing technology to realize a unique novel AI platform that simulates the whole brain functions and powers novel industry solutions.

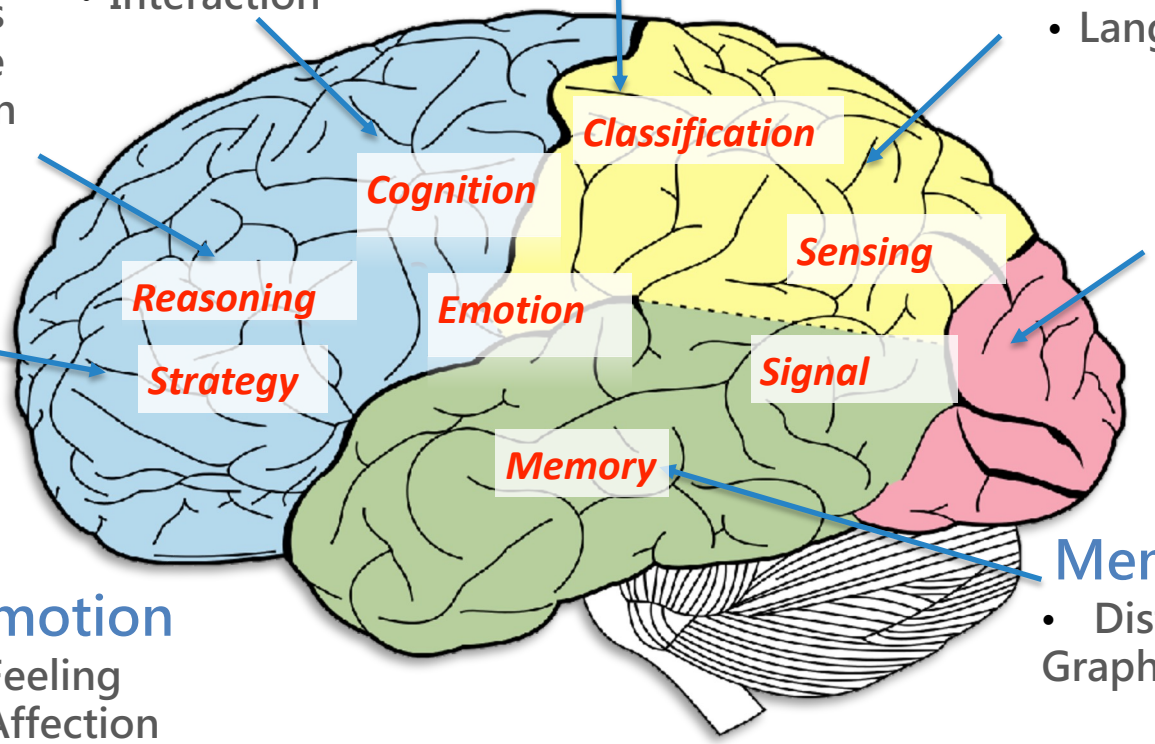
Ardi AI Platform

Learning
- Machine Learning and Deep Learning
- Autonomous Learning

Expression
- Big Data Visualization
- Explainability
- Interaction

Reasoning
- Bayesian Networks
- Causality Inference
- Behavior Prediction

Understanding
- Deep Vision Understanding
- Language Understanding



Perception
- Network Analysis
- Feature Analysis

Strategy
- Game Theory
- Optimal Solution

Personality
- Humanization
- Ethics

Knowledge
- Q&A, Conversation

Memory
- Distributed Native Graph Database

Pipeline
- Coordination
- Streamline Flow

Action
- Control
- Behavior

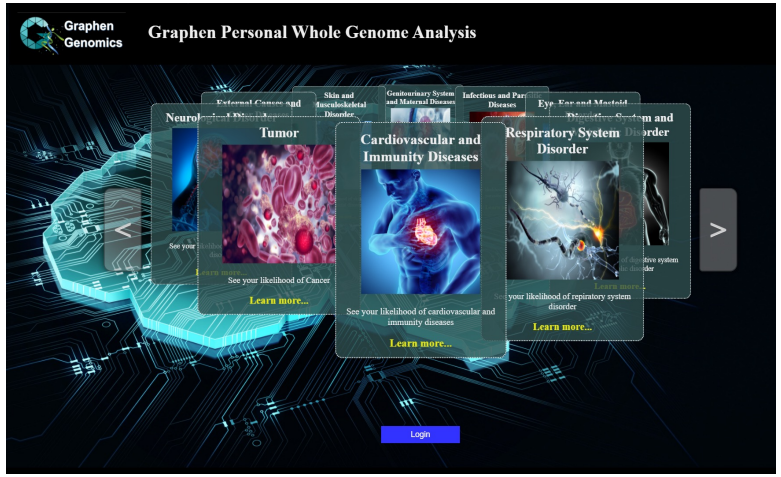
Emotion
- Feeling
- Affection

https://www.graphen.ai/products/ardi_feature.html

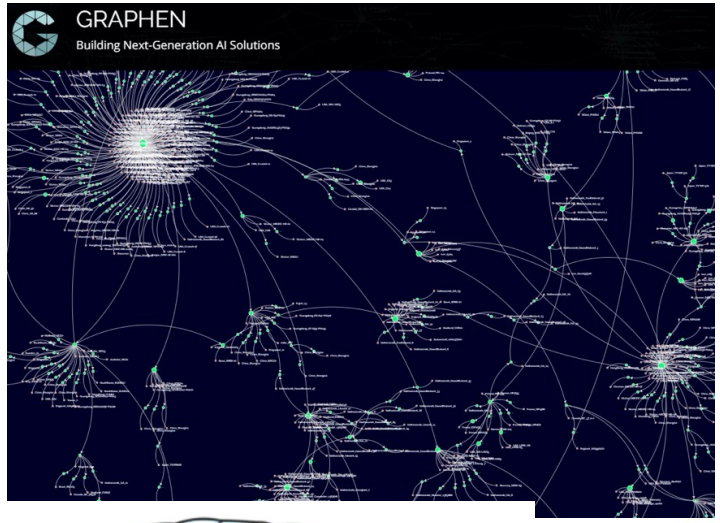
Graphen Medical AI Solutions

GRAPHEN
An Artificial Intelligence Company

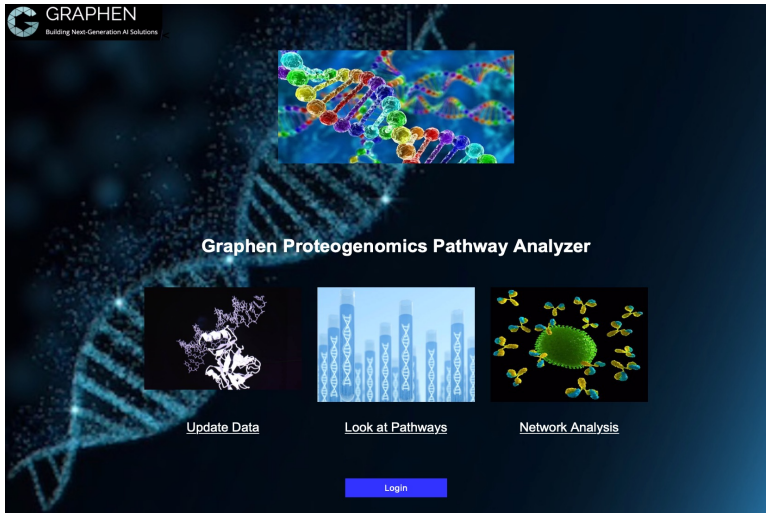
- Personalized Whole Genome Disease Analysis
- Large-Scale Medical Article Understanding by Machine

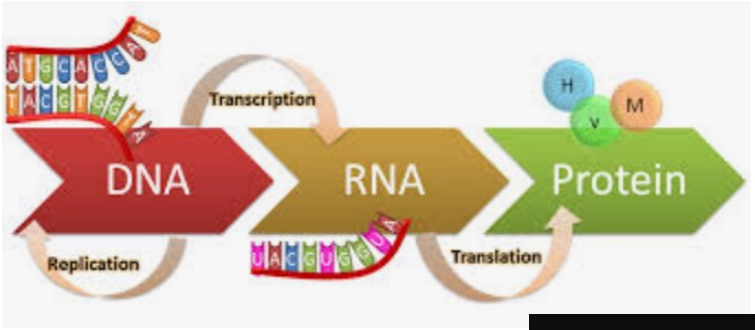


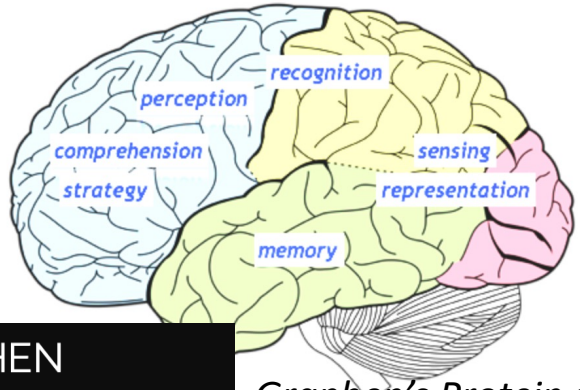
- AI Drug Developments – Protein Structure and Drug Binding Predictions
- Virus Genome Mutation Analysis

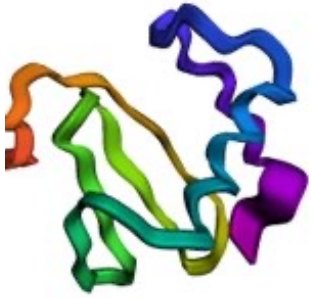


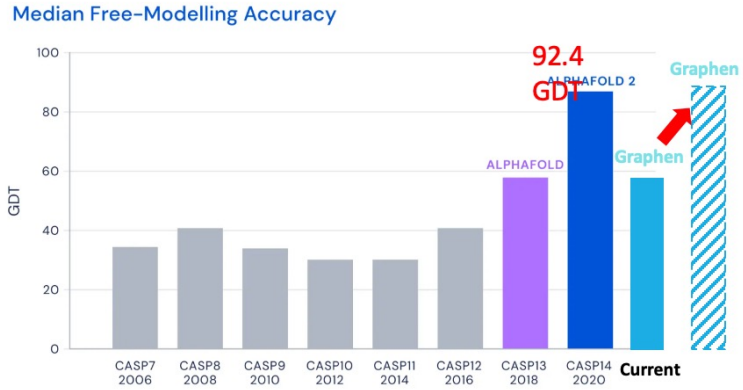
- Personalized Disease Pathway Analysis
- Precision Medicine Suggestion











GRAPHEN
Building Next-Generation AI Solutions

Graphen's Protein structure and functional prediction from DNA is like being able to create 'compiler' that translates software codes to executables.

Major AI Technologies behind our COVID-19 Monitoring works

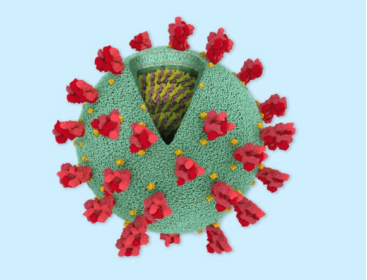
GRAPHEN
An Artificial Intelligence Company

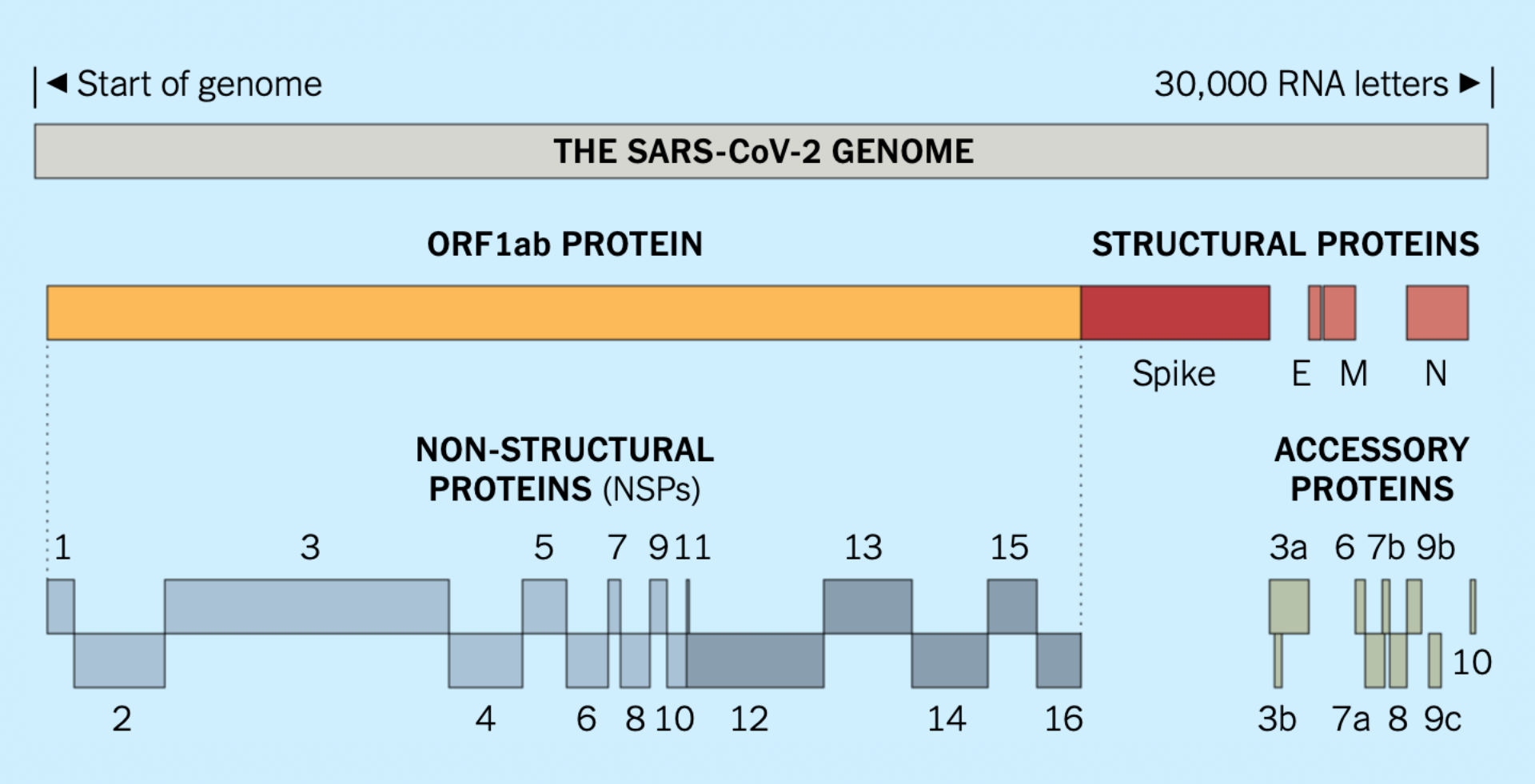
- Large-Scale Graph Database
- Real-Time System
- Graph Spectral Clustering
- Graph Visualization

28

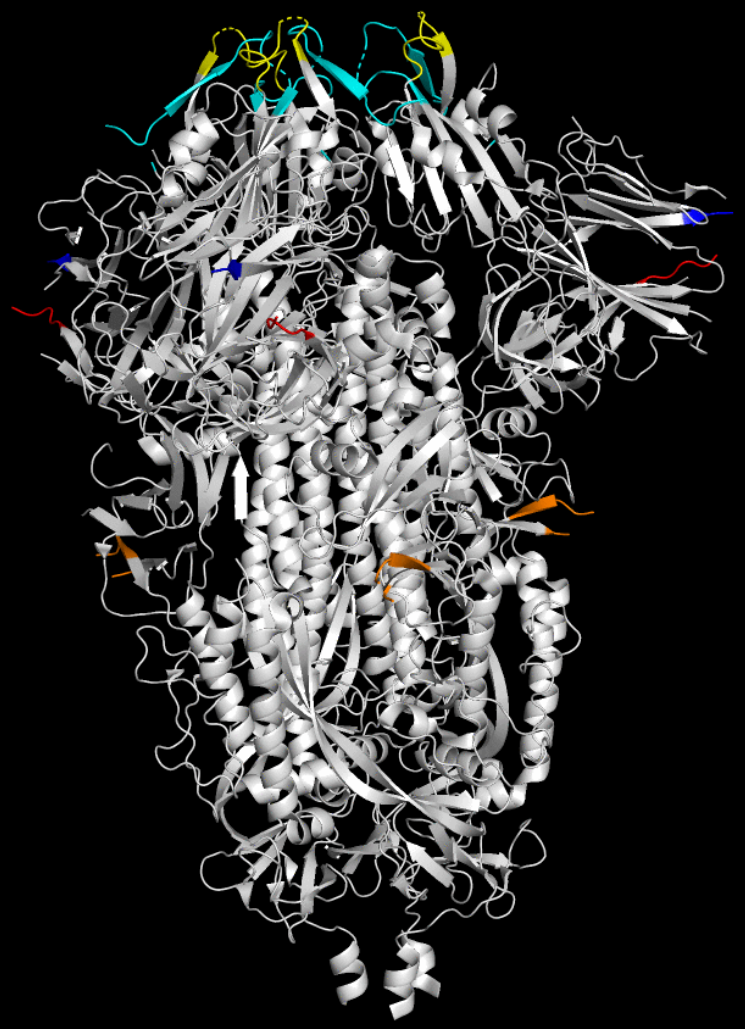
Genome is the Parts of Machine

Genome is the Software. Genome is the Hardware.





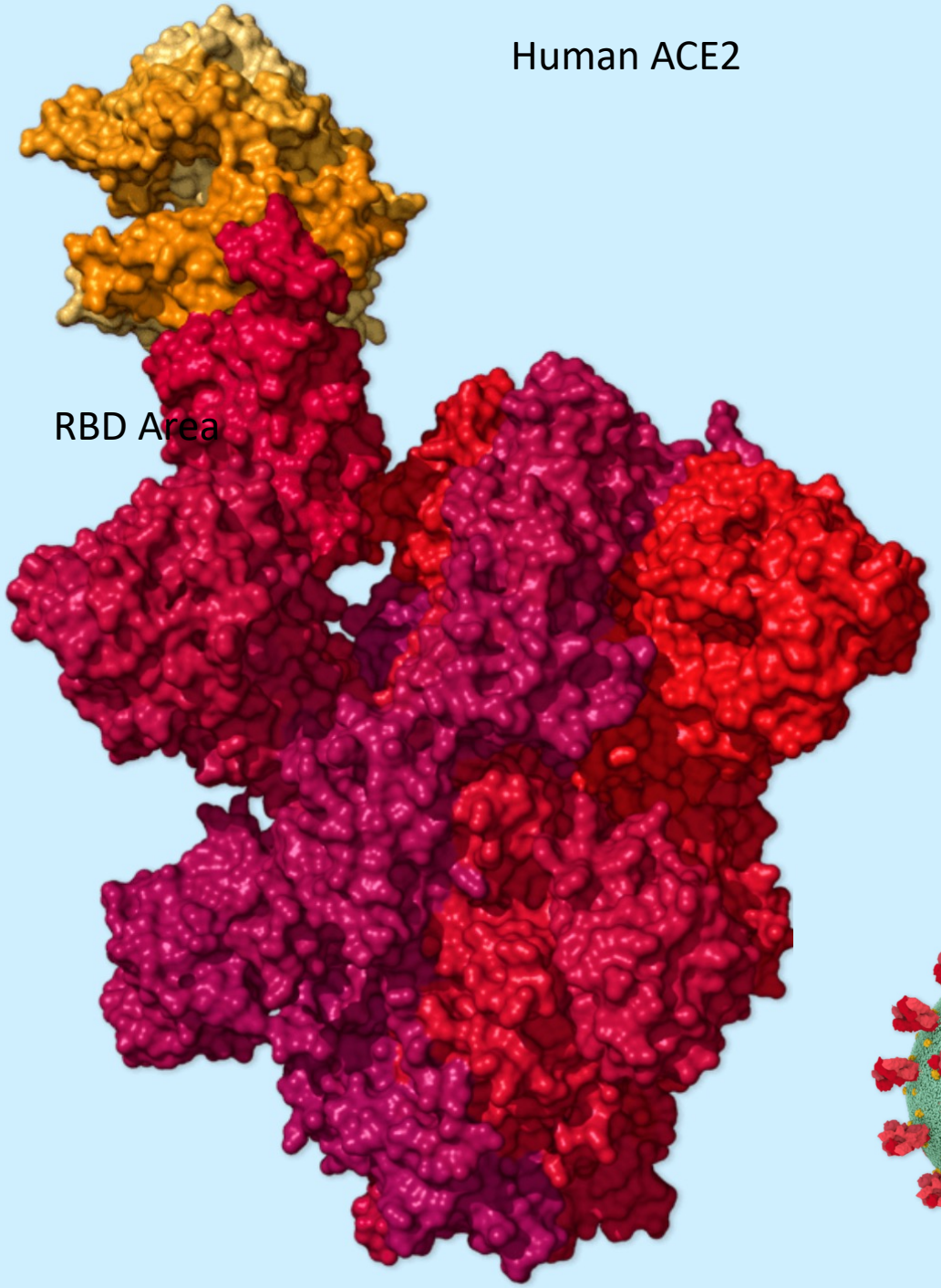
S protein Structure



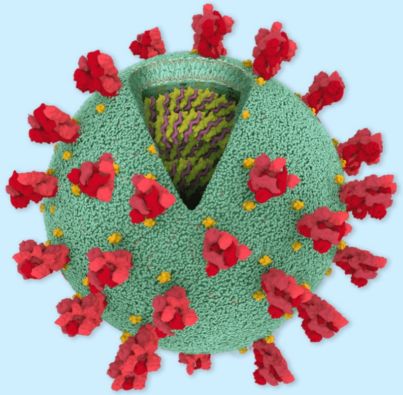
Colored areas are 6 significant different sections comparing the standard Wuhan HU-1 with two prior coronaviruses found in bats (ZC45 and ZXC21)

GRAPHEN
An Artificial Intelligence Company

Human ACE2

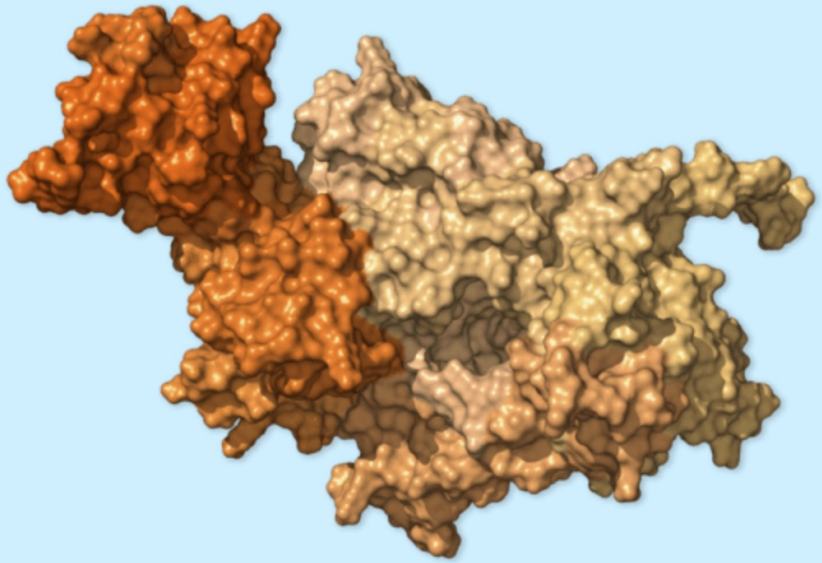


Each Spike is a combination of 3 Spike Proteins

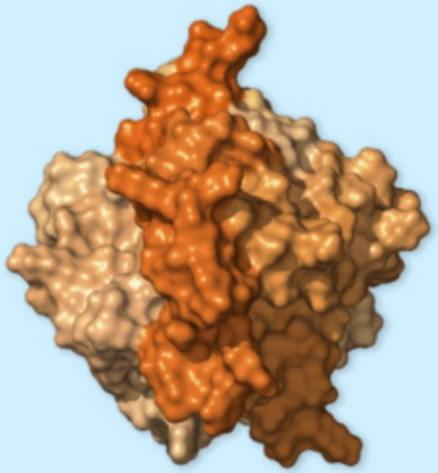


GRAPHEN
An Artificial Intelligence Company

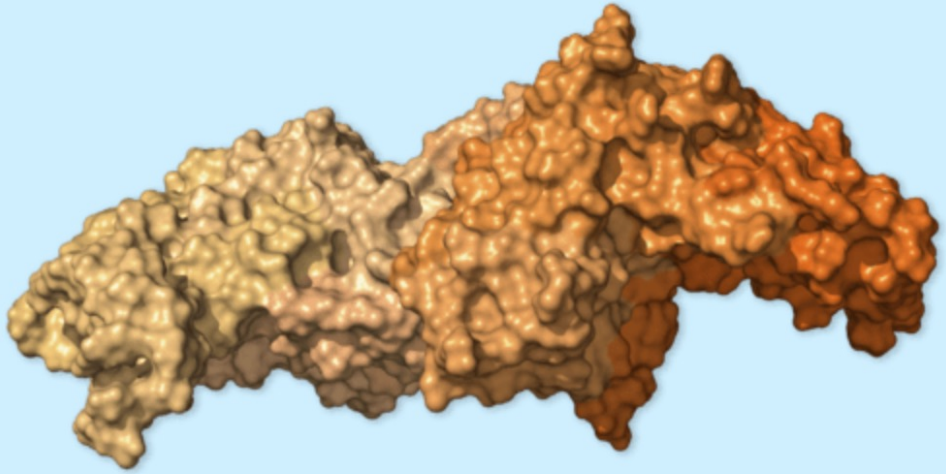
Copy Machine · NSP12



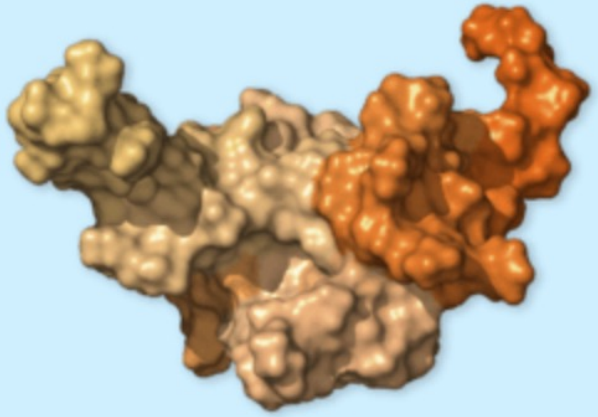
Escape Artist · ORF3a



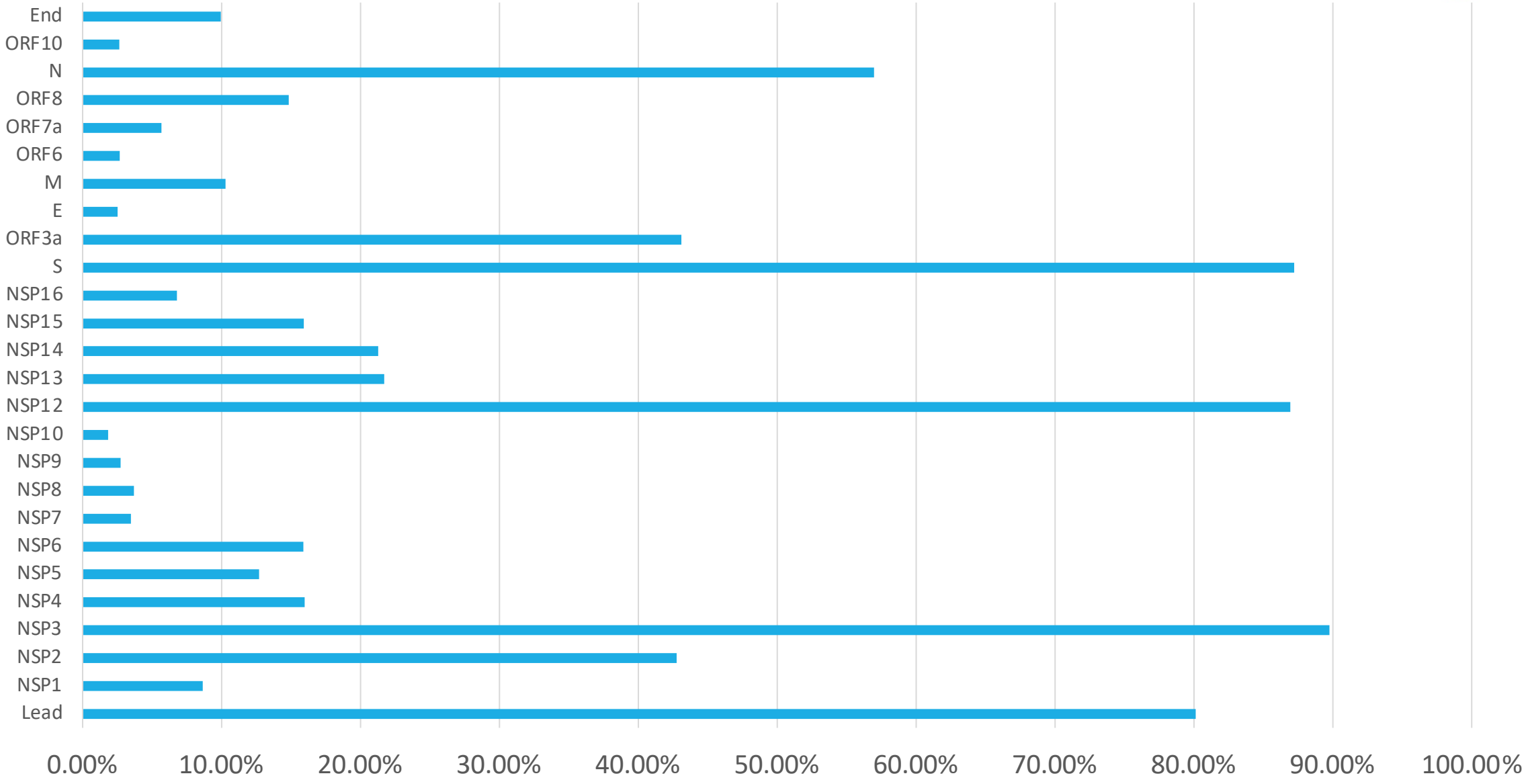
Viral Proofreader · NSP14



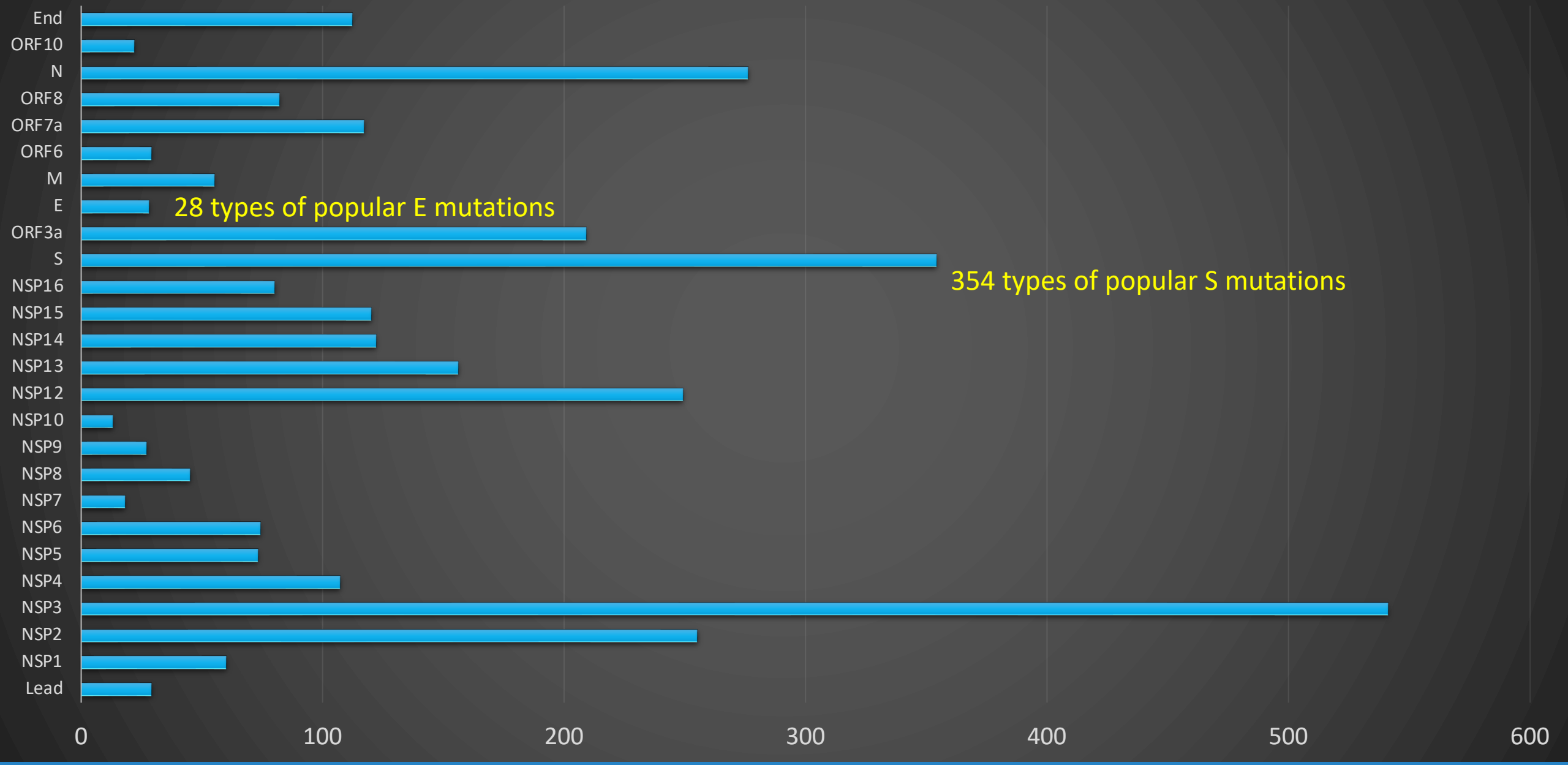
Genetic Camouflage · NSP10



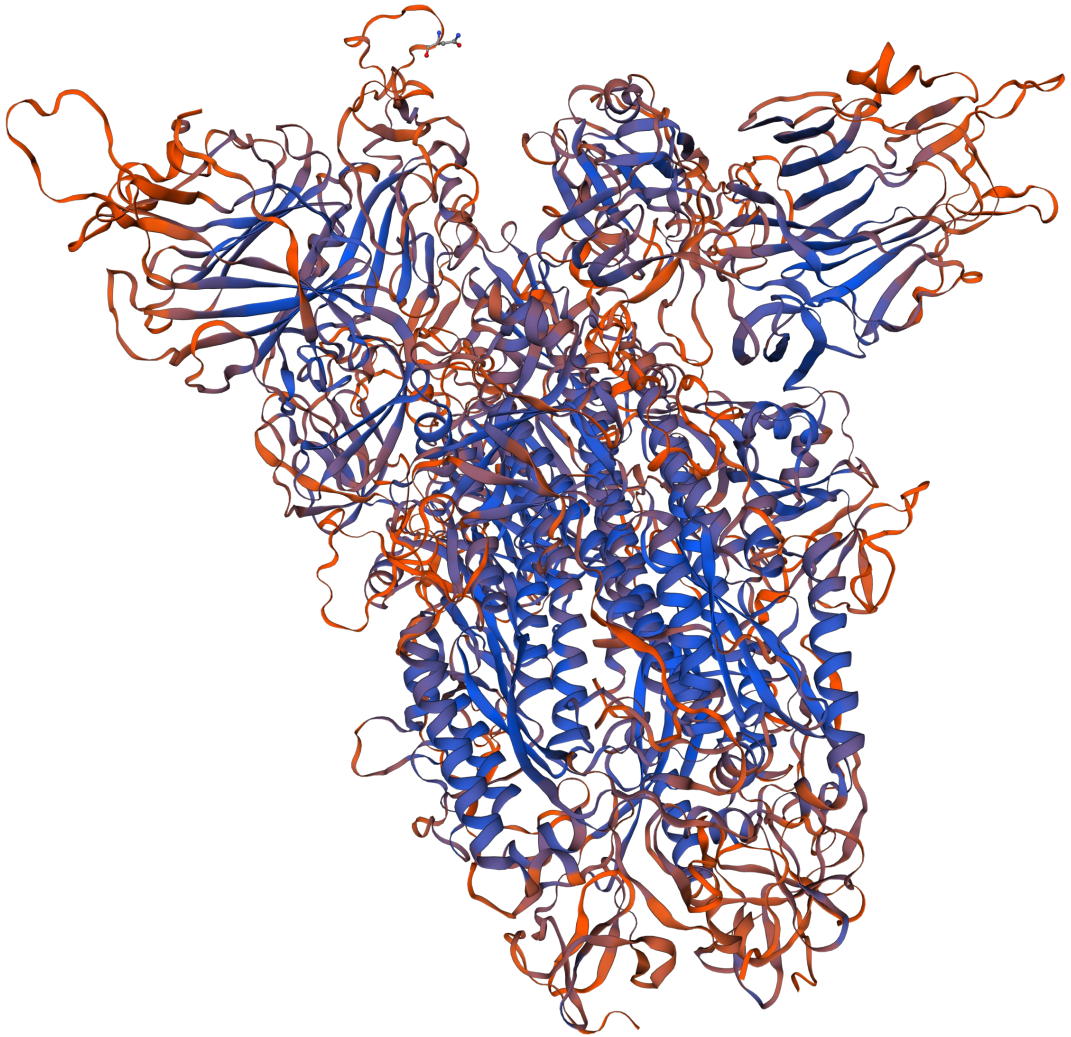
Mutated Strains Percentage (Total: 395,018, by 1/25/2021)

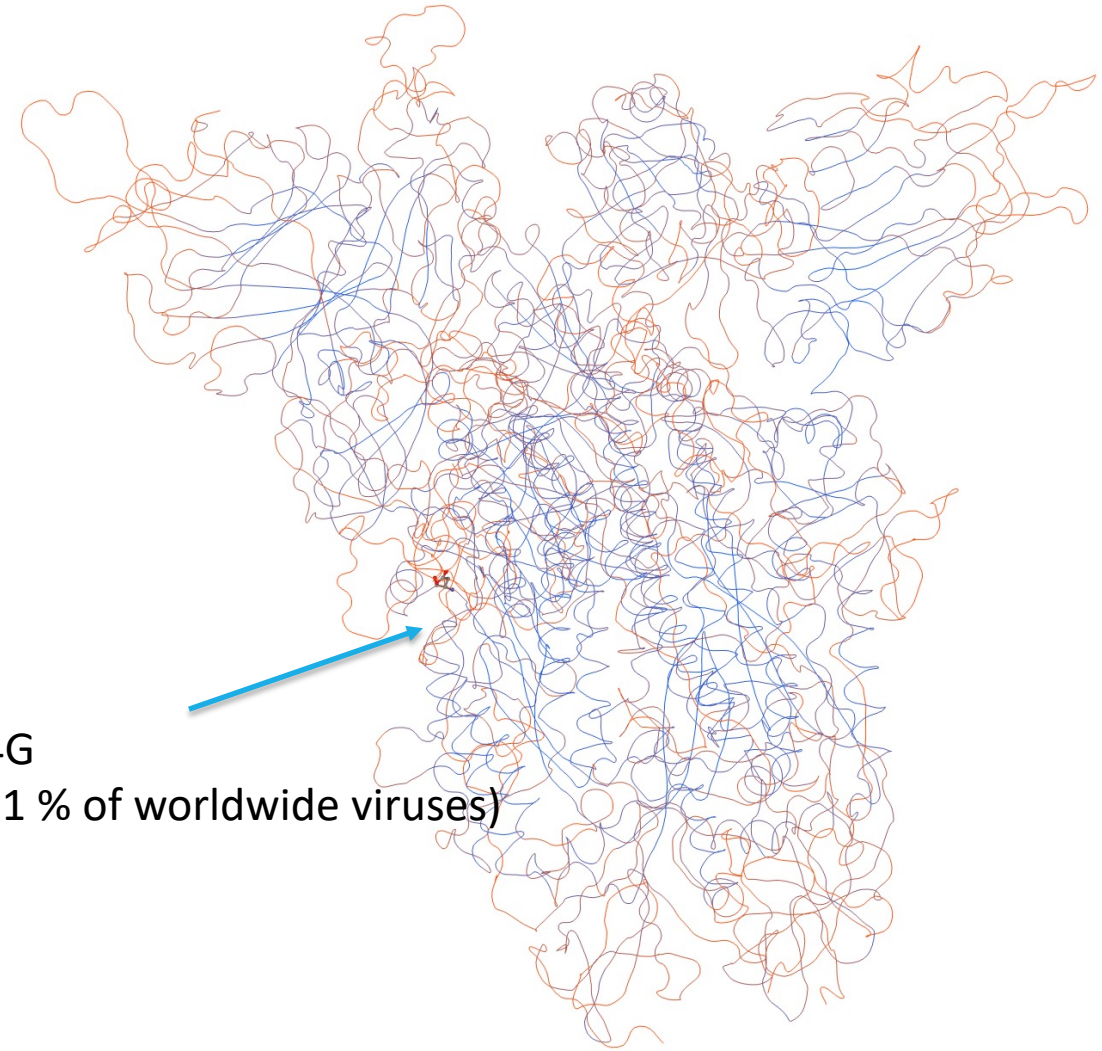


Types of Missense Mutations in each protein with at least 10 mutated strains



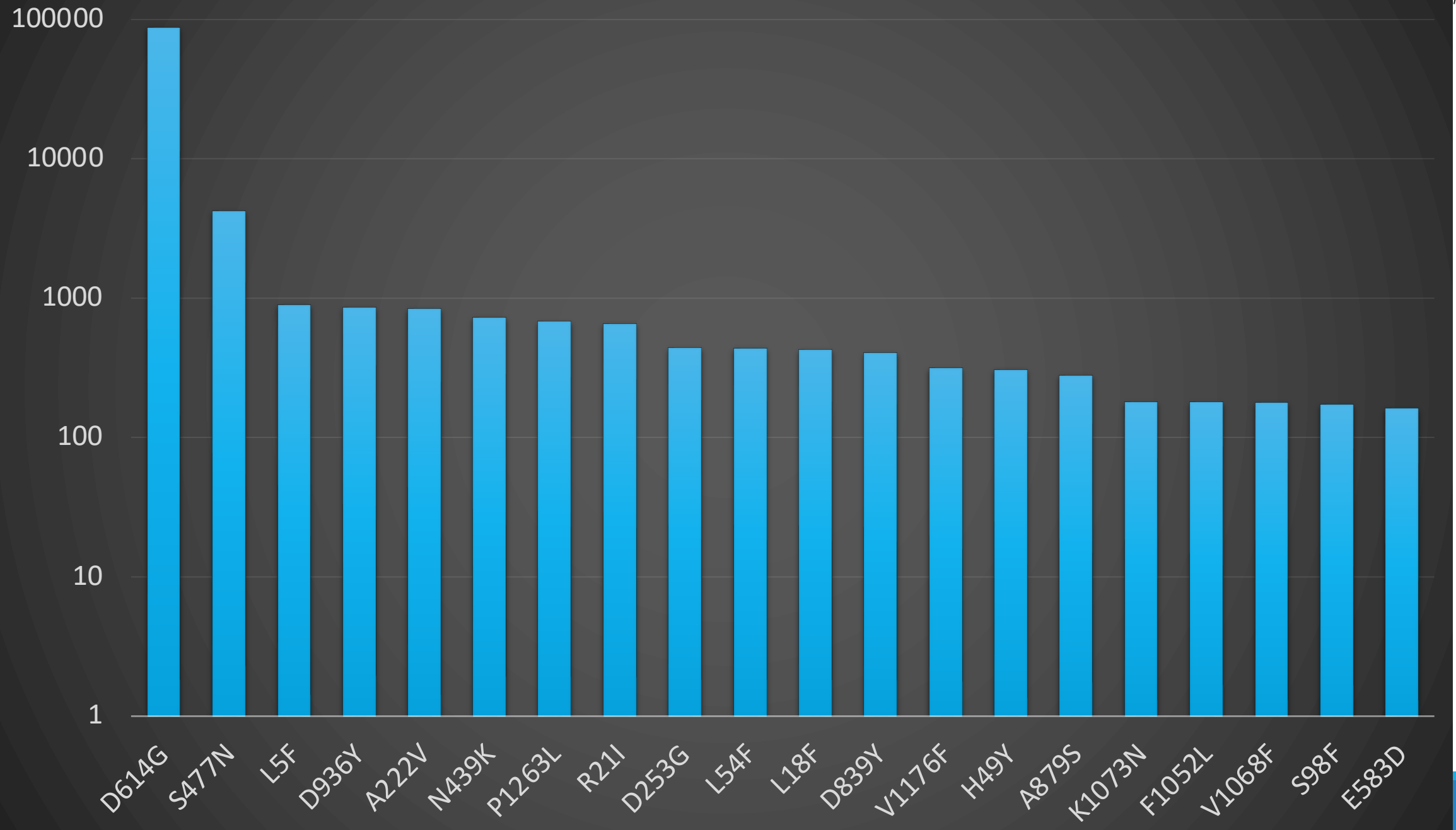
S protein Structure and its key Mutation





D614G
(86.11 % of worldwide viruses)

Number of Mutated Strains in Spike Protein (log scale)

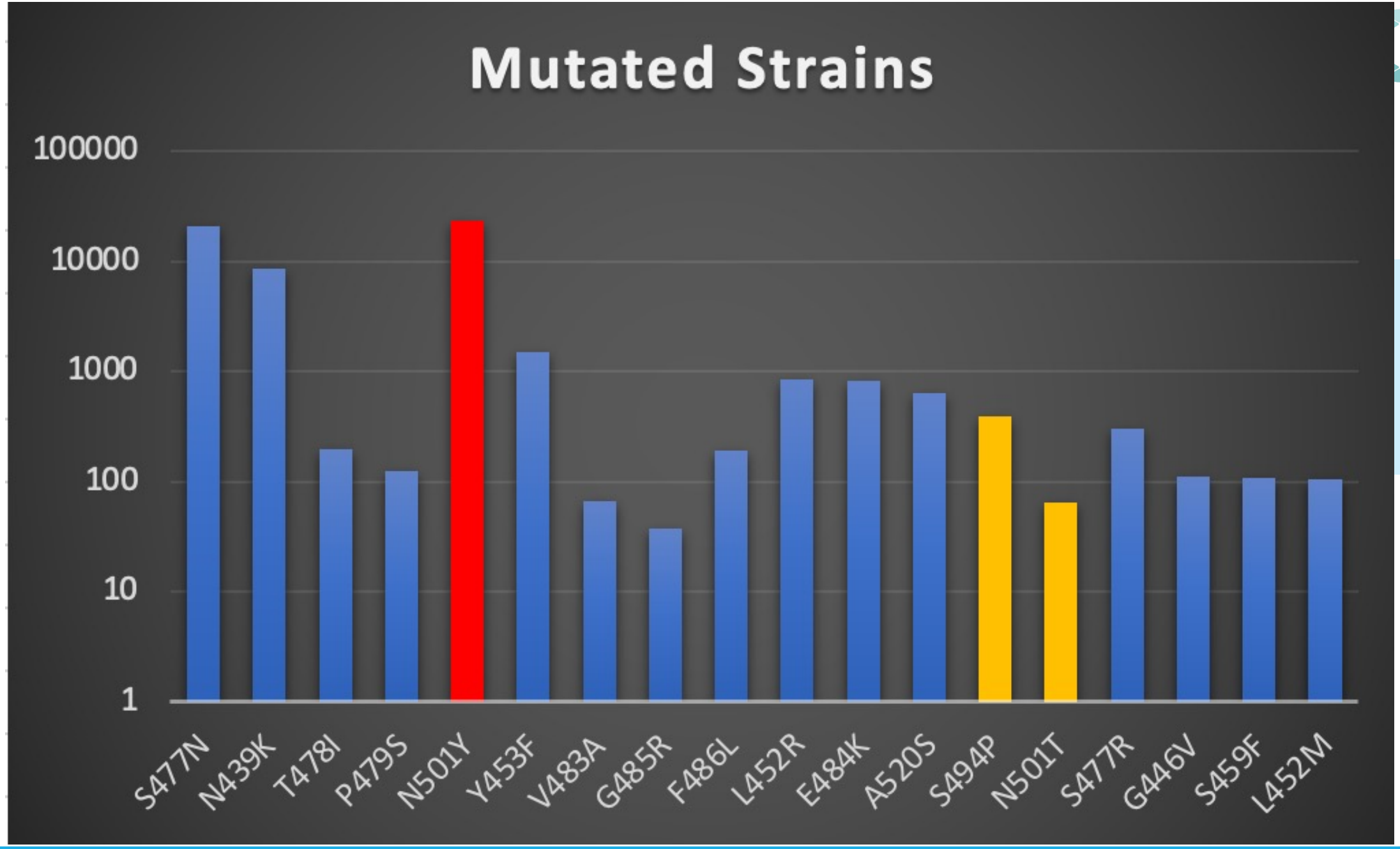


Amino Acids Residues 'Distance' of Similarity

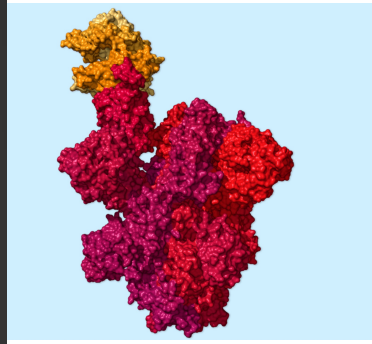
Arg	Leu	Pro	Thr	Ala	Val	Gly	Ile	Phe	Tyr	Cys	His	Gln	Asn	Lys	Asp	Glu	Met	Trp	
110	145	74	58	99	124	56	142	155	144	112	89	68	46	121	65	80	135	177	Ser
	102	103	71	112	96	125	97	97	77	180	29	43	86	26	96	54	91	101	Arg
		98	92	96	32	138	5	22	36	198	99	113	153	107	172	138	15	61	Leu
			38	27	68	42	95	114	110	169	77	76	91	103	108	93	87	147	Pro
				58	69	59	89	103	92	149	47	42	65	78	85	65	81	128	Thr
					64	60	94	113	112	195	86	91	111	106	126	107	84	148	Ala
						109	29	50	55	192	84	96	133	97	152	121	21	88	Val
							135	153	147	159	98	87	80	127	94	98	127	184	Gly
								21	33	198	94	109	149	102	168	134	10	61	Ile
									22	205	100	116	158	102	177	140	28	40	Phe
										194	83	99	143	85	160	122	36	37	Tyr
											174	154	139	202	154	170	196	215	Cys
												24	68	32	81	40	87	115	His
													46	53	61	29	101	130	Gln
														94	23	42	142	174	Asn
															101	56	95	110	Lys
																45	160	181	Asp
																	126	152	Glu
																		67	Met

Table 2. Difference D for each amino acid pair (*10*). The mean chemical distance from the three-property formula (see text) $\bar{D}_{cpv} = 100$ (D_{ij} values have been multiplied by 50.723 to make this mean possible). Linear regression of *RSF* and log *RSF* on these D values gives correlation coefficients of −.66 and −.72, respectively. Previous difference indexes give correlation coefficients against *RSF* of −.34 (minimum base changes), −.42 (Sneath difference), and −.49 (Epstein formula). In each case, correlation is between the two sets (difference and *RSF*) of 190 values (*3*, *4*, *7*).

Mutated Strains

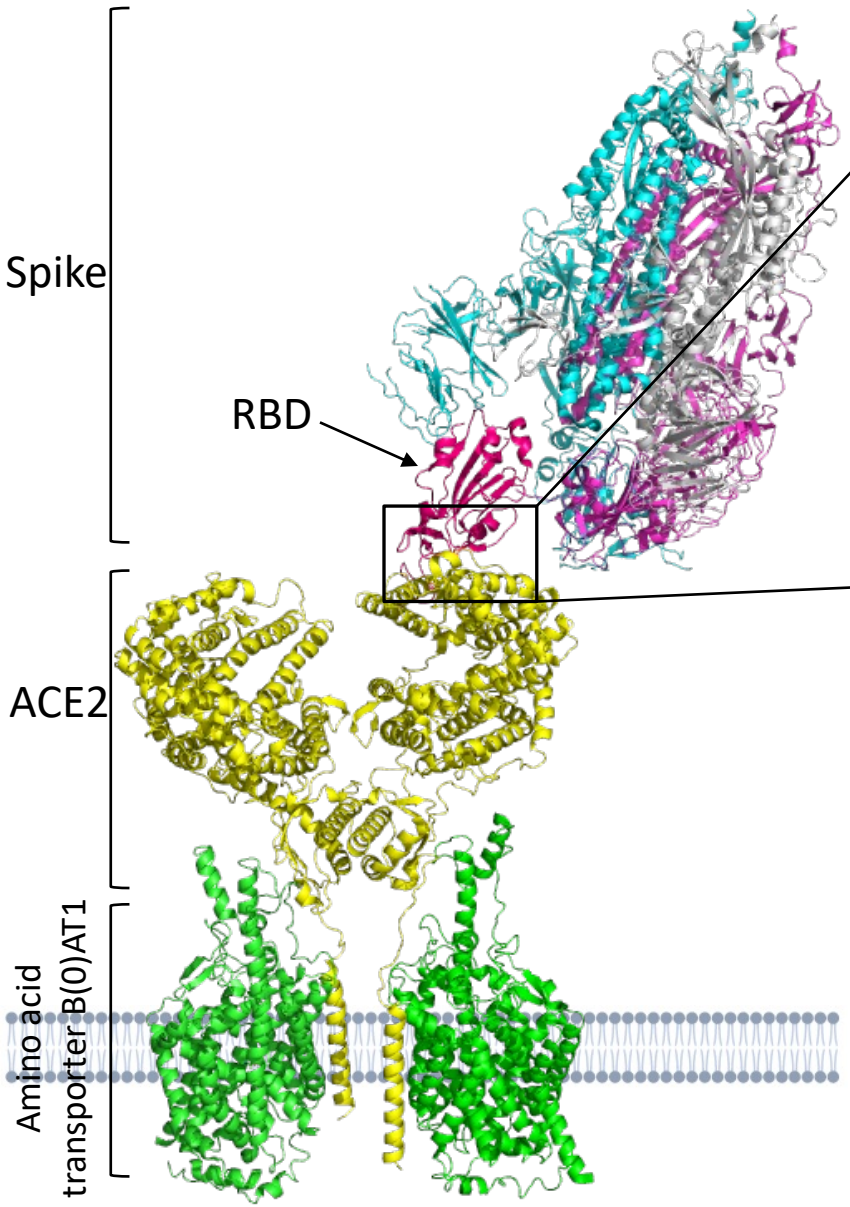


GRAPHEN
An Artificial Intelligence Company

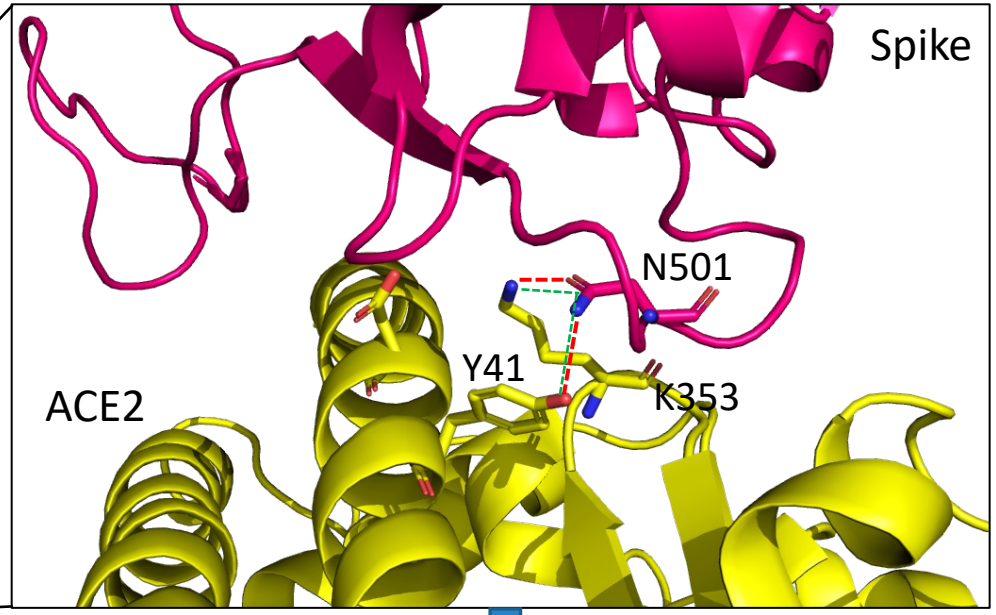


Spike N501Y increase the binding ability to ACE2

GRAPHEN
An Artificial Intelligence Company



90°

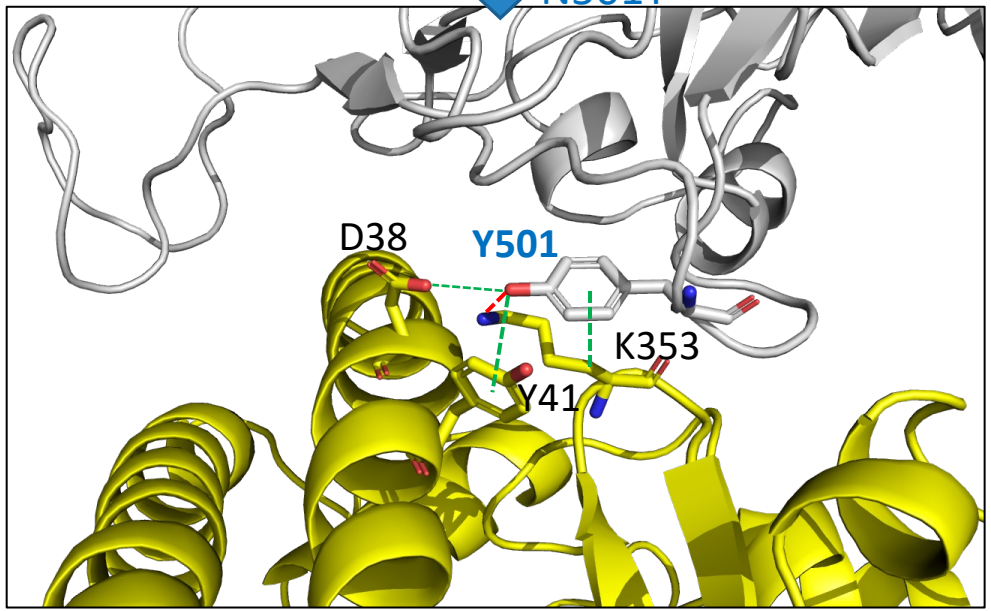


ACE2	Spike	Energy VDW	Energy HB
Y41	N501	-3.82	-2.46
K353	N501	-5.64	-3.5

Total energy: -15.42

--- Hydrogen bond
--- Van der Waals

N501Y

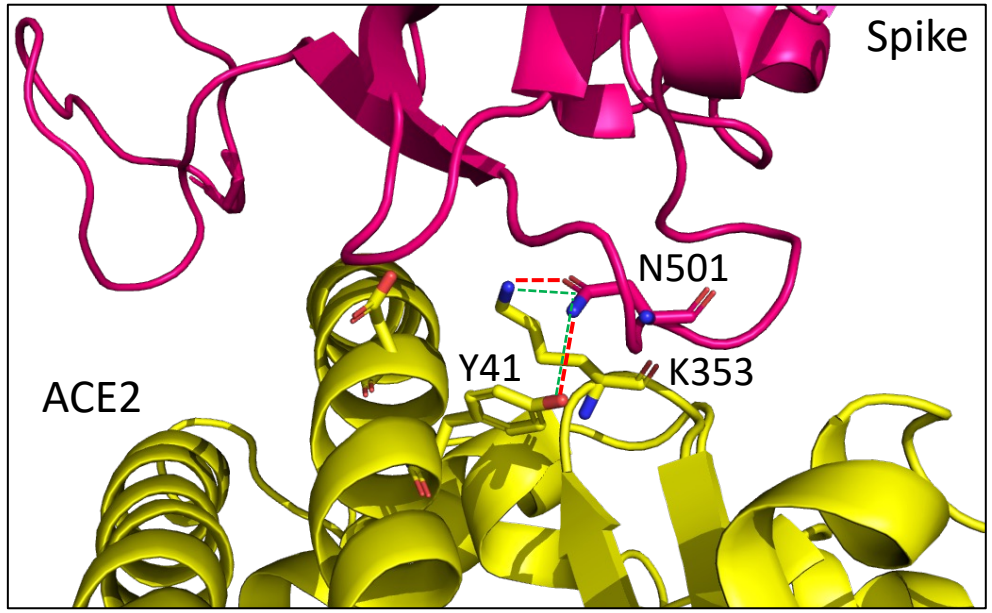


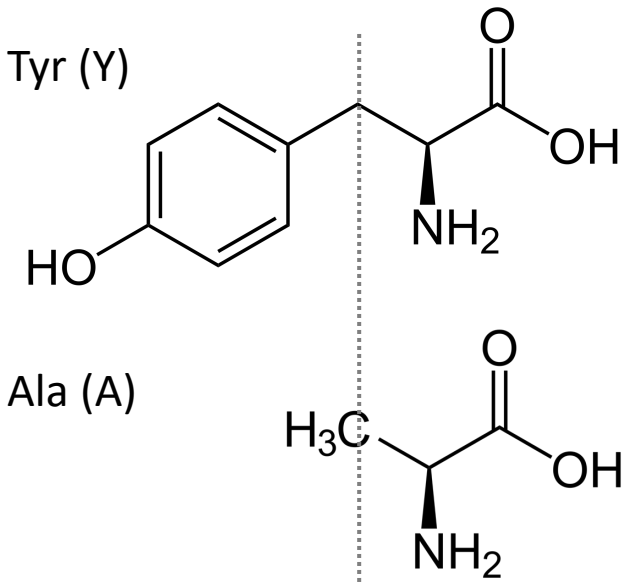
ACE2	Spike	VDW	HB
Y41	Y501	**-7.33**	0
K353	Y501	**-14.15**	-3.5
D38	Y501	-0.32	0

Total energy: -25.3

- Stabilize interface by increasing binding energy (1.6-fold) and number of contacting residues

ACE2 mutations and Spike binding effect





ACE2 mutations[1]

#N	Variation	Haddock Score	vdW (a.u.)	Elec (a.u.)	Effect
0	WT	-89.9 ± 5.2	-54.7 ± 2.4	-184.5 ± 9.3	WT
1	QAK24-26KAE	-95.2 ± 4.1	-52 ± 2.2	-196 ± 33	B
2	K31D	-85.7 ± 6.5	-56.4 ± 2.5	-176.9 ± 11.6	D
3*	E37A	-91.2 ± 3.7	-52.2 ± 1.1	-176.9 ± 30.6	A
4*	D38A	-86 ± 8.3	-55.5 ± 2.2	-174.8 ± 33.8	A
5*	Y41A	-92.1 ± 4.3	-52.7 ± 1.7	-171.7 ± 17	C
6	K68D	-85.9 ± 5.9	-55.5 ± 1.8	-188.8 ± 21	B
7	MYP82-84NFS	-99 ± 6.4	-53.9 ± 2.6	-176.8 ± 11.8	E
8	E110P	-90.2 ± 3.2	-56.2 ± 2.3	-208.3 ± 12.1	A
9	PD135-136SM	-102 ± 3.7	-56.6 ± 2.1	-210.3 ± 19.8	A
10	E160R	-90.5 ± 2	-55.5 ± 2.5	-182 ± 12.4	A

Effect C: Strongly inhibits interaction with SARS-CoV spike glycoprotein

1. Structural models of human ACE2 variants with SARS-CoV-2 Spike protein for structure-based drug design. *scientific data* (2020)

Monitoring Mutations is Critical to Drug / Vaccine Development

GRAPHEN
An Artificial Intelligence Company

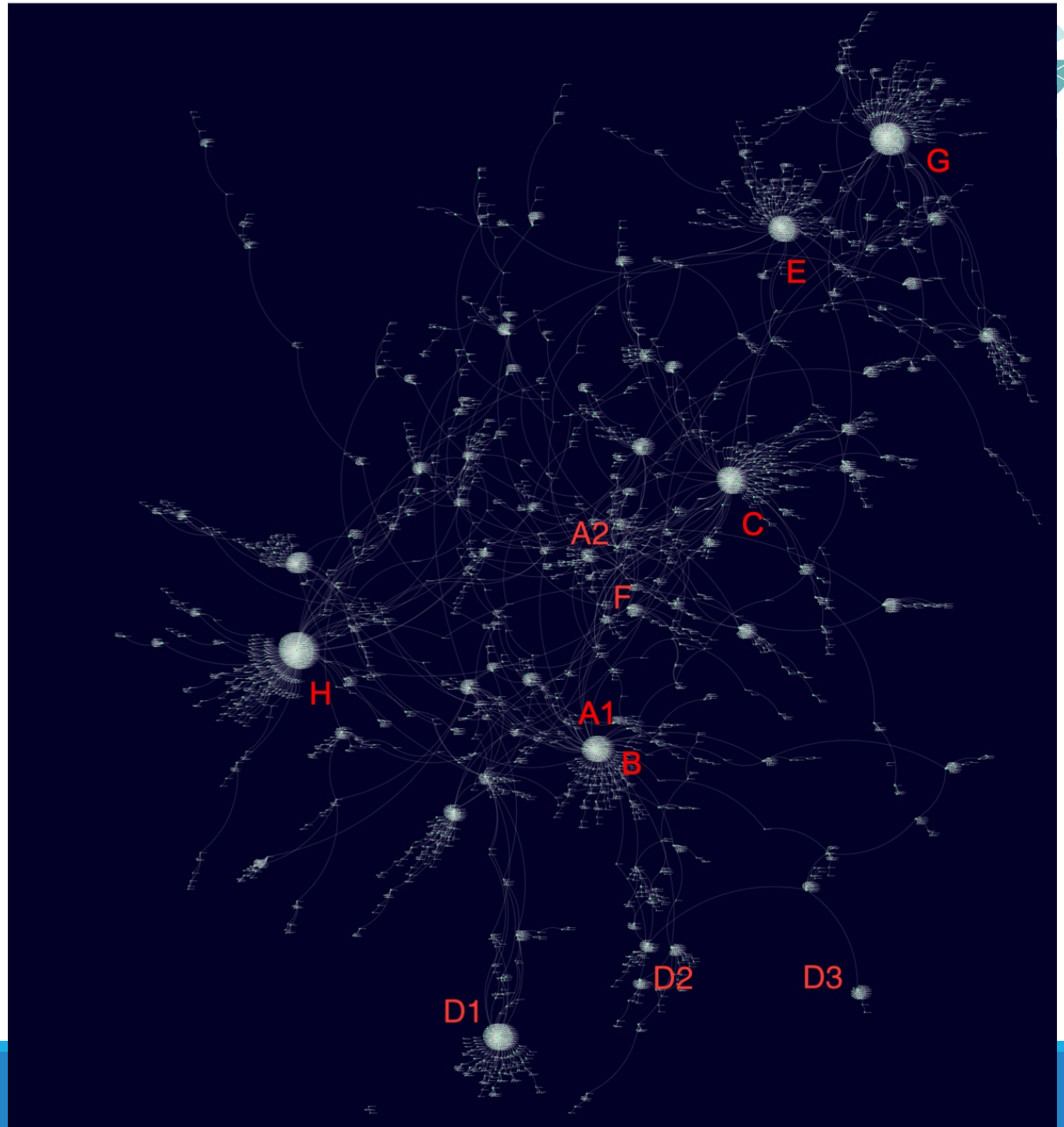
- PL protease
- M protease
- NSP12, NSP7, NSP8, and NSP14
- S protein:

-- RBD
-- RBM
-- Key Residues

Monitoring 4 types of mutations:

- Any Mutation
- Missense Mutations
- Silent Mutations
- Radical Mutations

Eight Major Types of Viruses Attacking the World



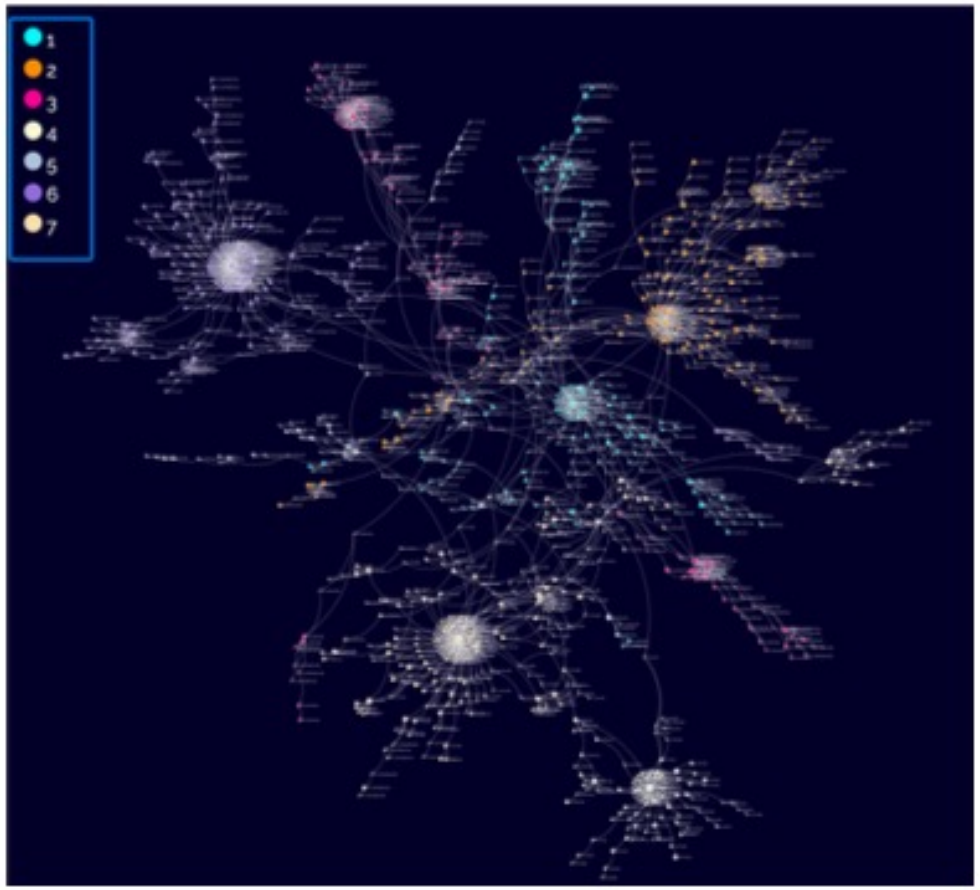
GRAPHEN
An Artificial Intelligence Company

Optimal Graph Spectral Analysis

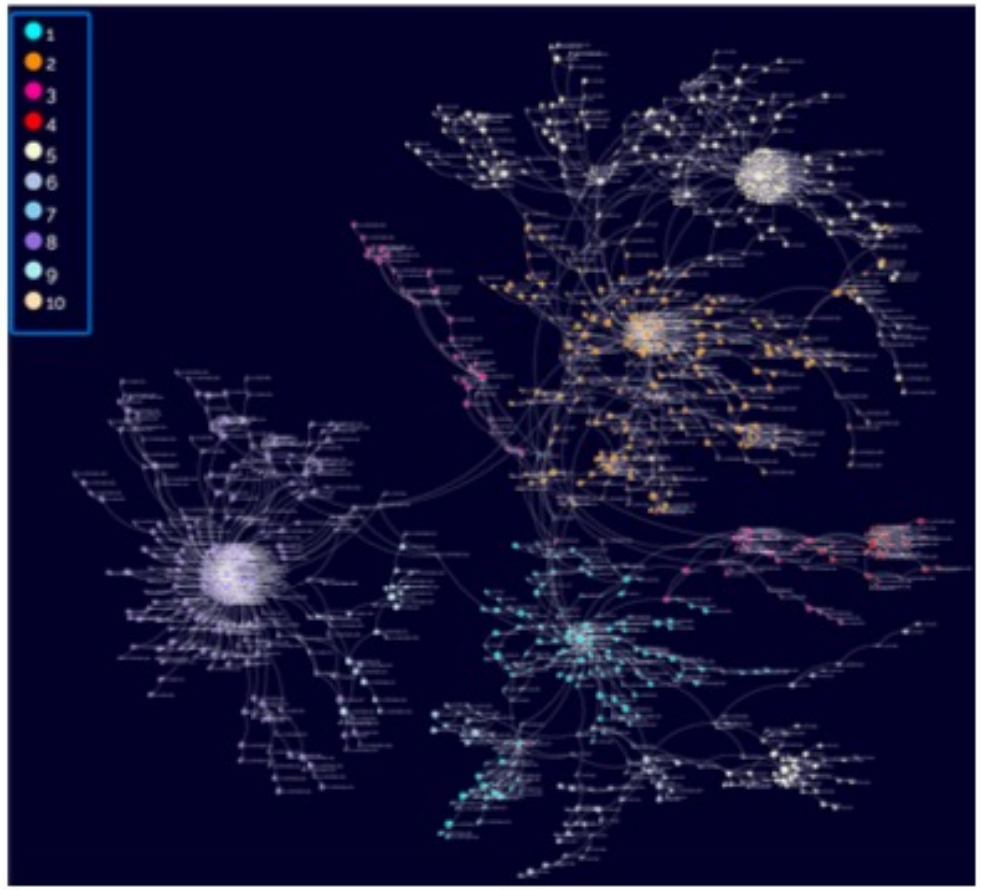
(a) Feb. 29, $K = 4$

(b) Mar. 31, $K = 7$

Optimal Graph Spectral Analysis

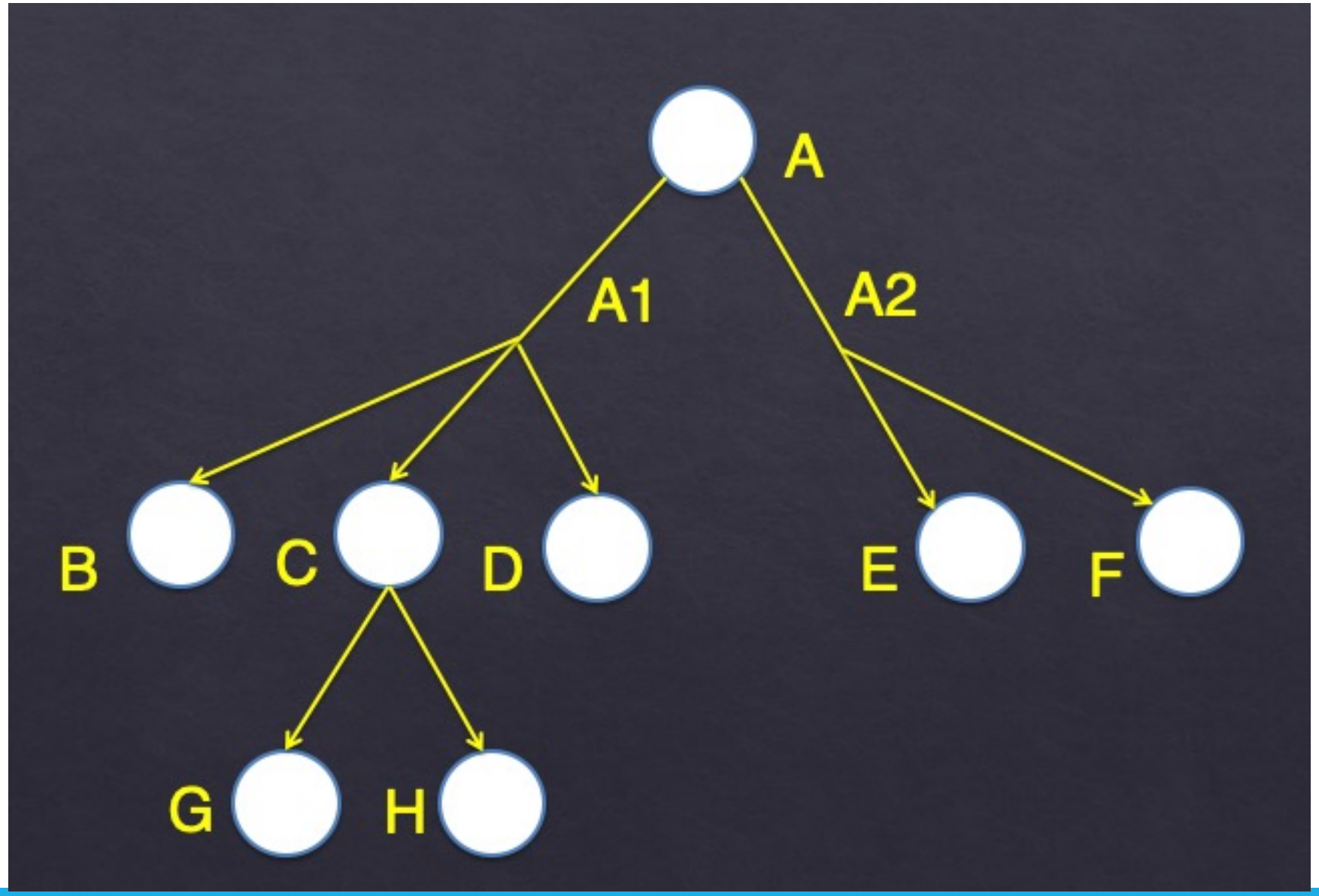


(c) April 30, $K = 7$



(d) May 31, $K = 10$

Eight Major Families

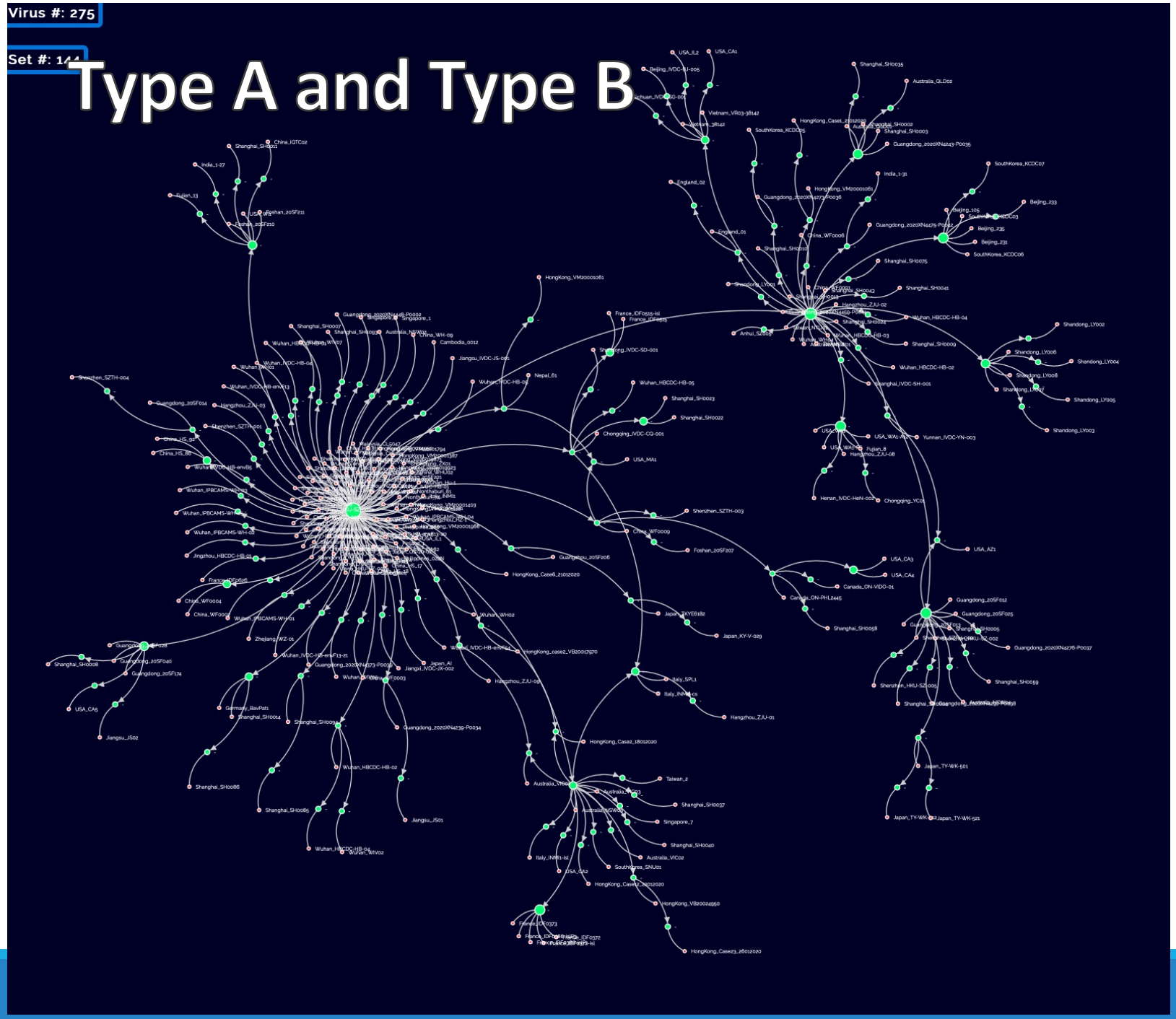


Distribution and Start of Eight Major Families

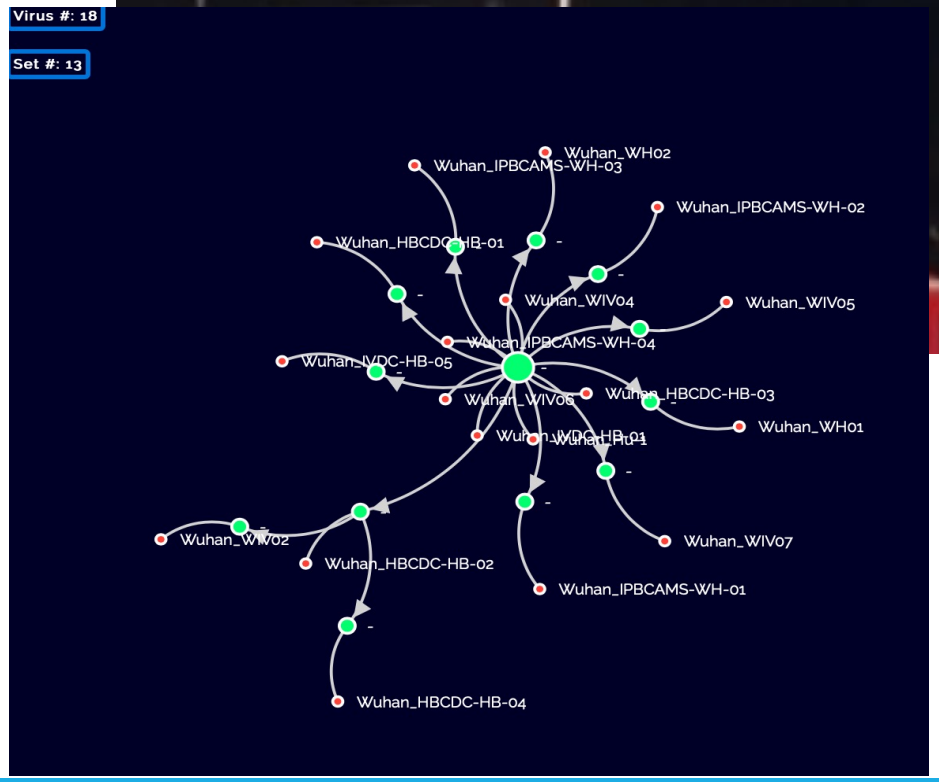
Cluster ID	Cluster Properties		
	Distribution region	***First Appearance Date***	***Features***
A	China	12/30/2019	All COVID-19 viruses may have evolved from here, among which the A2 virus is the closest to the virus genes on bats and pangolins.
B	China and other Asian countries	12/24/2019	Various virus strains evolved from the A1 virus in the A family.
C	Europe	1/28/2020	S protein variation.
D	United Kingdom, the Netherlands, Hong Kong	1/21/2020	NSP2, NSP3 and ORF3a protein variants.
E	United states of America West Coast, Canada	1/19/2020	Most of the ORF8 protein variants have two NSP13 protein variants.
F	Spain, Australia, South Korea, China	1/10/2020	All are viruses that have evolved directly from A2.
G	Europe, South America	2/16/2020	Variations of three consecutive sites on S protein and N protein.
H	France, United States of America East Coast	2/21/2020	S protein variation and ORF3a protein variation.

46

Type A and Type B

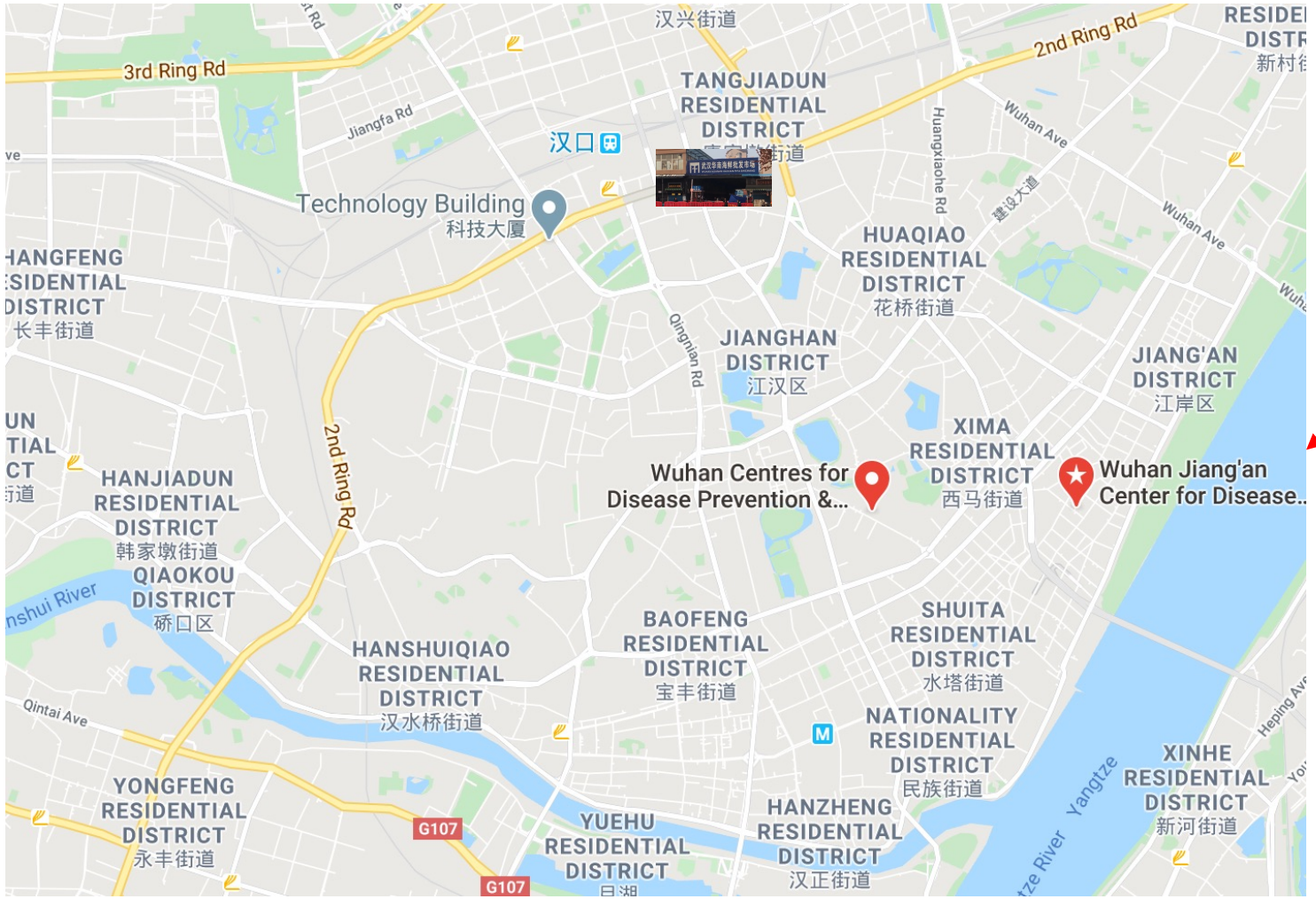


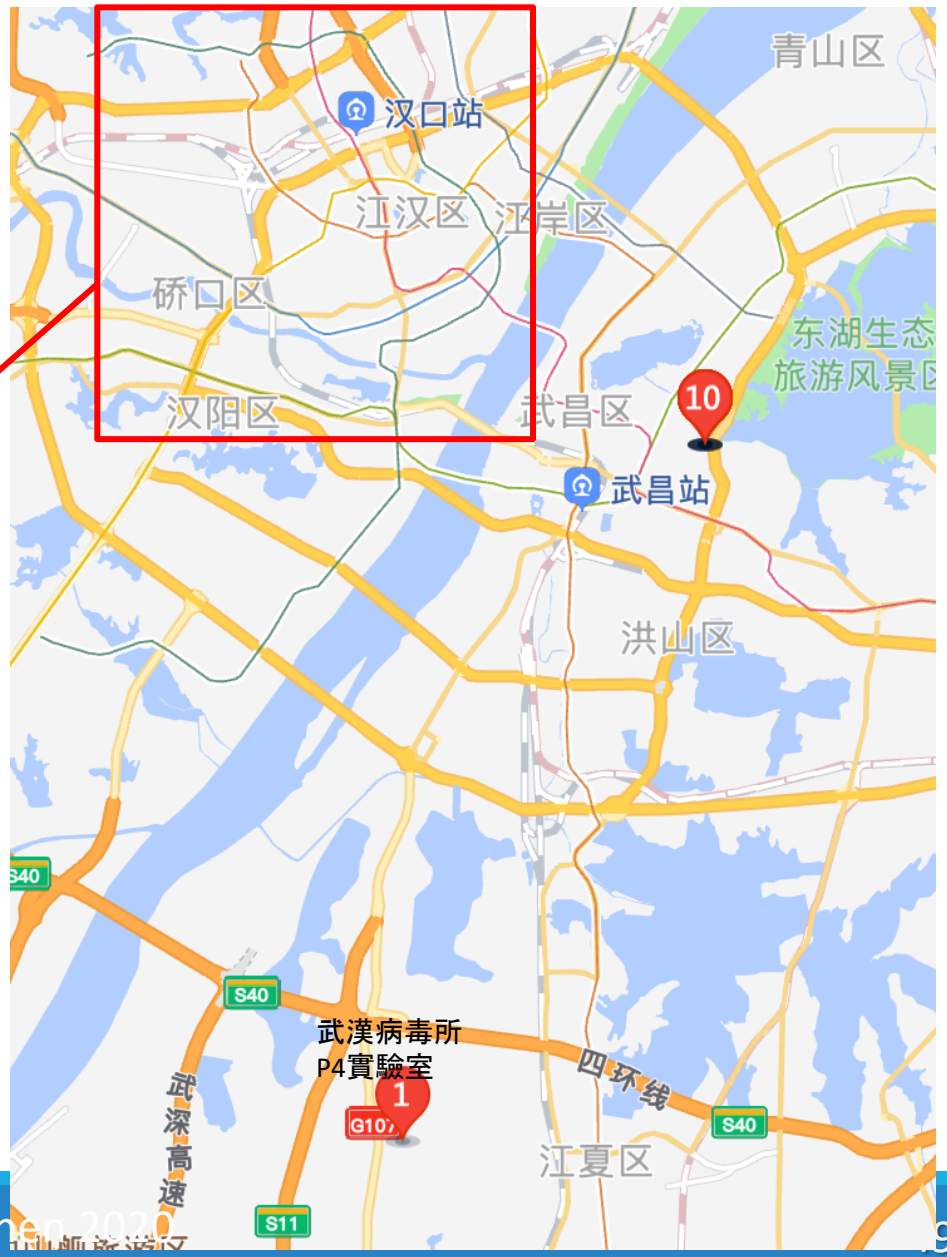
GRAPHEN
An Artificial Intelligence Company



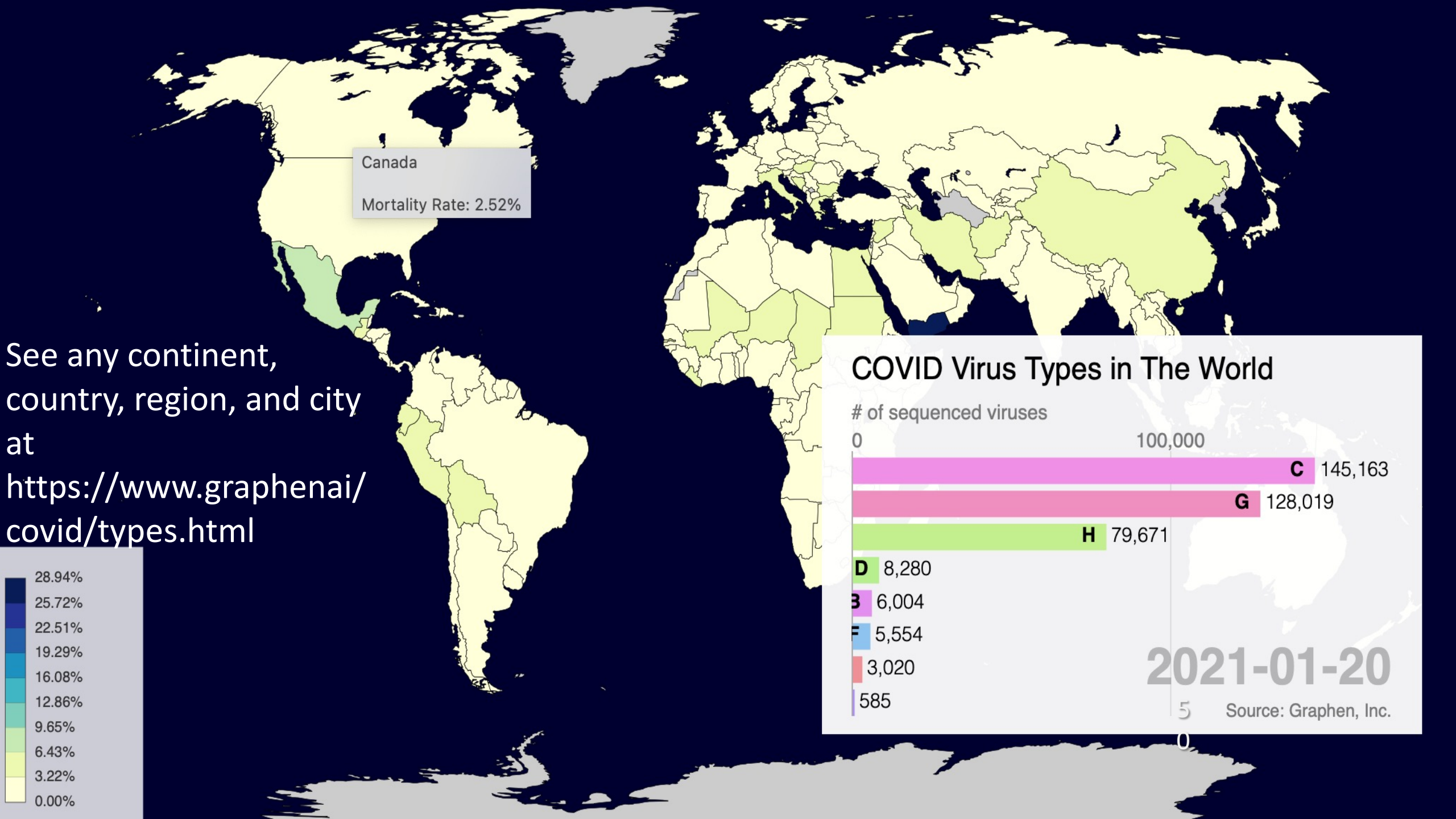
By 12/24/2019, there were 3-5 variants in various genomes ➔ Scientific evidence can almost certain the most widely spread COVID-19 started at late Nov or early Dec.

On average, SARS-CoV-2 adds a stable mutated point per week

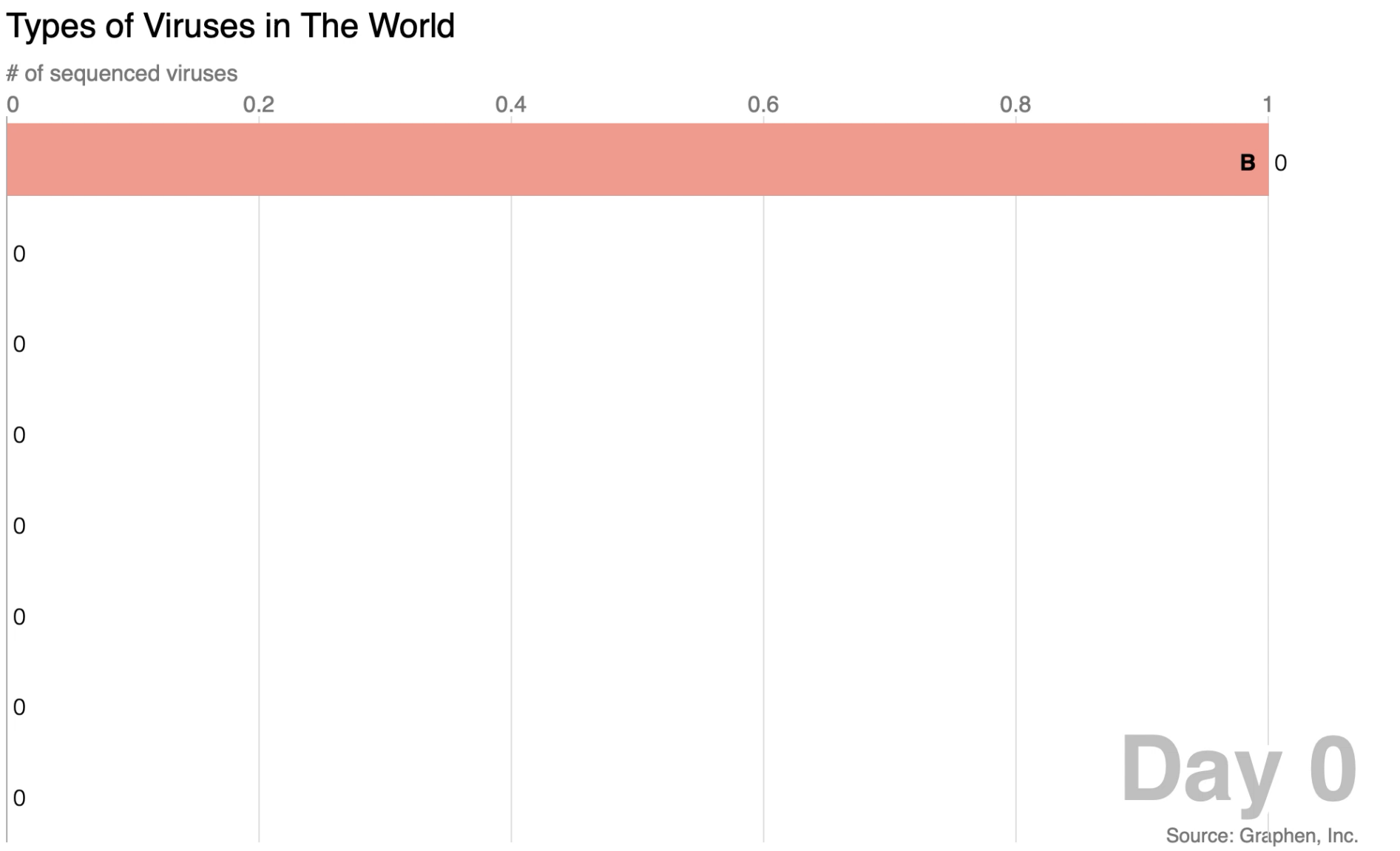

汉兴街道
3rd Ring Rd
2nd Ring Rd
Jiangfa Rd
TANGJIADUN RESIDENTIAL DISTRICT
汉口
Technology Building
科技大厦
Huangxiaohe Rd
Wuhan Ave
HUAQIAO RESIDENTIAL DISTRICT
花桥街道
JIANGHAN DISTRICT
江汉区
Qingnian Rd
JIANG'AN DISTRICT
江岸区
XIMA RESIDENTIAL DISTRICT
西马街道
Wuhan Centres for Disease Prevention &...
Wuhan Jiang'an Center for Disease..
HANJIADUN RESIDENTIAL DISTRICT
韩家墩街道
QIAOKOU DISTRICT
硚口区
HANSHUIQIAO RESIDENTIAL DISTRICT
汉水桥街道
BAOFENG RESIDENTIAL DISTRICT
宝丰街道
SHUITA RESIDENTIAL DISTRICT
水塔街道
NATIONALITY RESIDENTIAL DISTRICT
民族街道
XINHE RESIDENTIAL DISTRICT
新河街道
YONGFENG RESIDENTIAL DISTRICT
永丰街道
YUEHU RESIDENTIAL DISTRICT
HANZHENG RESIDENTIAL DISTRICT
汉正街道
Qintai Ave
Yangtze
G107


汉口站
青山区
江汉区
江岸区
硚口区
汉阳区
武昌区
东湖生态旅游风景区
10
武昌站
洪山区
S40
武漢病毒所
P4實驗室
1
四环线
武深高速
江夏区
S11

Map of Wuhan

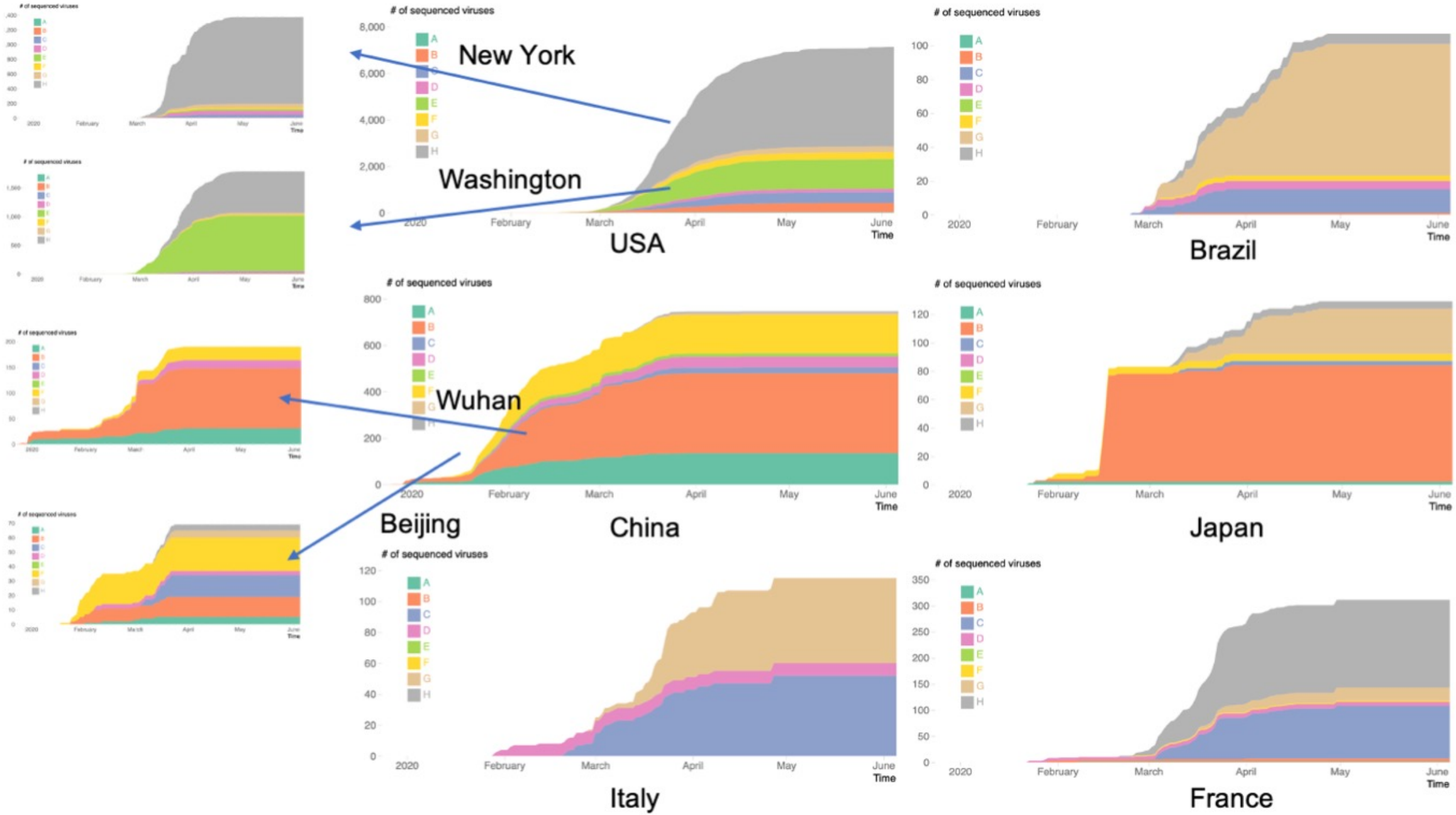


See any continent, country, region, and city at https://www.graphenai/covid/types.html

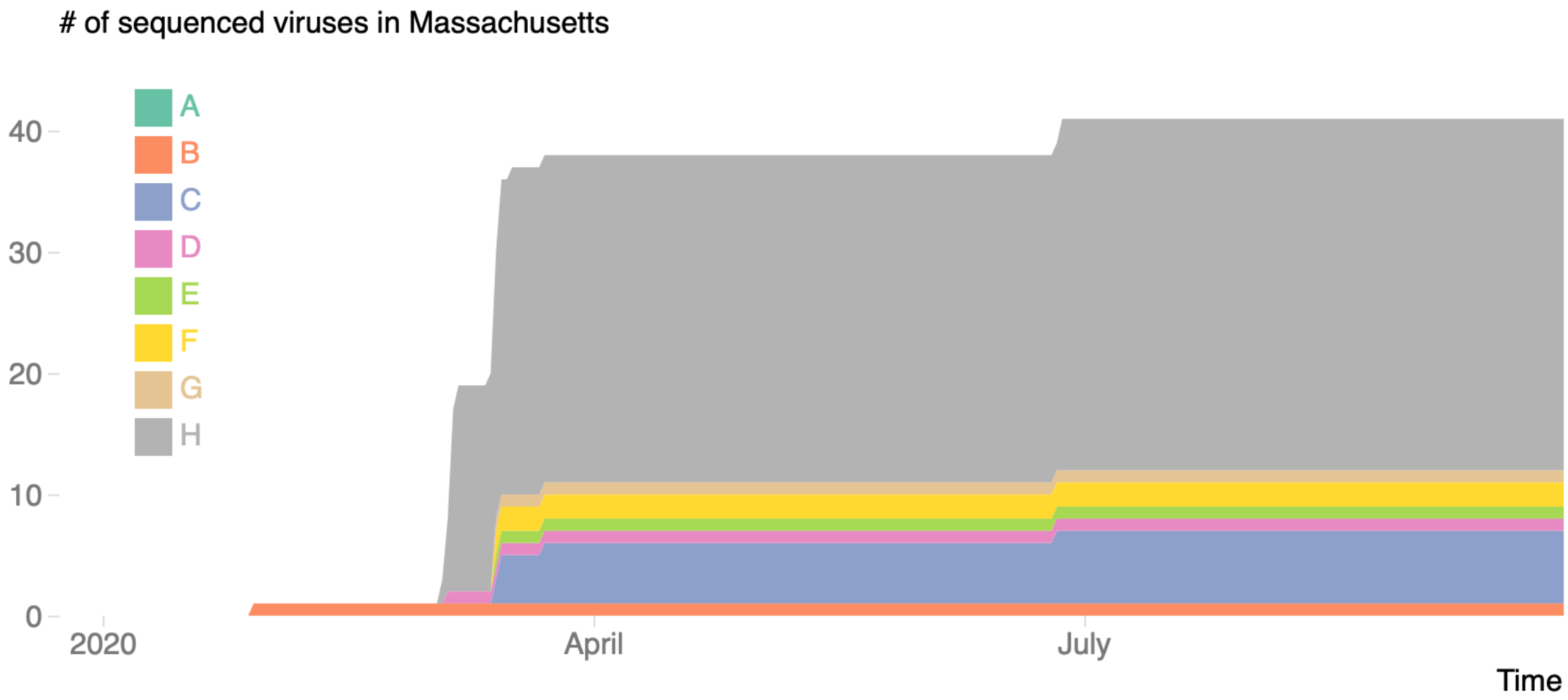


See any continent, country, region, and city at https://www.graphenai/covid/types.html

Timeline of Different Families in different countries



COVID-19 Virus Types in Massachusetts



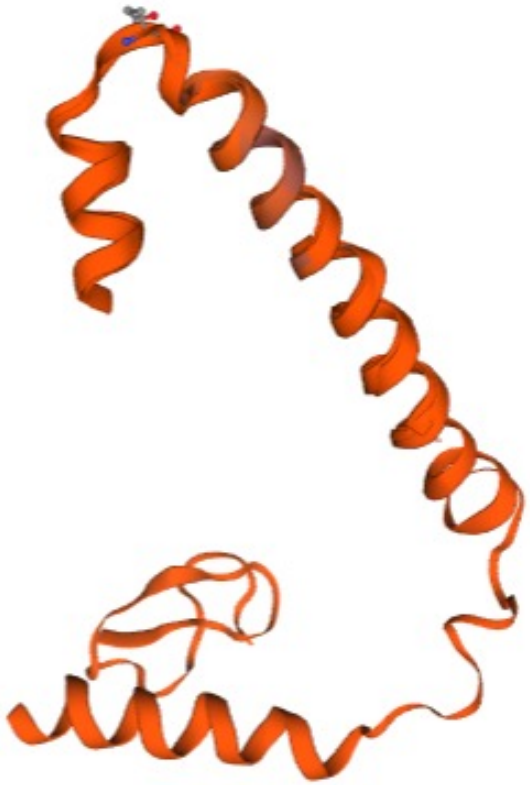
US is being affected by two families of SARS CoV-2

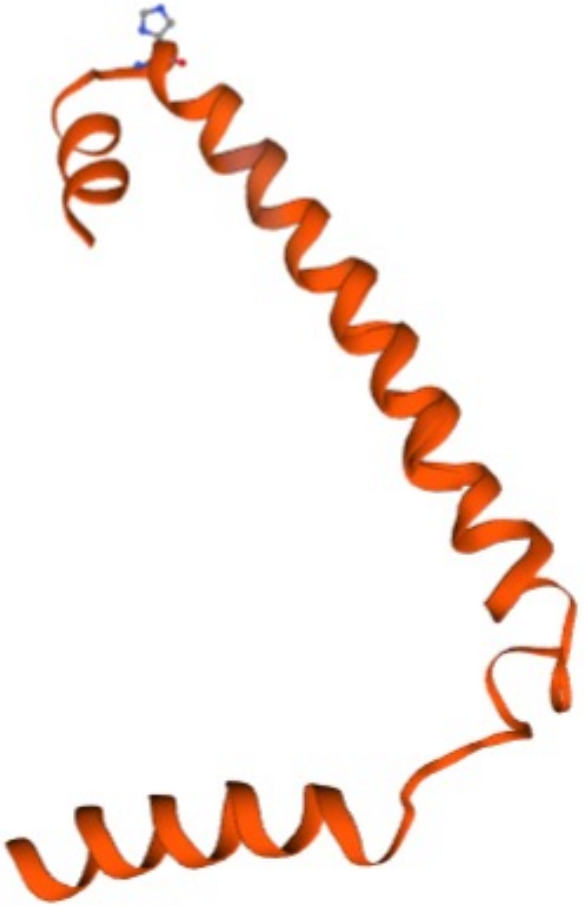
GRAPHEN
An Artificial Intelligence Company

Like this TV series – The Man in the High Castle

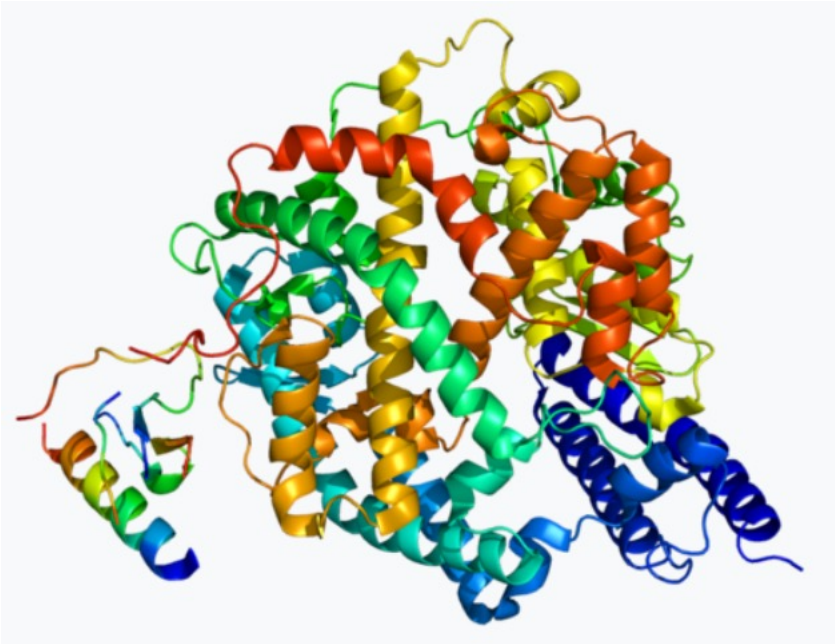
H-Strain Mutation: Escape Expert ORF3a Q57H

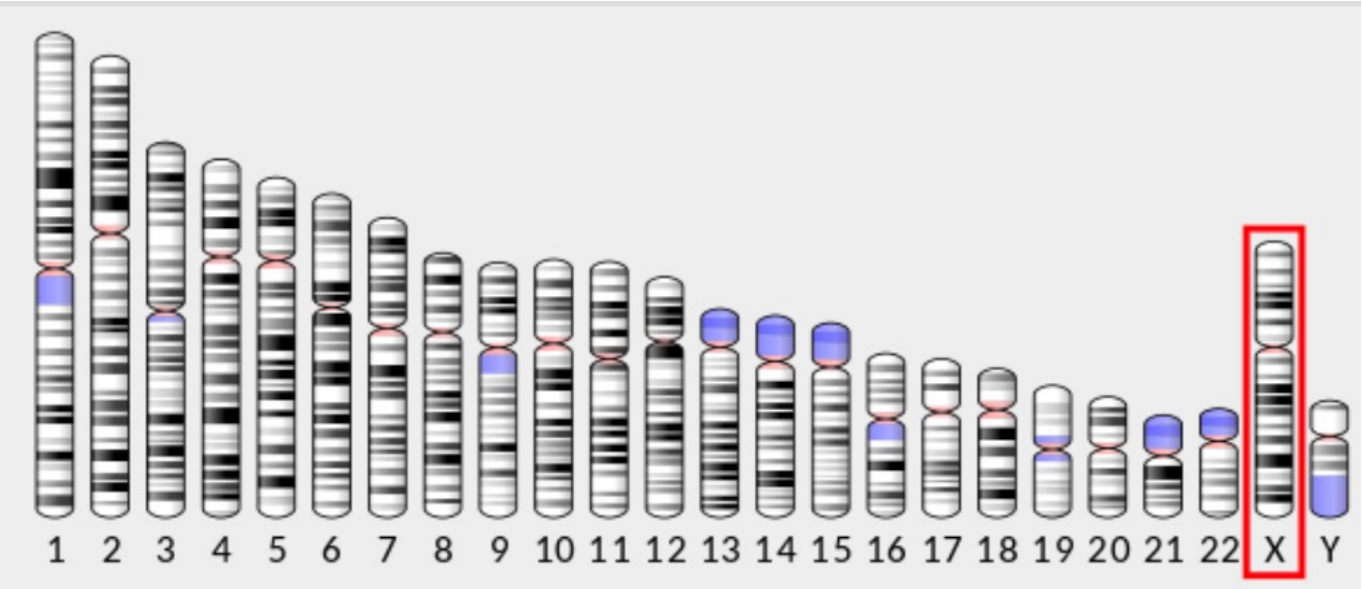
Overwhelmed East US and France

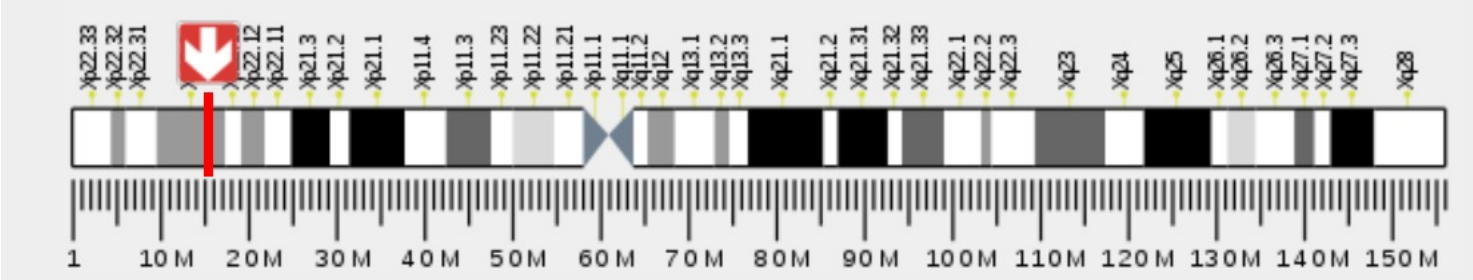




Human ACE2 Gene / Protein



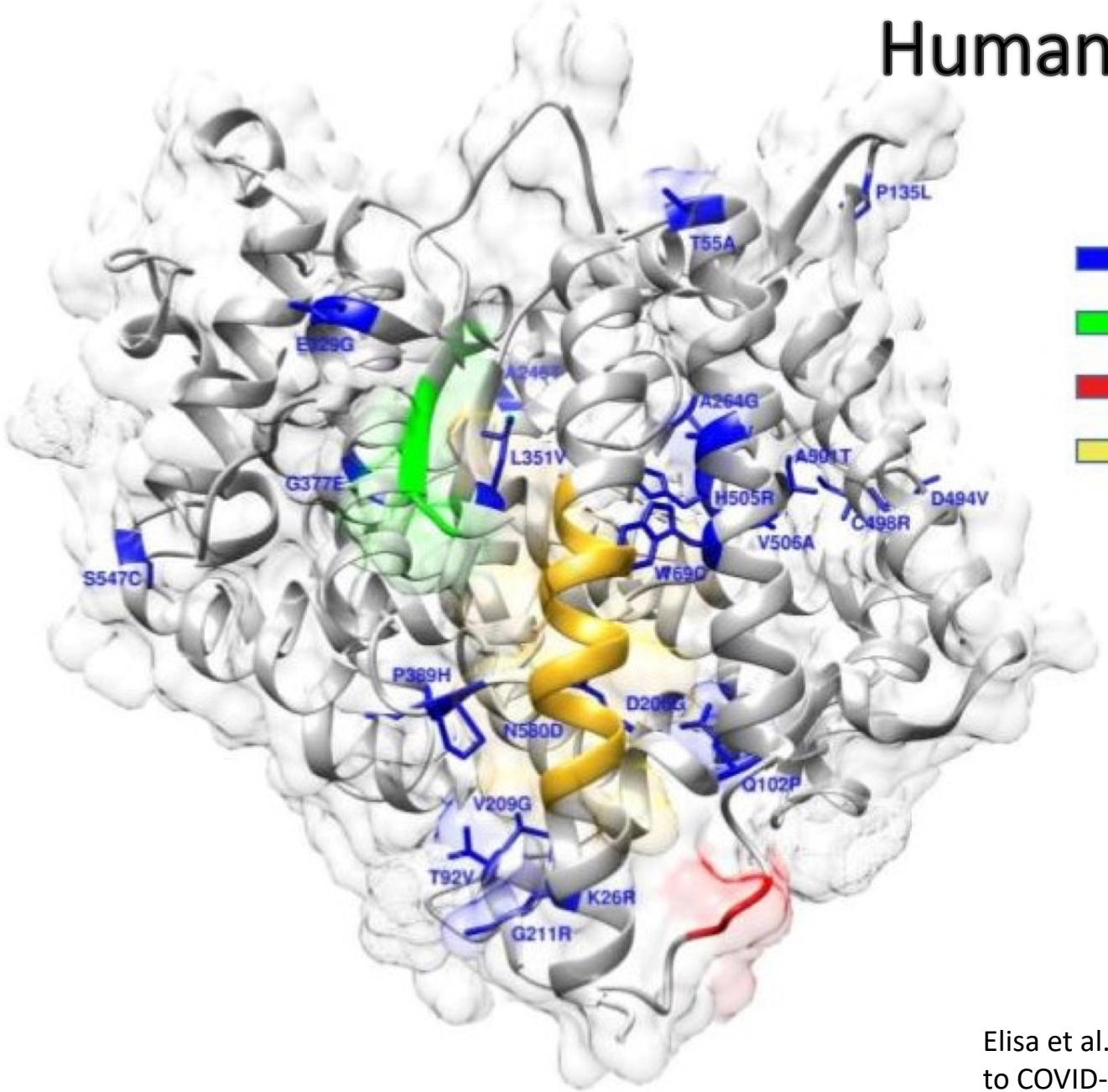




Xp22.1 15,561,033 bp to 15,602,148 bp 805 amino acids

Human ACE2 Mutations

GRAPHEN
An Artificial Intelligence Company



- Mutated positions
- ACE2 region aa (353-357)
- ACE2 region aa (82-84)
- ACE2 region aa (30-41)

Key receptor binding areas in ACE2: 31, 41, 82-84, and 353-357

Li et al, Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2, EMBO J. 2005

Mutations predicted to affect binding via COVID-19 patient study in Italy: L351V and P389H

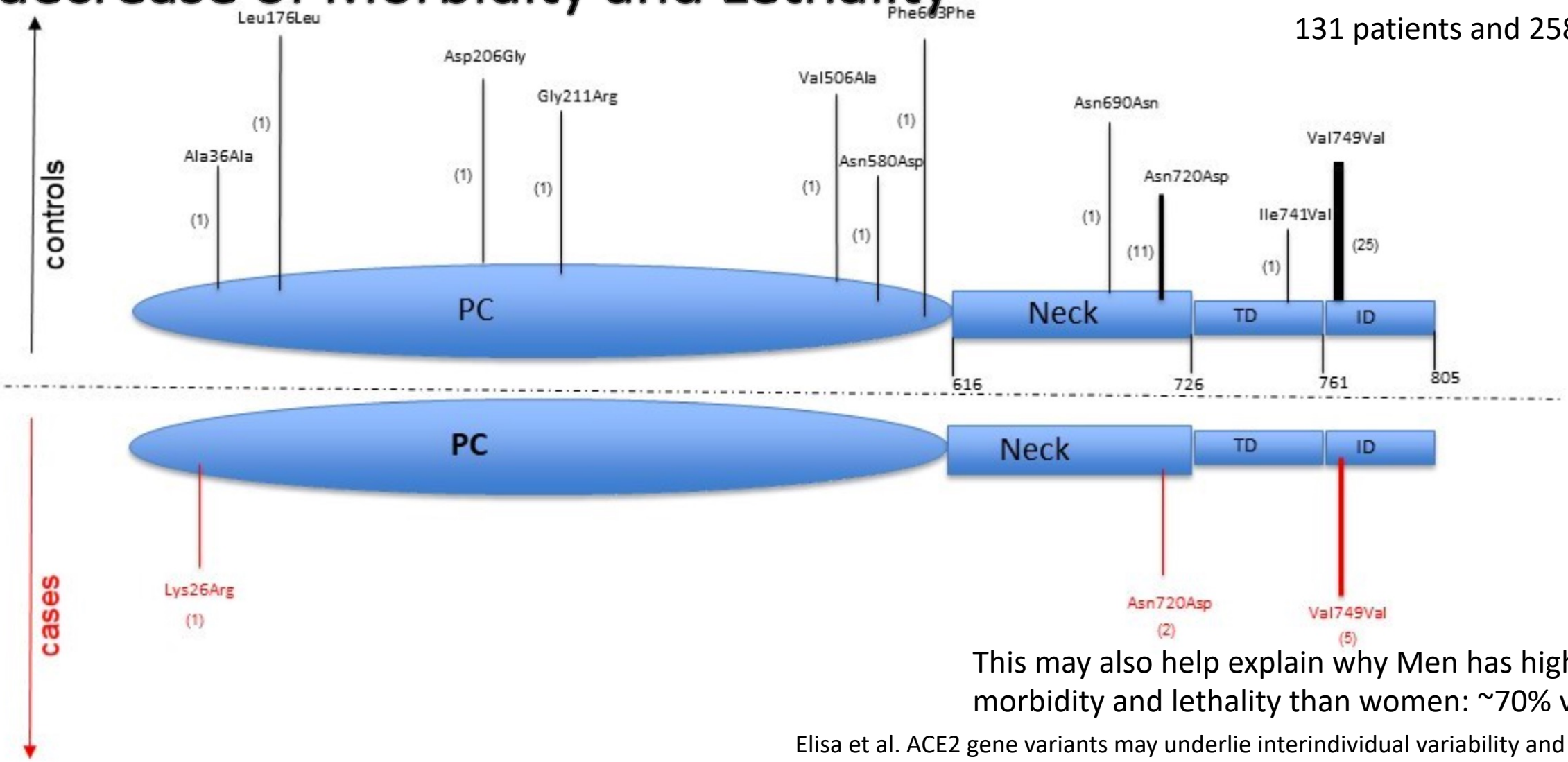
131 patients and 258 controls

Elisa et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population, European Journal of Human Genetics, July 2020

Variants in Human ACE2 may contribute to the decrease of Morbidity and Lethality

GRAPHEN
An Artificial Intelligence Company

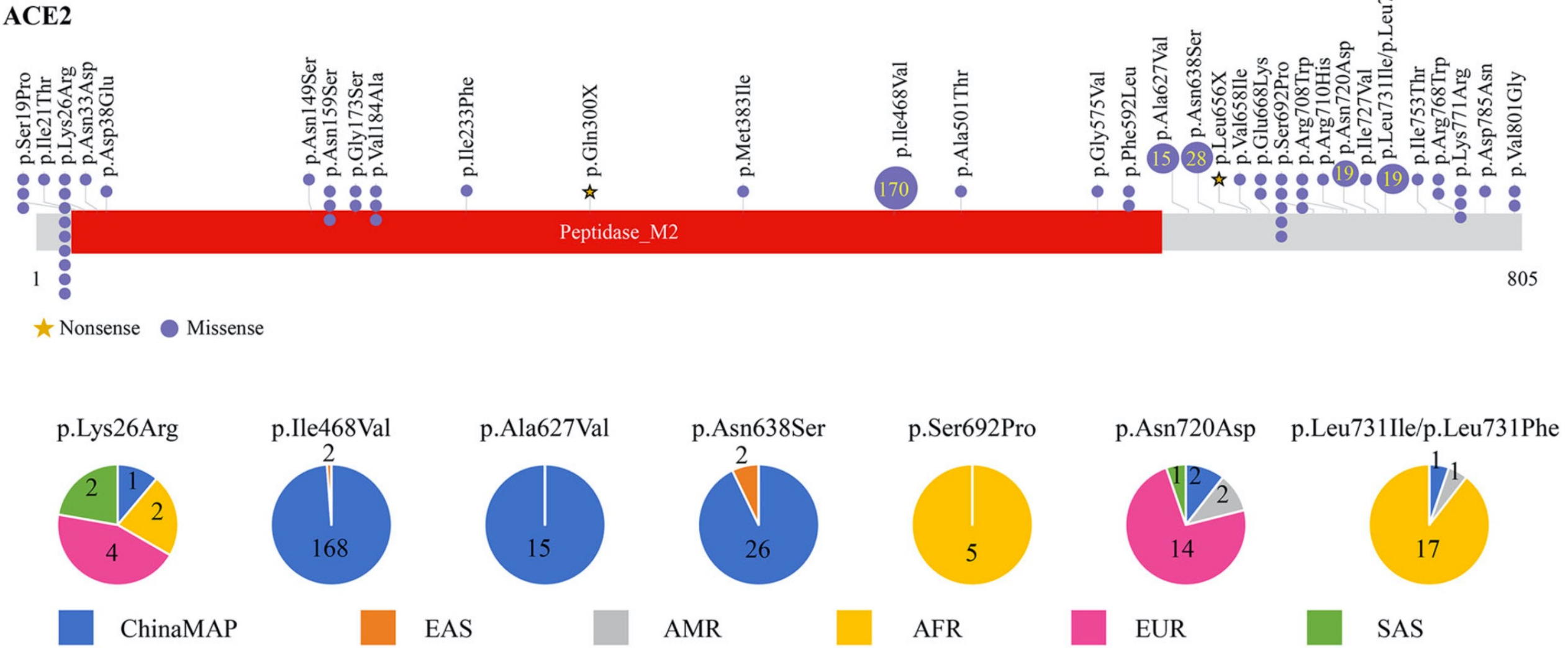
131 patients and 258 controls



This may also help explain why Men has higher morbidity and lethality than women: ~70% vs 30%

Elisa et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population, European Journal of Human Genetics, July 2020

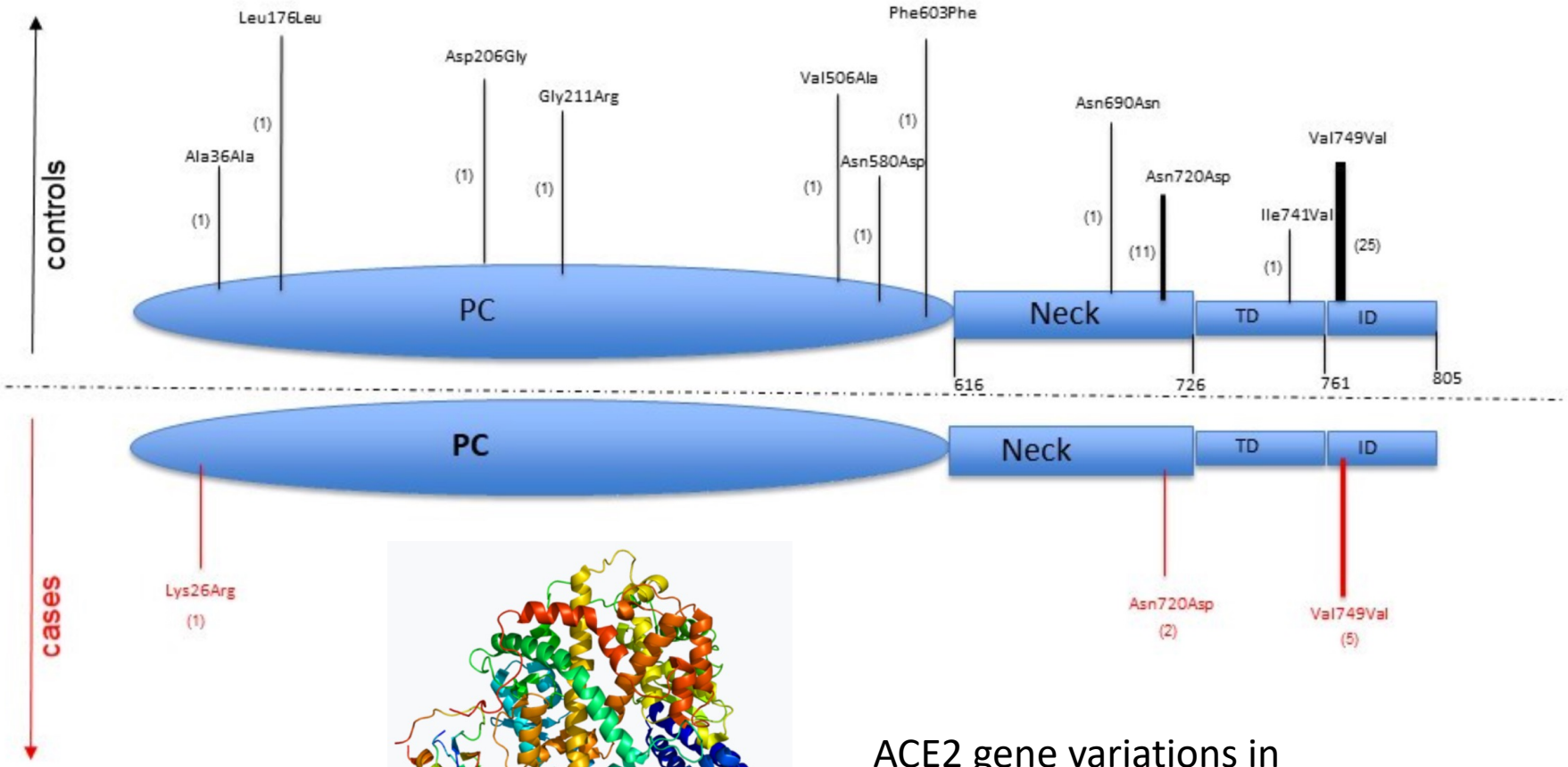
Common ACE2 Variants in different races



Cao et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations, Cell Discovery ,June 2020

Ongoing Work – Human Whole Genome Sequencing service for personal vulnerability detection

GRAPHEN
An Artificial Intelligence Company



ACE2 gene variations in Chromosome 22

Neanderthal gene in Chromosome 3

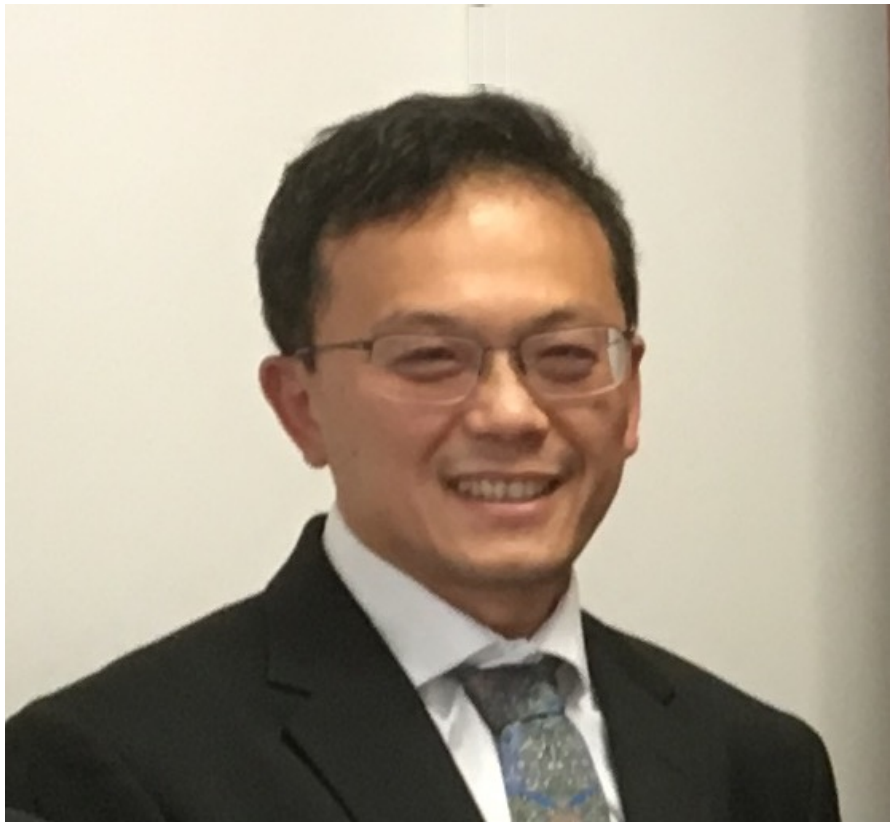
What does this mean?

You have <2% Neanderthal DNA

You inherited a small amount of DNA from your Neanderthal ancestors. Out of the 7,462 variants we tested, we found **297 variants** in your DNA that trace back to the Neanderthals.

All together, your Neanderthal ancestry accounts for less than **~2 percent of your DNA.**

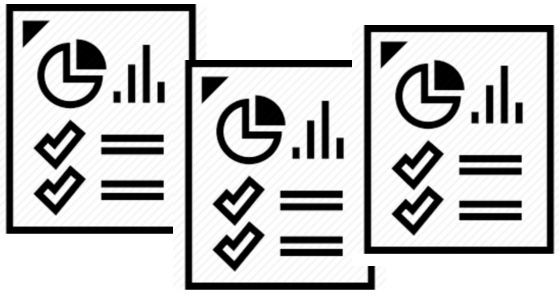
CY's ACE2 Variants



1. X,15582609,T,C,2,CC,ACE2,chrX,15579156,15620271
2. X,15582966,G,A,2,AA,ACE2,chrX,15579156,15620271
3. X,15583904,C,G,2,GG,ACE2,chrX,15579156,15620271
4. X,15586448,A,G,2,GG,ACE2,chrX,15579156,15620271
5. X,15586964,C,T,2,TT,ACE2,chrX,15579156,15620271
6. X,15588271,C,G,2,GG,ACE2,chrX,15579156,15620271
7. X,15590807,C,G,2,GG,ACE2,chrX,15579156,15620271
8. X,15597043,A,T,2,TT,ACE2,chrX,15579156,15620271
9. X,15597835,C,T,2,TT,ACE2,chrX,15579156,15620271
10. X,15598024,A,G,2,GG,ACE2,chrX,15579156,15620271
11. X,15600215,A,C,2,CC,ACE2,chrX,15579156,15620271
12. X,15600691,C,T,2,TT,ACE2,chrX,15579156,15620271
13. X,15600744,T,A,2,AA,ACE2,chrX,15579156,15620271
14. X,15601274,G,T,2,TT,ACE2,chrX,15579156,15620271
15. X,15601719,T,C,2,CC,ACE2,chrX,15579156,15620271
16. X,15603064,G,C,2,CC,ACE2,chrX,15579156,15620271
17. X,15604865,G,A,2,AA,ACE2,chrX,15579156,15620271
18. X,15607650,T,C,2,CC,ACE2,chrX,15579156,15620271
19. X,15608386,C,T,2,TT,ACE2,chrX,15579156,15620271
20. X,15608499,T,G,2,GG,ACE2,chrX,15579156,15620271
21. X,15608952,C,T,2,TT,ACE2,chrX,15579156,15620271
22. X,15610348,C,T,2,TT,ACE2,chrX,15579156,15620271
23. X,15614145,G,A,2,AA,ACE2,chrX,15579156,15620271
24. X,15615453,A,G,2,GG,ACE2,chrX,15579156,15620271
25. X,15616796,T,C,2,CC,ACE2,chrX,15579156,15620271
26. X,15617736,G,A,2,AA,ACE2,chrX,15579156,15620271
27. X,15618061,A,G,2,GG,ACE2,chrX,15579156,15620271
28. X,15618063,T,C,2,CC,ACE2,chrX,15579156,15620271

Predictive Health through Whole Genome Sequencing

GRAPHEN
An Artificial Intelligence Company



Knowing your 6.4B nucleotides. Once in a lifetime

150+ potential diseases risk reports and more…

Cancers

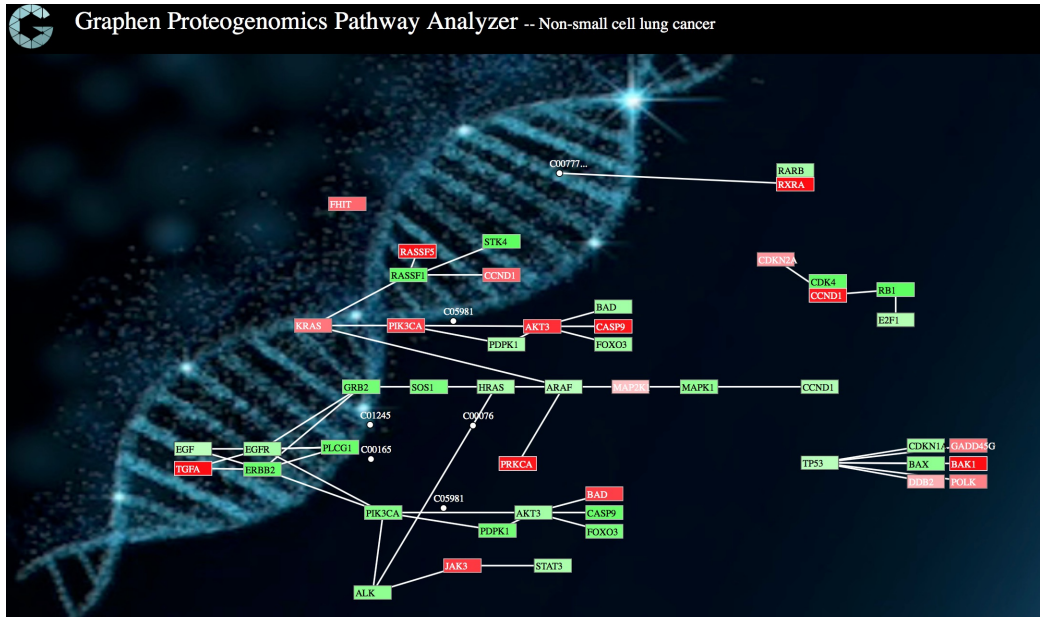
Immune Diseases

Neurodegnrative Diseases

Cardiovascular Diseases

Digestive System Diseases

Muscle-related Diseases, tec.



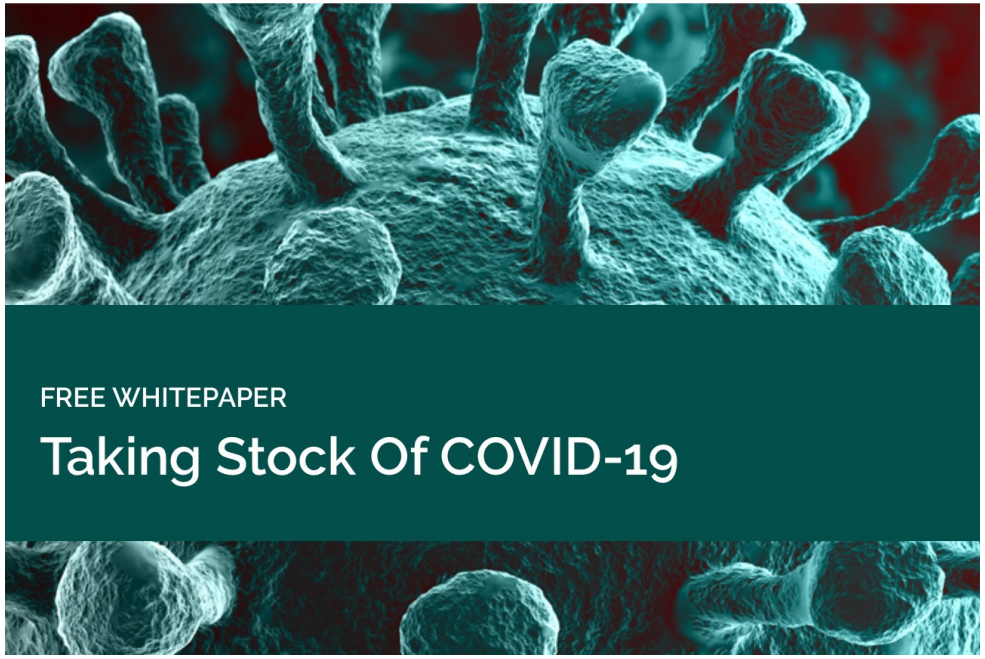
When new situations happen, e.g., COVID-19's new knowledge is available, a pre-sequenced individual could get analytics report immediately

=> A Medical Challenge as well as Big Data and AI Challenge

Service shall start at 4Q2020 or 1Q2021

COVID-19 Pandemic Impact: Global R&D Spending For AI Medical will be 4x to more than US$2 Billion By 2025

ABIresearch
for visionaries

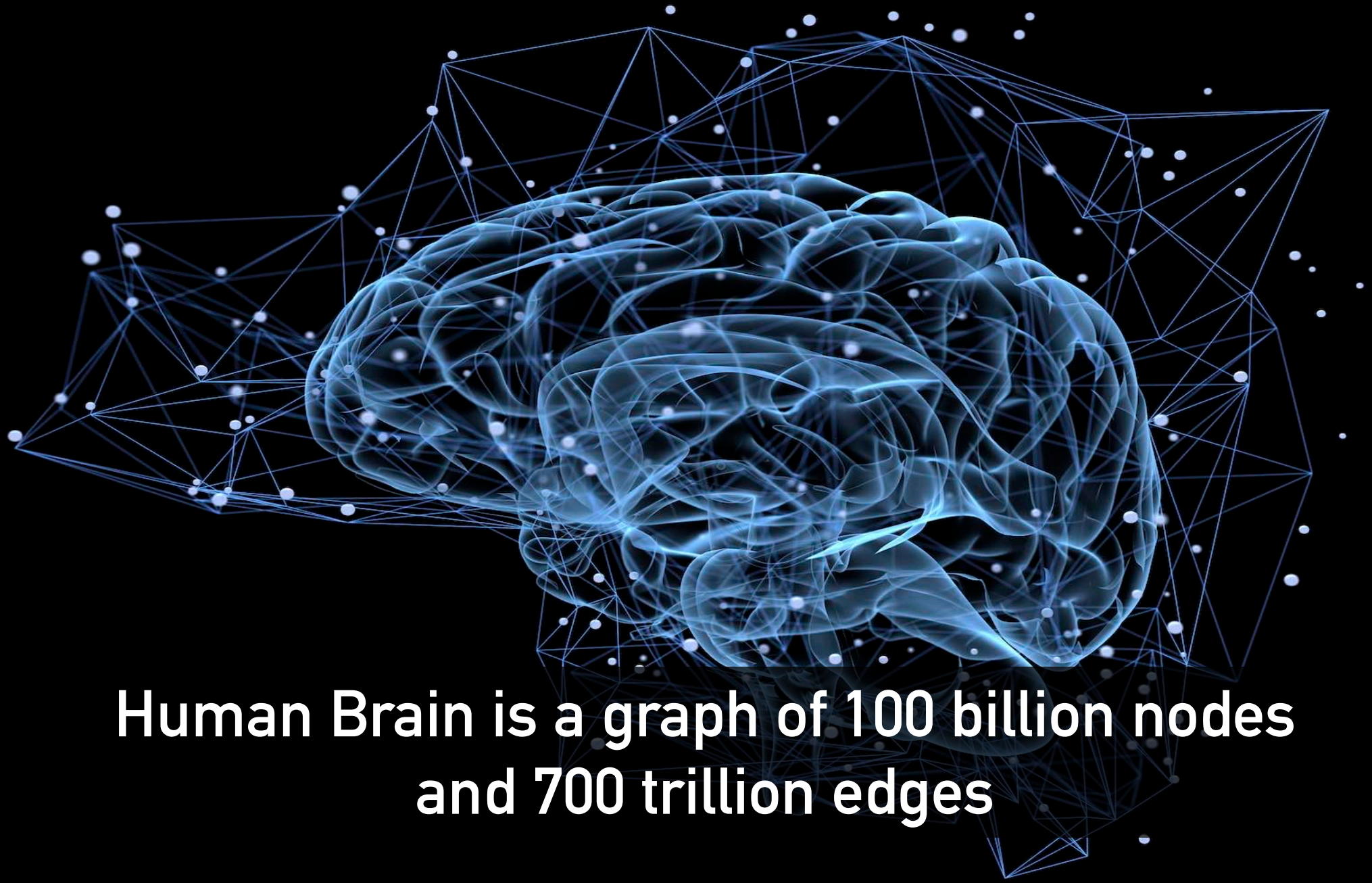


*Several companies, including **Alibaba, YITU, Graphen**, and **Google** DeepMind, are already developing AI tools to help:*

- *detect the virus,*
- *diagnose its evolution,*
- *track its geographical footprint, project its future, and*
- *even predict its potential protein structure to find a vaccine for it."*

COVID-19 Pandemic Impact: Global R&D Spending For AI Medical will be 4x to more than US$2 Billion By 2025

*Tools from established companies like **Google** DeepMind, startups like **Graphen**, and AI chipsets from vendors like **NVIDIA** and **Intel** will help accelerate the speed of drug discovery, development, and testing, allowing pharmaceutical companies and healthcare authorities to combat the pandemic.*



Human Brain is a graph of 100 billion nodes and 700 trillion edges

Graphen Core -- Full-Brain AI Platform

GRAPHEN
An Artificial Intelligence Company

Developing advanced Graph Computing technology to realize a unique novel AI platform that simulates the whole brain functions and powers novel industry solutions.

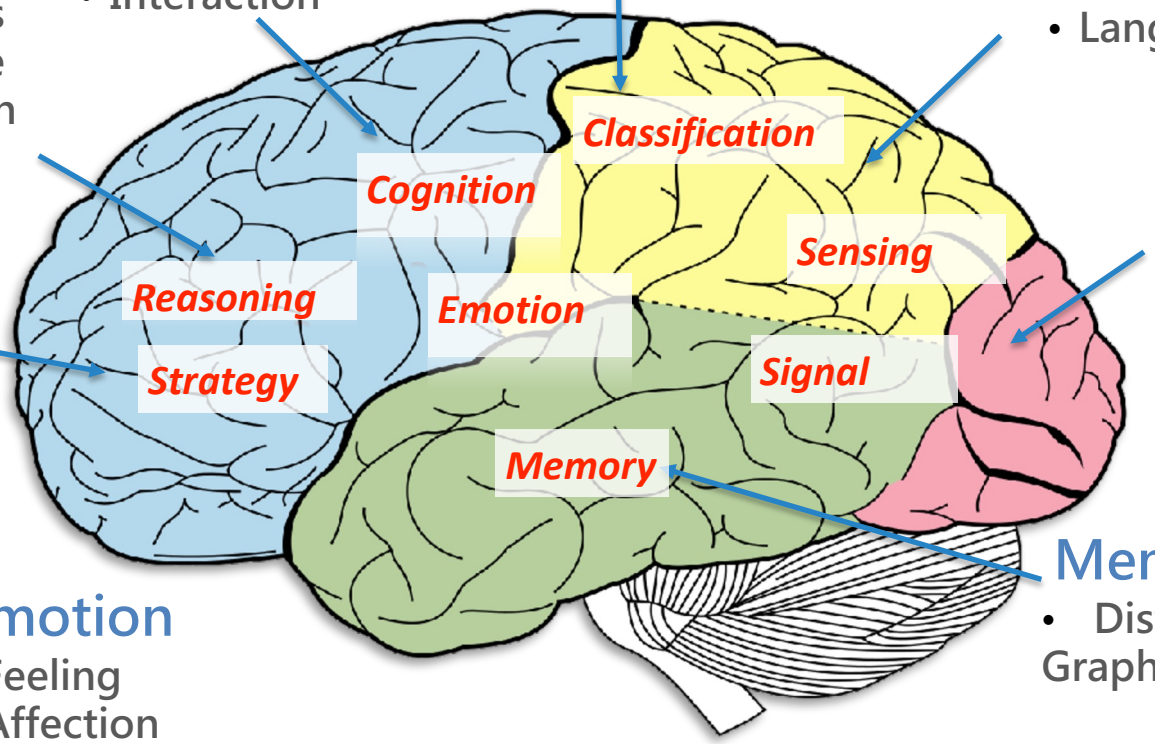
Ardi AI Platform

Learning
- Machine Learning and Deep Learning
- Autonomous Learning

Expression
- Big Data Visualization
- Explainability
- Interaction

Reasoning
- Bayesian Networks
- Causality Inference
- Behavior Prediction

Understanding
- Deep Vision Understanding
- Language Understanding



Perception
- Network Analysis
- Feature Analysis

Strategy
- Game Theory
- Optimal Solution

Personality
- Humanization
- Ethics

Knowledge
- Q&A, Conversation

Memory
- Distributed Native Graph Database

Pipeline
- Coordination
- Streamline Flow

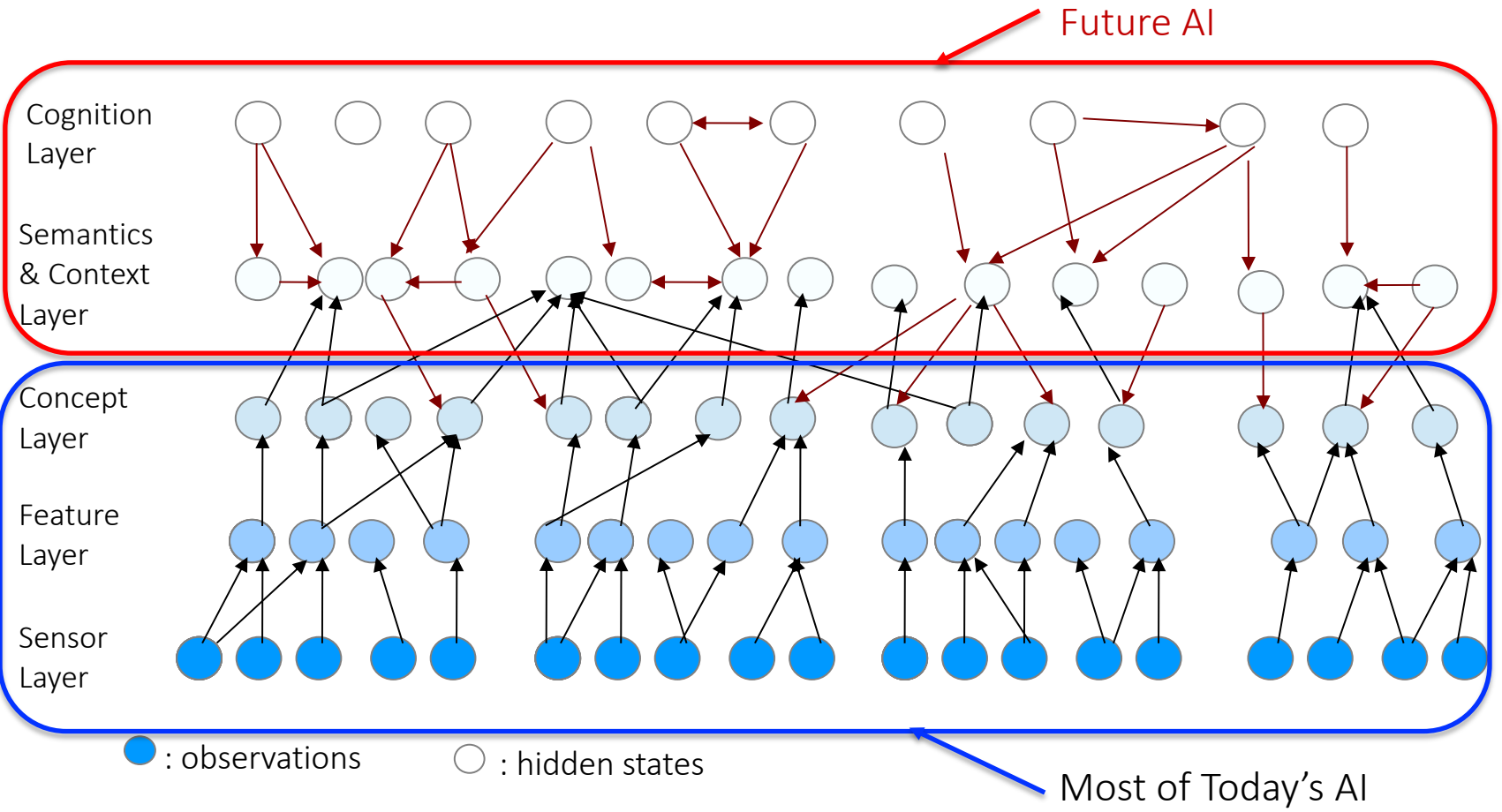
Action
- Control
- Behavior

Emotion
- Feeling
- Affection

https://www.graphen.ai/products/ardi_feature.html

Evolution of AI – Ardi represents the Future AI

GRAPHEN
An Artificial Intelligence Company



Today's "AI"	Future AI
Processing	Perception
Classification	Reasoning
	Strategy
Individuals	Context
Built on Open Sources	Built on Graph Computing

Machine Learning / Deep Learning targets on "Concept Classification" layer, which is only one of many full brain functions!!

Graphen Medical

GRAPHEN
An Artificial Intelligence Company

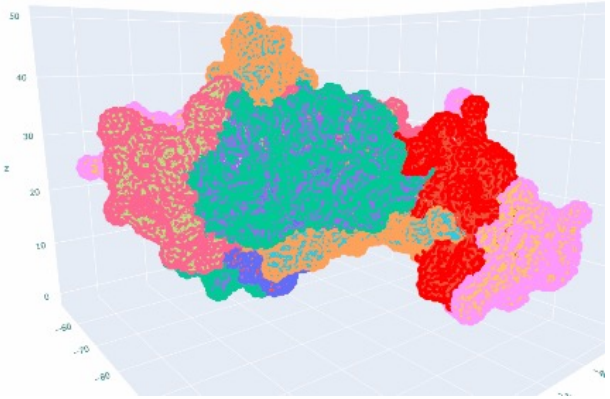
Each Person knows Biological Digitized-Self, with Personalized Precision Treatment and Drugs

-- Graphen's Vision to Help Making Future of Medicine

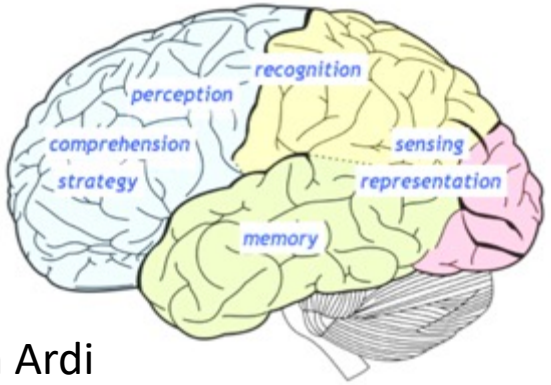
Graphen Drugomics

Graphen Genomics

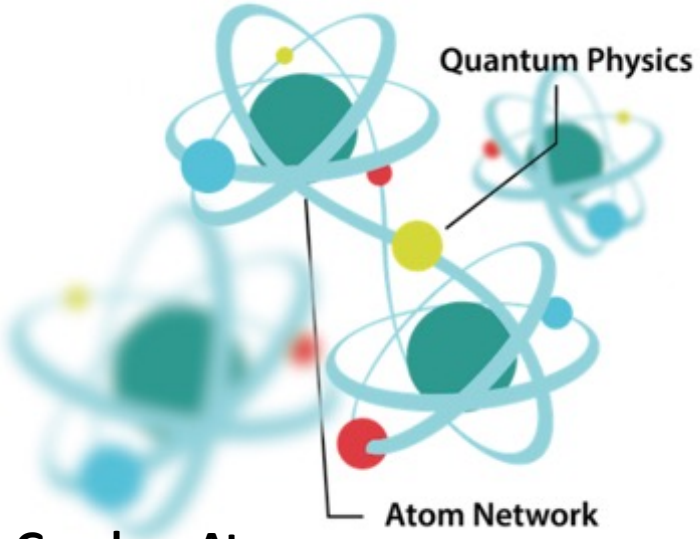
Life is composed of graphs of atoms



https://www.graphen.ai/products/medical.html

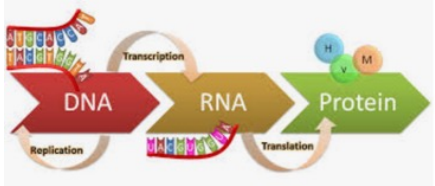


Graphen Ardi

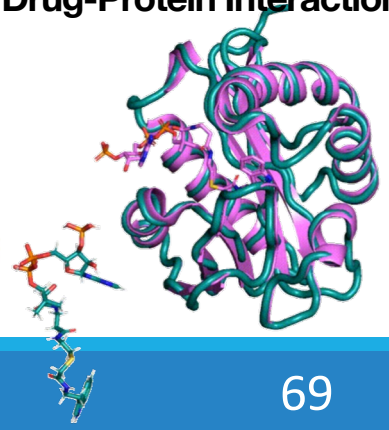


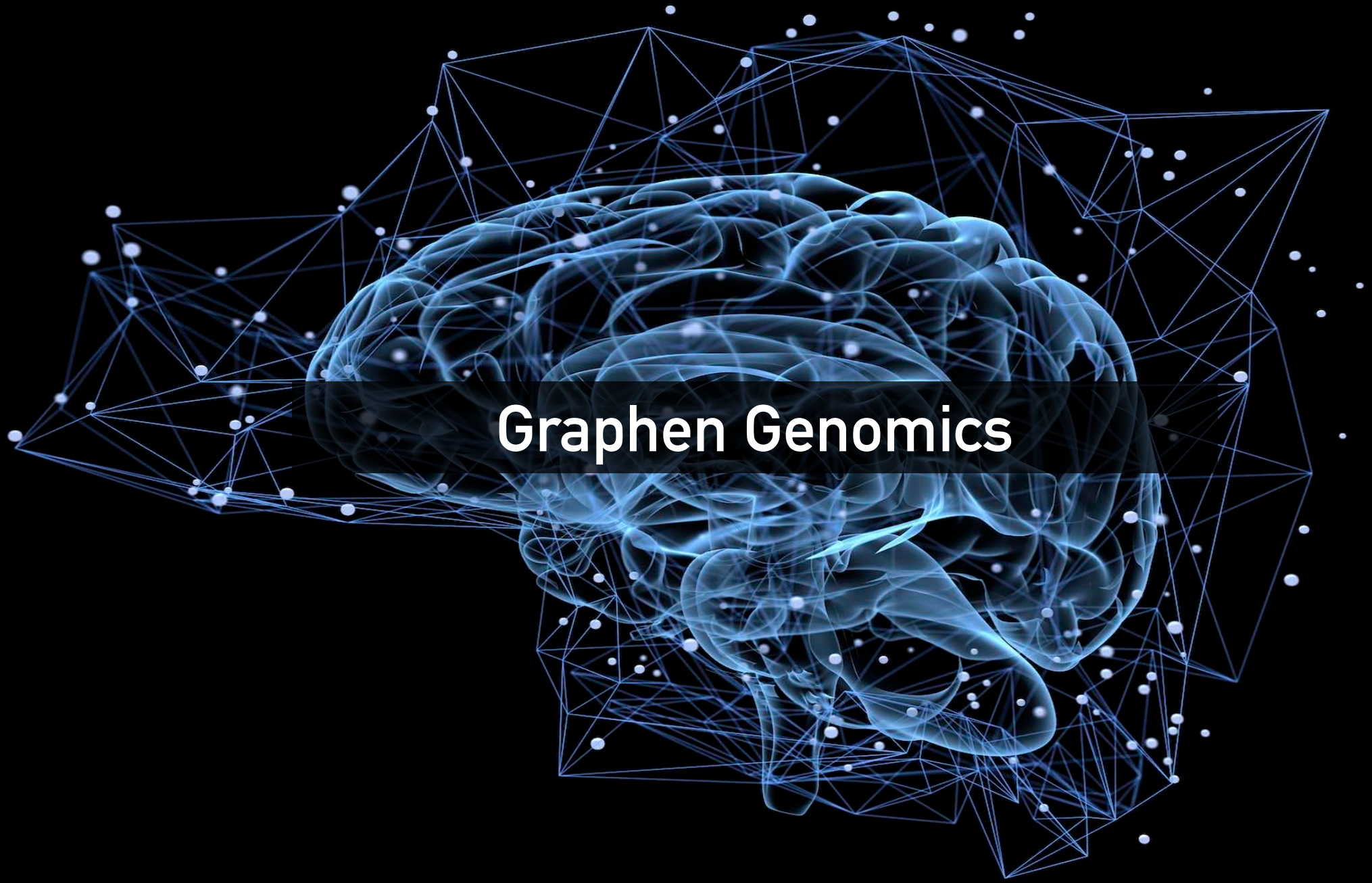
Graphen Atom

Central Dogma of Biology



Drug-Protein Interaction





Graphen Genomics

Graphen Genomics

Home Products COVID-19 About Graphen

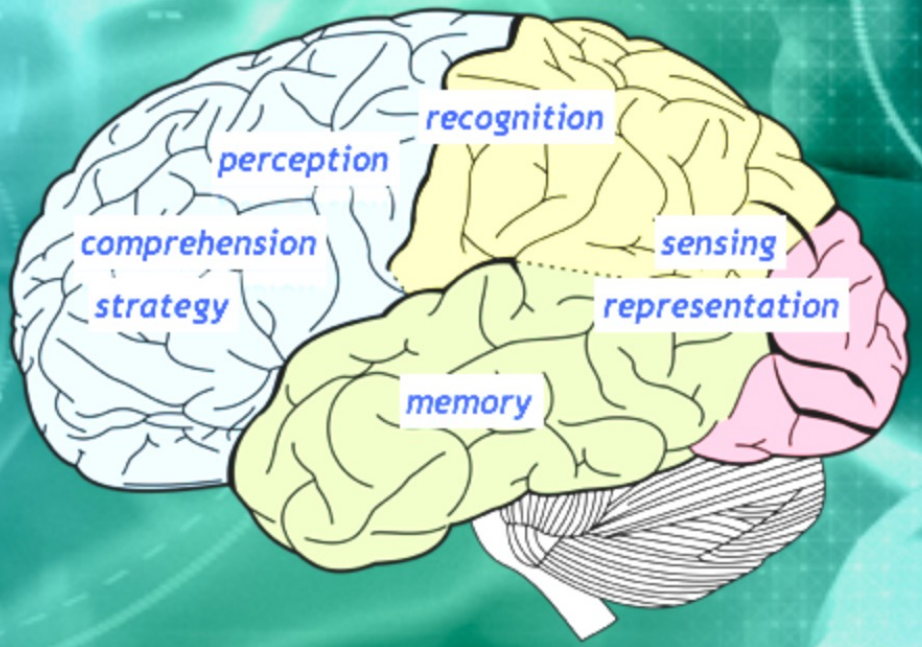
What am I?

What made me genetically? How am I doing today? What can happen in the future?

Congenital Me

Today's Me

Future Me



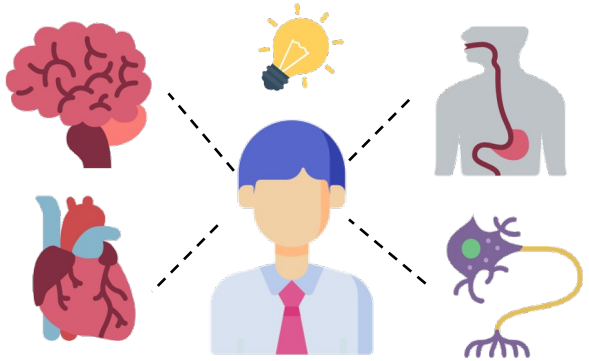
Graphen Whole Genome Analysis

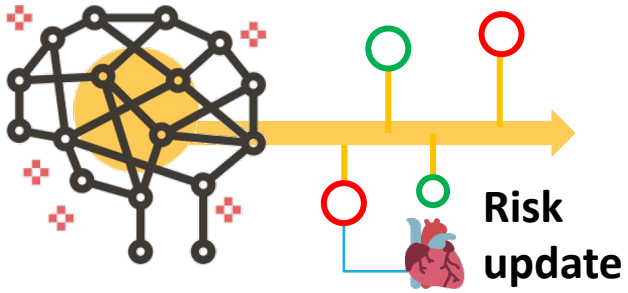
Do you want to know yourself? What does your blueprint say about you? Graphen Personal Whole Genome Analytics System analyzes your entire 6.4B genome. It provides your risk likehood of 350+ diseases in 10 major categories:

- **Tumor**
- **Cardiovascular and Immunity Diseases**
- **Respiratory System Disorder**
- **Digestive System and Metabolic Disorder**
- **Eye, Ear and Mastoid Disorder**
- **Infectious and Parasitic Diseases**
- **Genitourinary System and Maternal Diseases**
- **Skin and Musculoskeletal Disorder**
- **Neurological Disorder**
- **External Causes and Other Diseases**

Graphen Whole Genome Product Features

GRAPHEN
An Artificial Intelligence Company





Personal digitization:
Starts with your 3 billion DNA pairs

By testing the entire, 100% innate genetic composition, Graphen helps customers understand their genetic profile. Combine the research and development of biomedicine to understand yourself. Other testing products in the market, only analyze 0.1% to 2% of gene base pairs, which means one need to be tested again every time it is needed - time-consuming and lots of hassle. Graphen Whole genetic testing is completely digital, just one test and you get to know all your 3 billion DNA pairs.

Single test, focus on the future:
Know all your disease risks

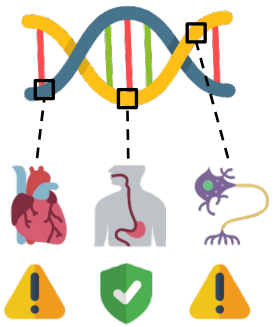
Graphen genetic testing only needs to be done once in your life to understand the risks of over 350 diseases. The test results show clear scientific evidence at present and explain the 10 major diseases including cardiovascular and immune diseases, cancer, digestive system diseases, nervous system diseases, etc. Test results provide personal risk factors to help customers grasp the distance to the disease and predict the need for follow-up exams.

Advanced AI technology:
Risk control, life-long renewal

Artificial intelligence has the ability to automatically read a large number of documents and judge the credibility. It makes it easy and efficient conduct our risk assessments and keep up with the latest medical research results. In addition to quantifying the risks of various diseases, customers only need to test once to subscribe to risk assessment updates. Risk control, life-long renewal.

Graphen Whole Genome Product Advantages

GRAPHEN
An Artificial Intelligence Company



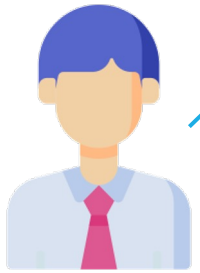
Zero pre-disease risk detection

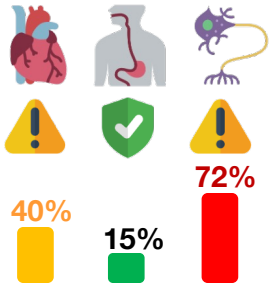
Full genetic testing allows you to know the risk of possible future illness before you become ill, and to prevent it early and accurately.

350+
Diseases

350+ diseases test completed at once

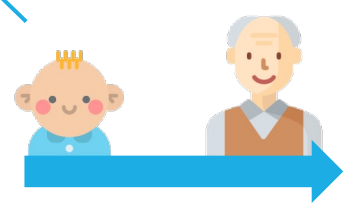
Without complicated testing items and procedures, you can understand 350+ possible disease risks in advance.





Multi-faceted disease risk quantitative assessment

Multi-index quantification of disease-gene risk, giving priority to assessment of the degree of urgency, early prevention and no panic .

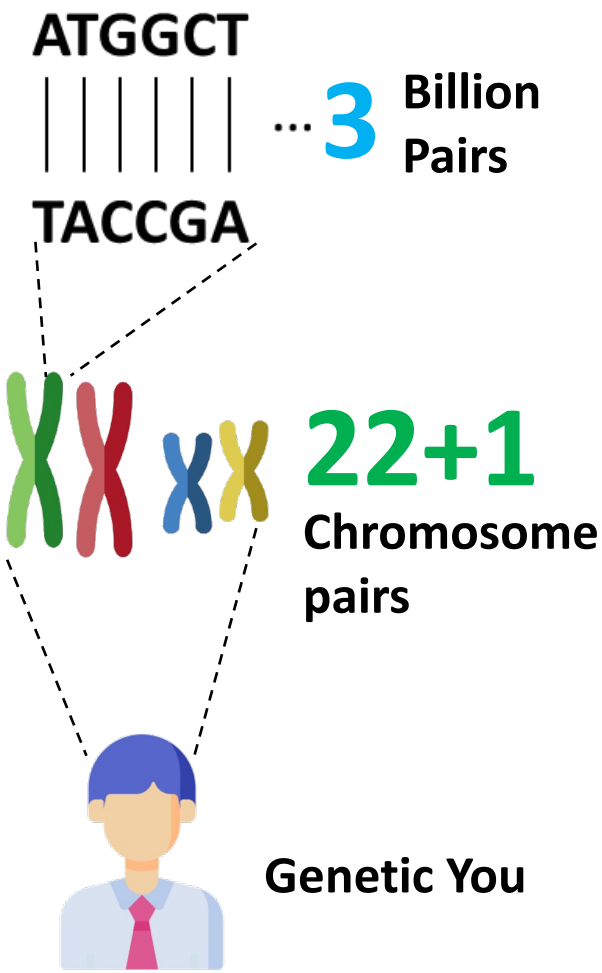


Report Update

Lifetime updates of risk assessment

As long as the test is done once, the future new risk indicators do not need to be tested again, and your risk assessment is continuously updated.

Whole Genome Sequencing



Genes dominate the physiological development of human life. We form our own unique genetic makeup by inheriting the genes of our parents. In addition to the acquired environment and other factors, humans can be said to be composed of up to 3 billion base pair "codes", and everyone's "codes" are slightly different.

In addition to making each person an independent individual, these small differences may also make us innately prone to certain diseases, which we used to call "physique" and can be understood through "family medical history." And now we can understand the potential risk of disease from a person's genetic information. Such as genetic diseases, diabetes, cardiovascular diseases, cancer, etc., by digitizing genetic information, we have the opportunity to understand the current process before the disease develops, and prevent and control it early, so as to regain the possibility of future health.

10 categories of 350+ diseases-a comprehensive understanding of your disease risk

GRAPHEN
An Artificial Intelligence Company

With only one full genetic test, Graphen can analyze more than 350 disease risks and summarize your Digital Risk Profile in ten categories:

- Respiratory diseases
- Nervous system disease
- Digestive system diseases
- Cardiovascular and immune diseases
- Cancer, etc.

Let customers fully understand the possible future diseases, as well as the doubts about the past family medical history, and immediately take response measures and adjustment strategies.

Disease risk quantification and accurate assessment of multi-geographical races

GRAPHEN
An Artificial Intelligence Company

Display Language: 中文

Select Race: East Asian

Low Risk (0 - 30) | Mid Risk (30 - 60) | High Risk (60 - 100)

Potential Disease	潛在疾病	Category	疾病類別	Known Risky Alleles	Your Matched Alleles	Allele Match Ratio	Risk Score Total	Risk Score East Asian⇩
age-related macular degeneration	年齡相關性黃斑變性	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	63	19	30%	56.57	49.96
exfoliation syndrome	剝脫綜合徵	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	6	1	17%	41.25	34.53
wet macular degeneration	濕性黃斑變性	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	17	4	24%	27.38	22.08
vogt-koyanagi-harada disease	沃格特-小柳-原田病	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	4	1	25%	20.45	20.45
central serous retinopathy	中央性漿液性視網膜病	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	2	2	100%	19.85	20.16
low tension glaucoma	低張力青光眼	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	2	2	100%	8.61	9.23
open-angle glaucoma	開角型青光眼	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	83	35	42%	8.45	7.73

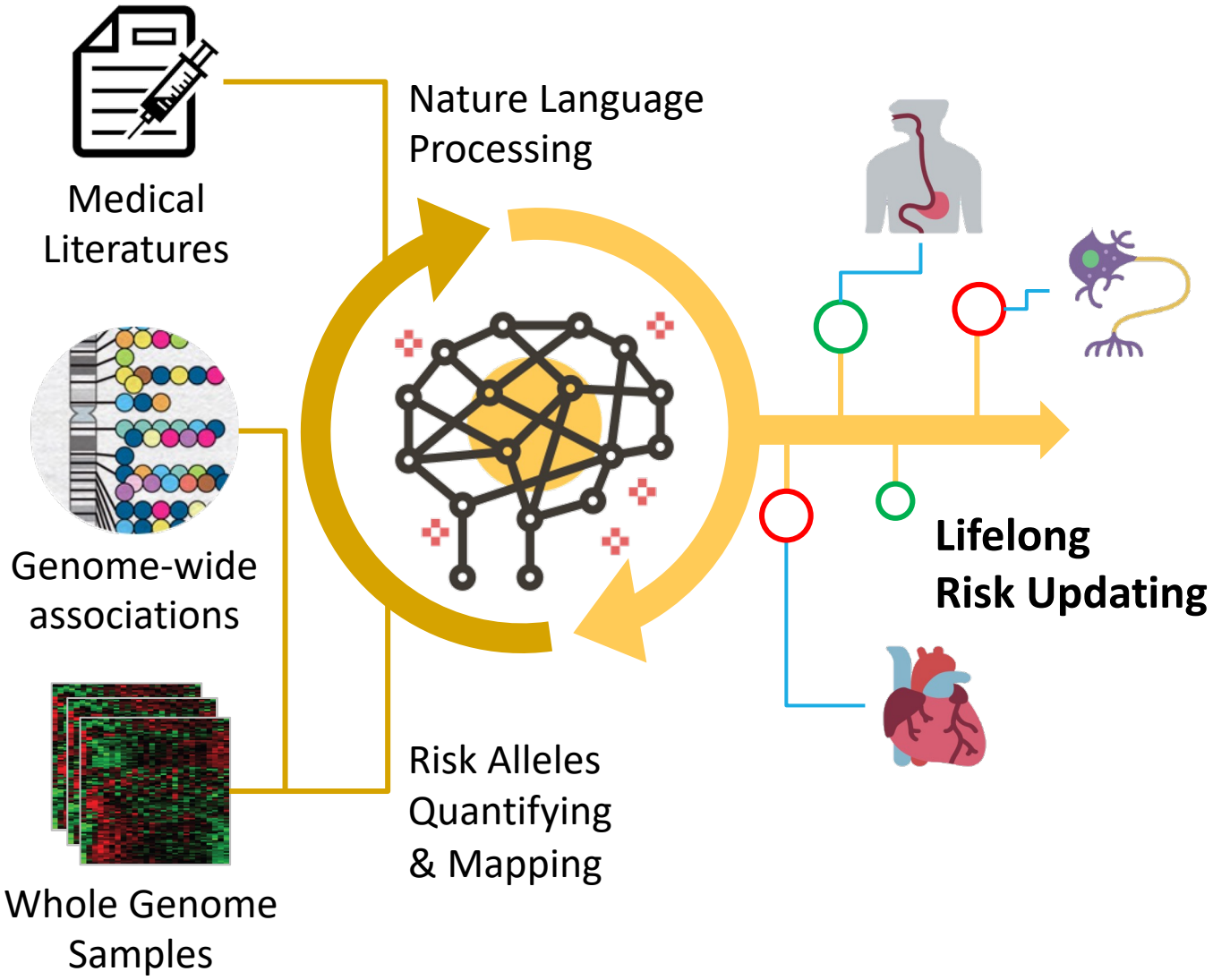
Graphen quantifies the risk between individuals and diseases through a number of indicators, indicating that the risks of different diseases are acute:

- Genome-wide Association (GWAS) statistical indicators
- Number of risk genetic coincidences
- Risk gene frequency
- Number of papers published on risk genes, IF, etc.

In addition to quantifying the degree of disease risk, Graphen can also assess the degree of risk based on race and geolocation, and apply accurate testing to multiple ethnic groups.

AI Empowered Lifetime of Disease Risk Management

GRAPHEN
An Artificial Intelligence Company



Graphen developed this genetic analysis technology based on artificial intelligence, allowing customers to subscribe to life-long disease risk assessment services by providing full genetic testing data once, and the customer's risk control is always up to date with the latest research results.

Technical content：

- Big data collection of gene-disease association and testing data
- Artificial intelligence technologies such as natural language processing fully support gene-disease research literature to be continuously updated
- Literature citation and influence index to establish gene-disease association confidence level

Graphen Genomics

Graphen Personal Whole Genome Analysis

Cardiovascular and Immunity Diseases

See your likelihood of cardiovascular and immunity diseases

Learn more...

Respiratory System Disorder

See your likelihood of repiratory system disorder

Learn more...

Digestive System and Metabolic Disorder

See your likelihood of digestive system and metabolic disorder

Learn more...

Welcome, CY!

Graphen Genomics

CY's Whole Genome Analysis Results

Display Language: English

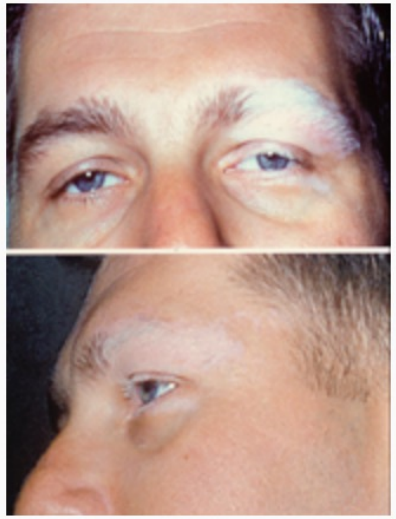
Select Race: Any

Low Risk (0 - 30) | Mid Risk (30 - 60) | High Risk (60 - 100)

Potential Disease	潛在疾病	Category	疾病類別	Known Risky Alleles	Your Matched Alleles	Allele Match Ratio	Risk Score Total
age-related macular degeneration	年齡相關性黃斑變性	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	63	19	30%	56.57
exfoliation syndrome	剝脫綜合徵	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	6	1	17%	41.25
wet macular degeneration	濕性黃斑變性	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	17	4	24%	27.38
vogt-koyanagi-harada disease	沃格特-小柳-原田病	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	4	1	25%	20.45
central serous retinopathy	中央性漿液性視網膜病	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	2	2	100%	19.85
low tension glaucoma	低張力青光眼	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	2	2	100%	8.61
open-angle glaucoma	開角型青光眼	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	83	35	42%	8.45

Display Langu

Vogt–Koyanagi–Harada病（VKH）是一種推測為自身免疫病的多系統疾病，會影響有黑色素的色素組織。最顯著的表現是雙眼瀰漫性葡萄膜炎，會影響眼睛。 VKH也可能會累及內耳，對聽力，皮膚和中樞神經系統的腦膜產生影響。



Close

Potential Dis							our Risk core ⇓
exfoliation syn							16.69897
vogt-koyanagi- disease							8
age-related ma degeneration							4.89818527
wet macular de							0.53566688
glaucoma							5.41672368
open-angle glaucoma	開角型青光眼	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	83	35	0.421686747	13.19312867
atrophic macular degeneration	萎縮性黃斑變性	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	52	1	0.019230769	13.15490196
primary angle closure glaucoma	原發性閉角型青光眼	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病	8	2	0.25	13.150515

Graphen Genomics

Ching-Yung's Risky Genome Mutations

Chrom.	Position	Your Genotype	In Gene	In Coding Area	Reference_SNP ID	Risk Allele	Your Risk Allele Match	Risk Allele Frequency	P-Value (mlog) ⇩	Potential Disease	潛在疾病	Category	疾病類別
6	32575658	GC			rs3021304-G	G	1	0.35	118	vogt-koyanagi-harada disease	沃格特-小柳-原田病	Eye, Ear and Mastoid disorder	眼耳鼻喉疾病

Graphen Genomics

Ching-Yung's Whole Genome Analysis by Risk Allele

Genome-wide association analysis of Vogt-Koyanagi-Harada syndrome identifies two new susceptibility loci at 1p31.2 and 10q21.3.

Hou S *et al.*

Nat Genet, 2014-08-10, *citation #*: 31

KNOW MORE..

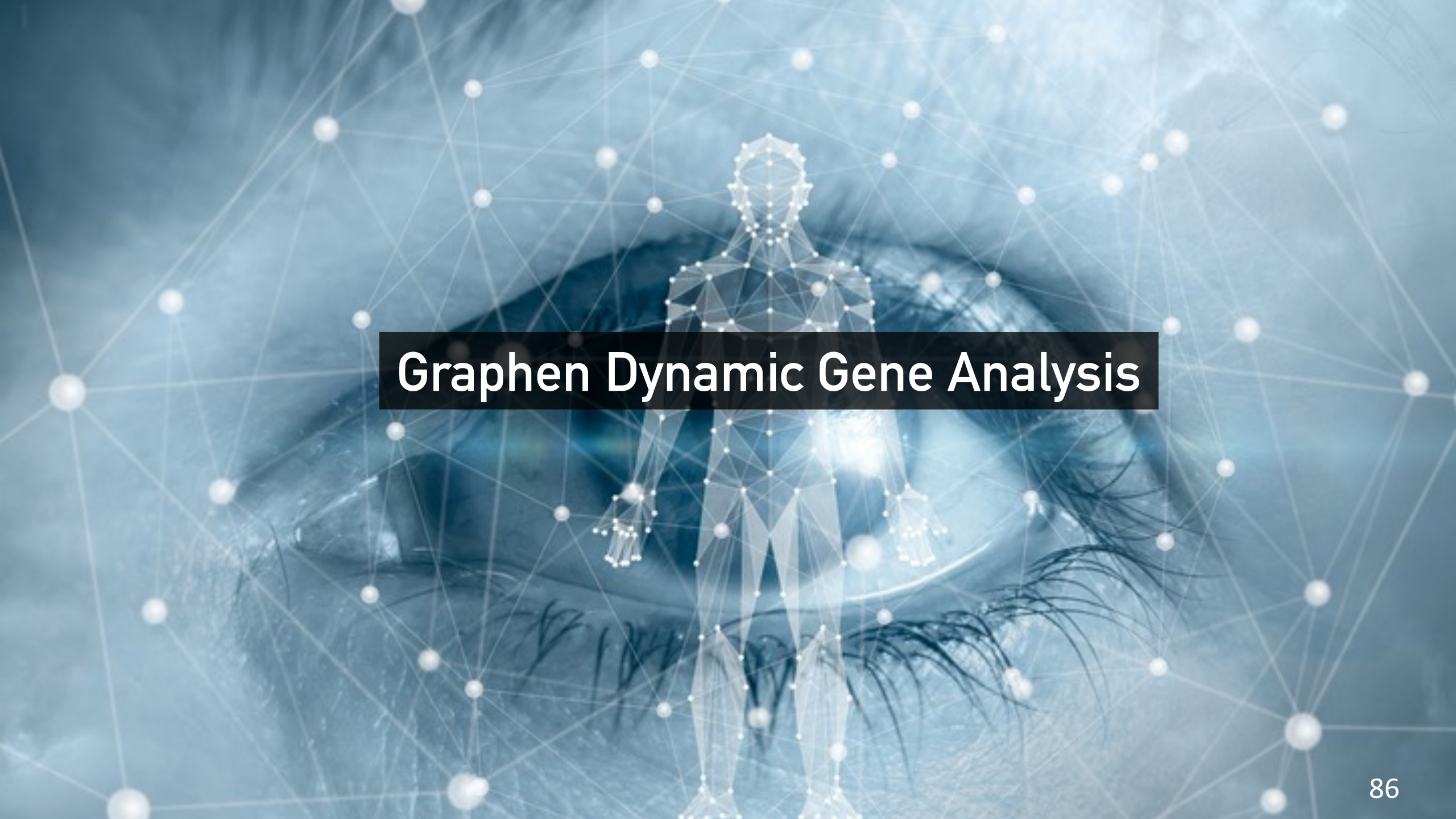
Genome-wide association analysis of Vogt-Koyanagi-Harada syndrome identifies two new susceptibility loci at 1p31.2 and 10q21.3.

×

To identify new genetic risk factors for Vogt-Koyanagi-Harada (VKH) syndrome, we conducted a genome-wide association study of 2,208,258 SNPs in 774 cases and 2,009 controls with follow-up in a collection of 415 cases and 2,006 controls and a further collection of 349 cases and 1,588 controls from a Han Chinese population. We identified three loci associated with VKH syndrome susceptibility (IL23R-C1orf141, rs117633859, P(combined) = 3.42 10(-21), odds ratio (OR) = 1.82; ADO-ZNF365-EGR2, rs442309, P(combined) = 2.97 10(-11), OR = 1.37; and HLA-DRB1/DQA1, rs3021304, P(combined) = 1.26 10(-118), OR = 2.97). The five non-HLA genes were all expressed in human iris tissue. IL23R was also expressed in the ciliary body, and EGR2 was expressed in the ciliary body and choroid. The risk G allele of rs117633859 in the promoter region of IL23R exhibited low transcriptional activation in a cell-based reporter assay and was associated with diminished IL23R mRNA expression in human peripheral blood mononuclear cells.

Read Paper

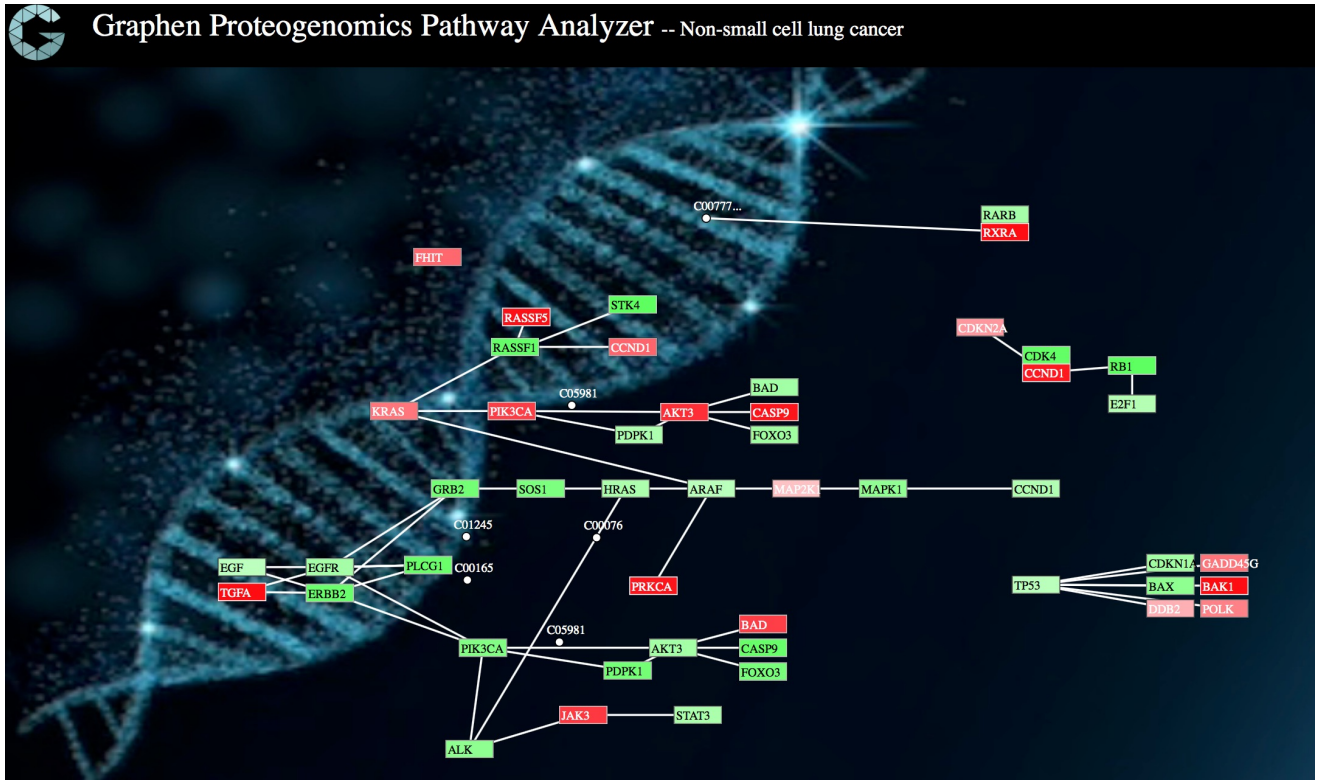
CLOSE



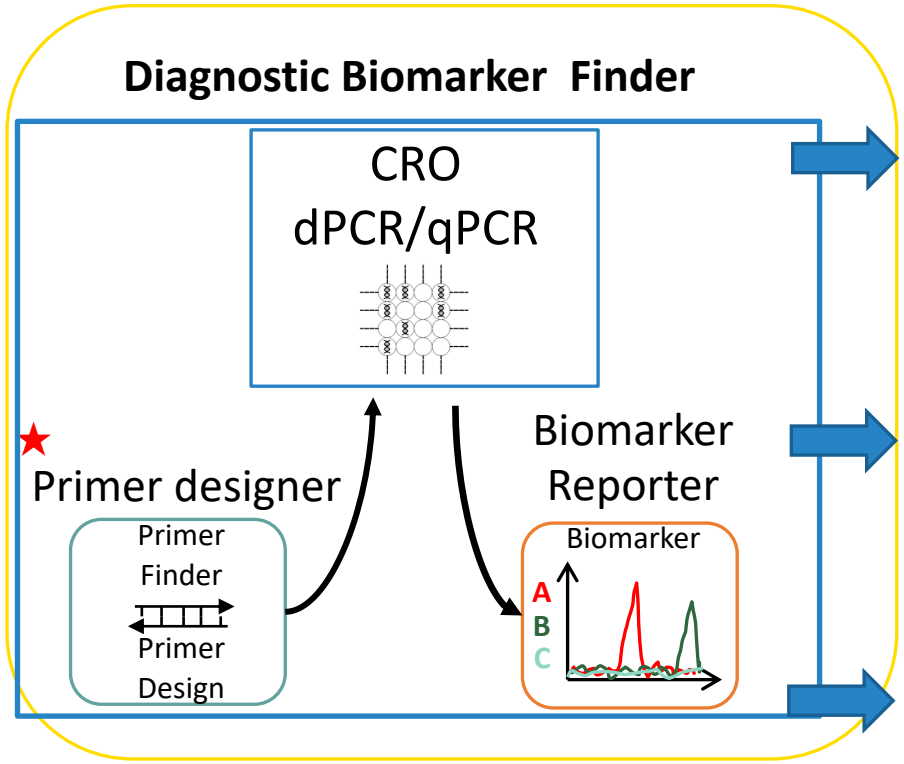
Graphen Dynamic Gene Analysis

Detection Platform to Understand ”where you are today”

GRAPHEN
An Artificial Intelligence Company



Detecting

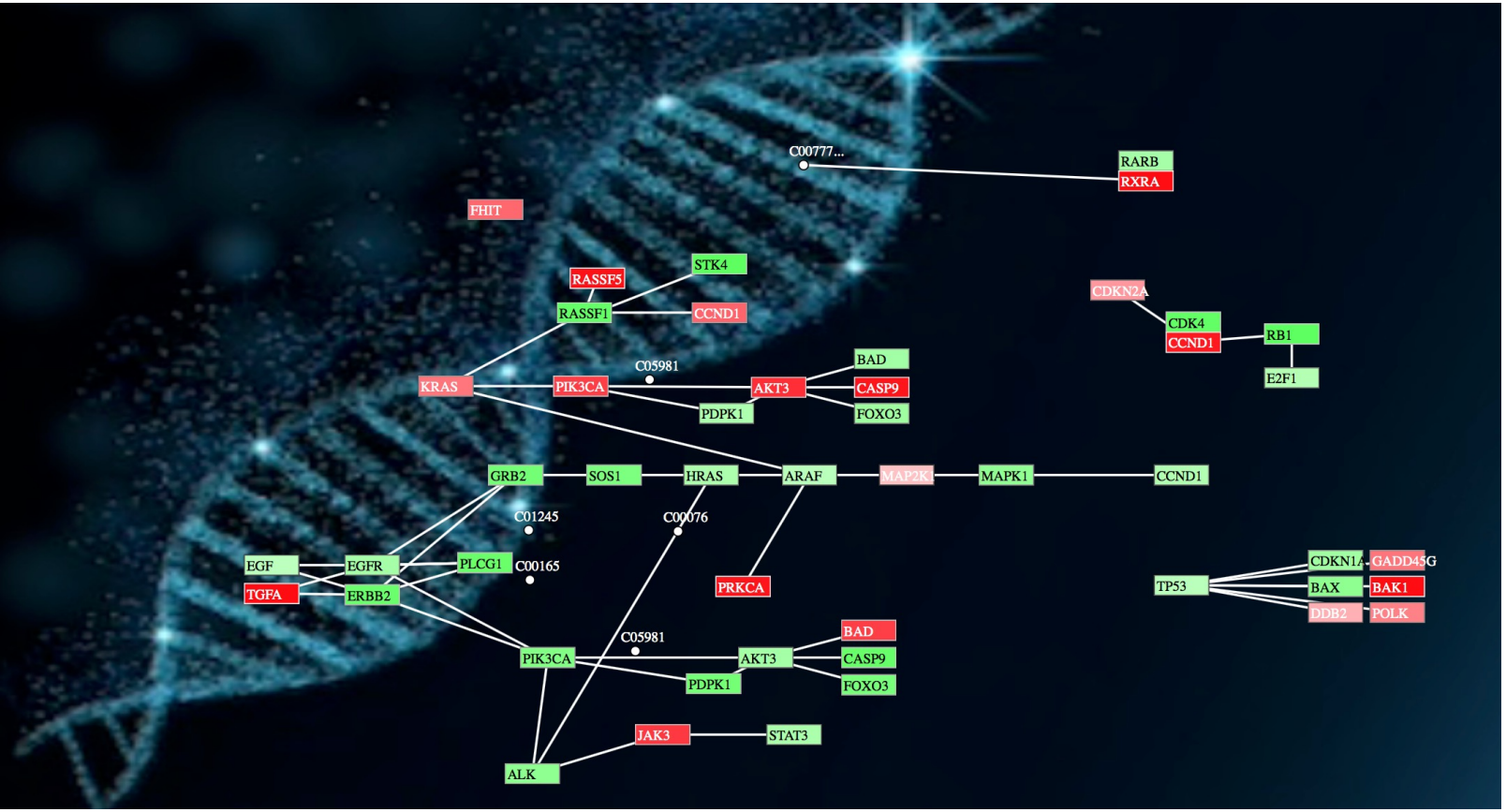


Graphen Genomics

Optimize the Primers by Recurrent Neural Networks

Gene dynamic performance and biological signal transmission chain-integrated gene interaction evaluation

GRAPHEN
An Artificial Intelligence Company



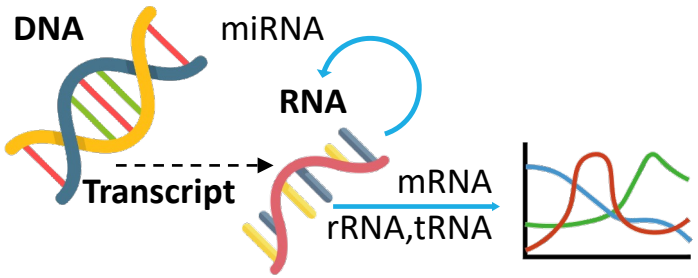
Graphen integrates the Biological Signaling Pathways network formed by the interaction between genes into the evaluation process, linking current gene performance with future potential disease risks:

- Fold change rate of gene expression
- Upstream and downstream associated gene expression rate
- The delivery chain initiates the classification of biological functions

Graphen not only observes the change trend of individual genes, but also evaluates your health-disease progress based on the overall changes in your personal gene network, which is accurate and constructive.

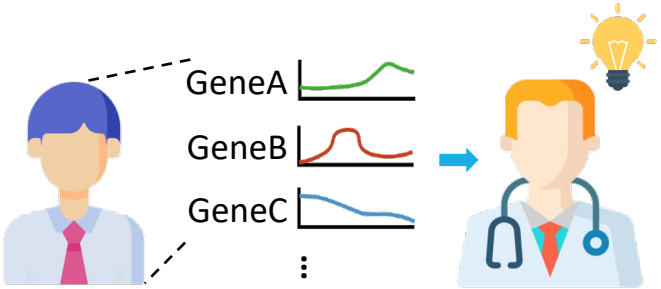
Graphen Gene Dynamic Detection Product Feature

GRAPHEN
An Artificial Intelligence Company



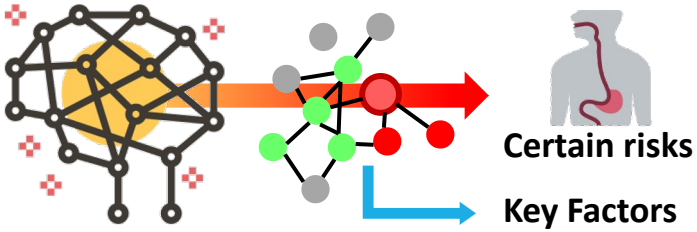
Genetic performance over time: Understand your genetic performance trends

Innate genetic composition contains possible future disease risks, but factors such as environmental factors, living habits, and bacterial virus infections cause changes in acquired gene expression over time, which are the main reasons for triggering diseases. Use genetic tracking to understand the trend of your genetic performance and grasp the safe distance between you and the disease.



Long-term tracking genetic resume: Your prevention and treatment strategy protection

Differences in gene expression determine which lifestyle can reduce your risk of disease, and also affect the most appropriate treatment strategy when you are sick. Through long-term monitoring of your genetic performance trends, Graphen can provide tracking resumes for the first time to help doctors choose the best strategy for you, increase the quality of life and prevent the occurrence of diseases.

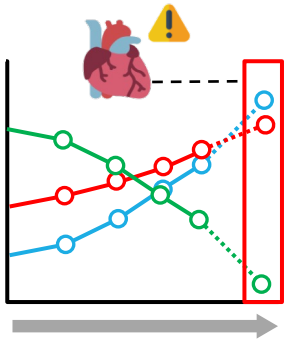


Advanced artificial intelligence technology: uncover the key factors and accurately prevent

Graphen's AI technology analyzes the human's physiological transmission chain, gene and protein mutations, can judge how your gene performance causes specific disease risks, and design precise gene probes to find out key potential factors, see the best, and provide you with precise prevention strategies.

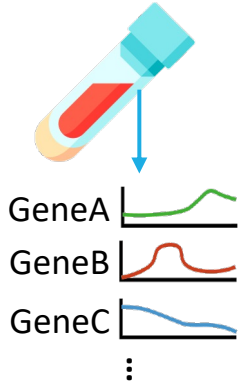
Graphen Gene Dynamic Detection Product Advantages

GRAPHEN
An Artificial Intelligence Company



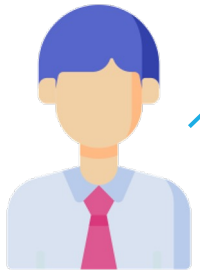
Pre-zero disease progression tracking

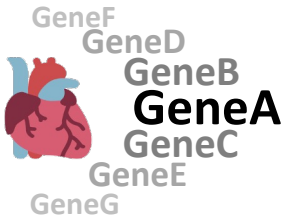
Continue to track changes in gene expression, allowing you to grasp the distance to the underlying disease, and accurate prevention in advance.



Traceable by blood draw

There is no need for complicated testing items, procedures and high costs, and blood draws can accurately track your health progress on a regular basis.

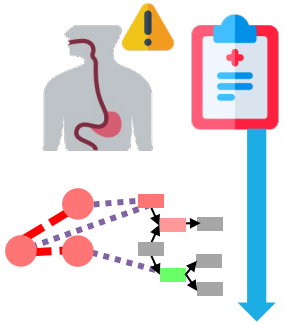




~100 plex

Up to one hundred kinds of target diversity detection

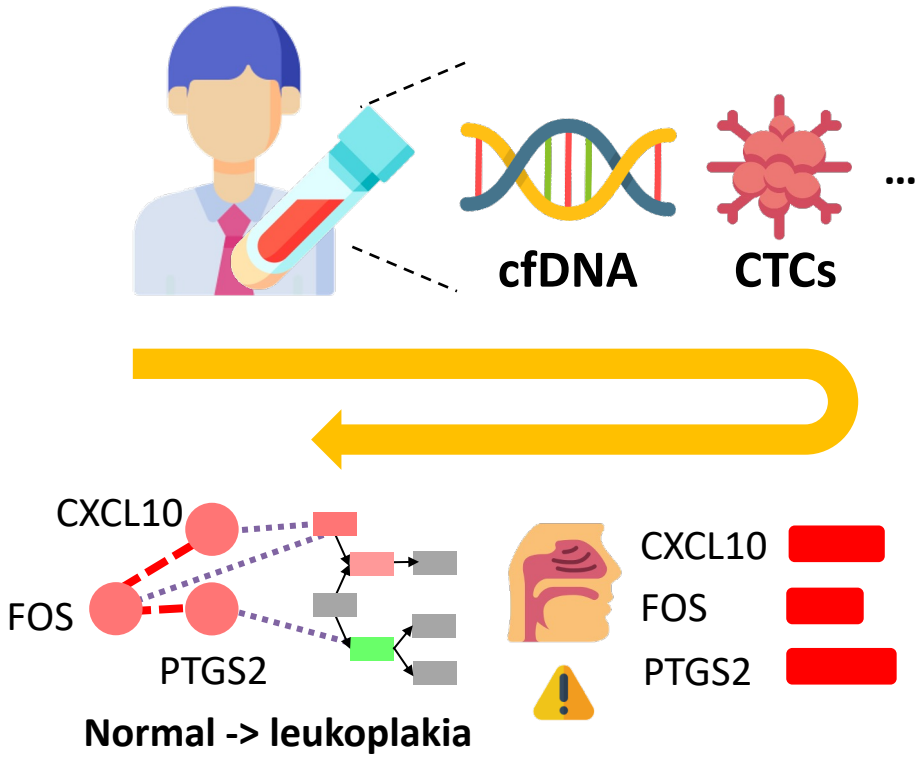
Up to one hundred targets can be evaluated in a single time, without repeated testing, and you can accurately understand your genetic performance trends.



Support health check follow-up evaluation

It can continue to track the risks of specific diseases after the health check, assisting individuals and corporate employees in the health management assessment.

Gene performance dynamics and liquid biopsy



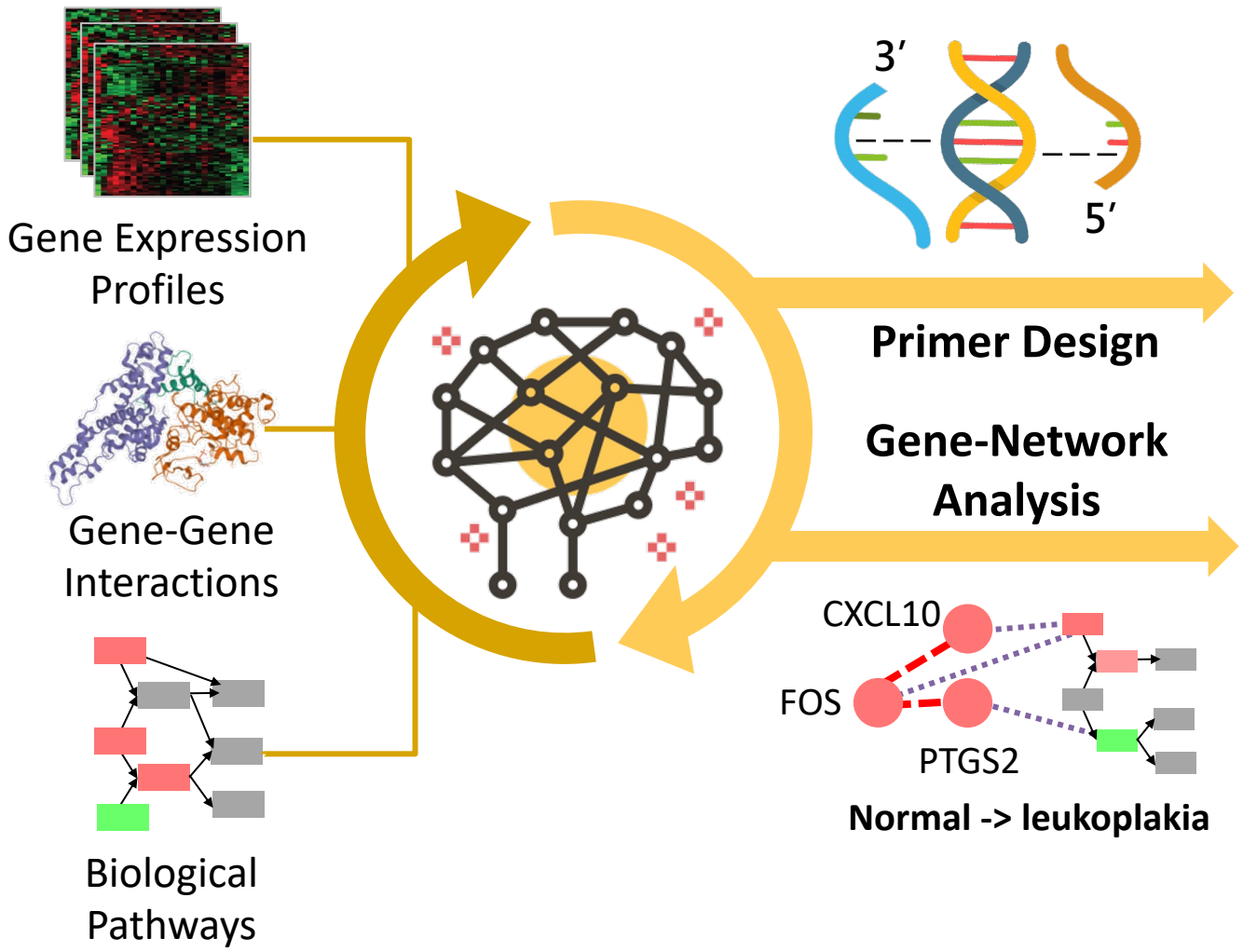
In order to adapt to factors such as living environment, daily habits, and coping with diseases and infections, the human body will adjust the way and degree of gene expression over time and physiological state to form gene dynamics. For example, when infected by viruses or bacteria, immune-related genes will be more active than usual to resist the invasion of foreign pathogens.

In terms of diseases, liquid biopsy/liquid biopsy captures circulating blood nucleic acid (cfDNA) and circulating tumor cells (CTCs) by collecting blood, saliva, urine and other easily obtained body fluid samples. In a simple and less invasive way, analyze the subject's current genetic performance status.

After the subject understands his innate genetic risk, through the dynamic trend of genetic expression, we can evaluate the distance between the subject and the disease state and take effective preventive measures.

Advanced artificial intelligence technology: precise primer design to identify key risk factors

GRAPHEN
An Artificial Intelligence Company



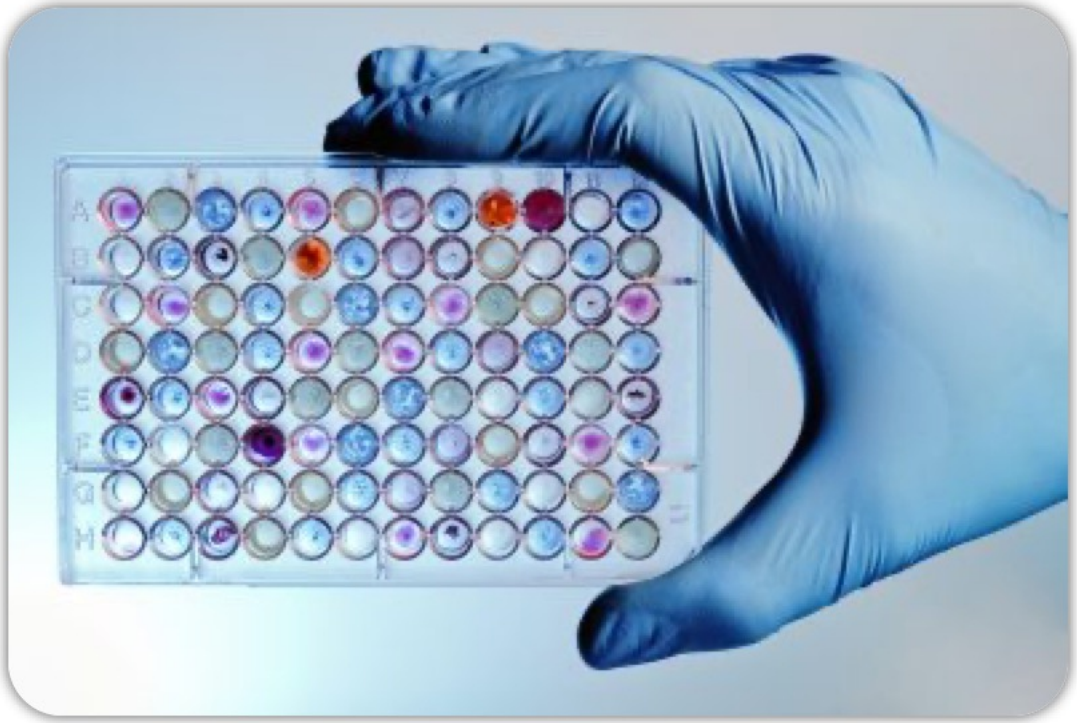
Graphen develops advanced artificial intelligence technology, applied to gene expression network analysis and primer design specially used for multi-target detection, to find out the key factors that cause your future disease risk.

Technology

- Network analysis/prediction of gene-biological delivery chain
- Gene/Translated Protein Mutation Analysis
- Gene target primer design/optimization
- Multi-target primer adhesion prediction

Graphen provides high-precision, low-false positive detection results, and at the same time to find out the key risk factors at multiple levels.

Nearly a hundred target detections in a single test-accurately and quickly solve the needs of multiple diseases

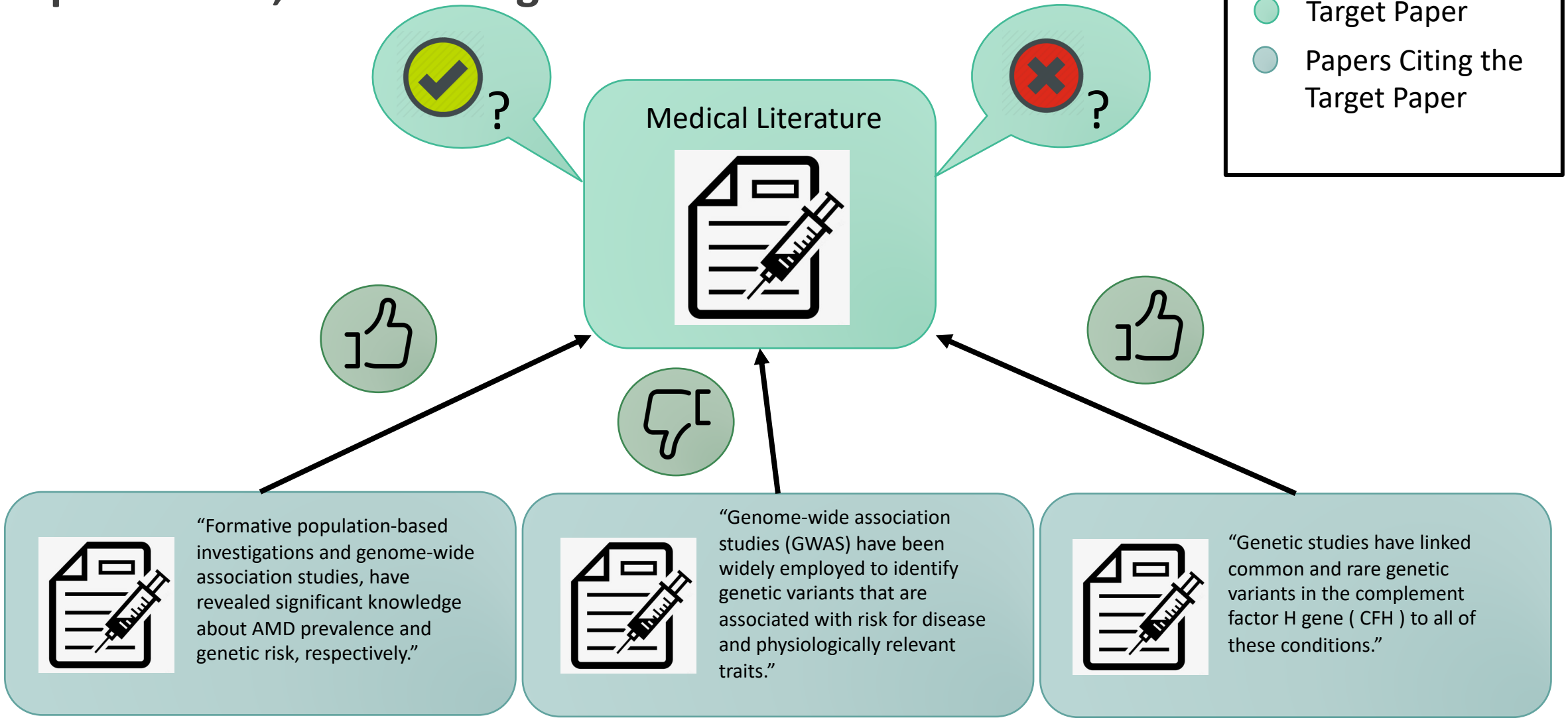


Graphen develops artificial intelligence and combines Multiplex PCR technology to break through the traditional PCR's limitation in the number of single target detections; a single detection can test up to nearly a hundred targets, expanding the depth and depth of gene mutations in disease risk analysis. The breadth of multiple disease-related genes. With faster detection speed and lower detection cost, it meets the needs of modern personalized precision medicine.

Collective Intelligence through Medical Literature AI Reading, Comprehension, and Ranking of Information Networks

GRAPHEN
An Artificial Intelligence Company

- Target Paper
- Papers Citing the Target Paper



Personal massive variants risk likelihood quantification

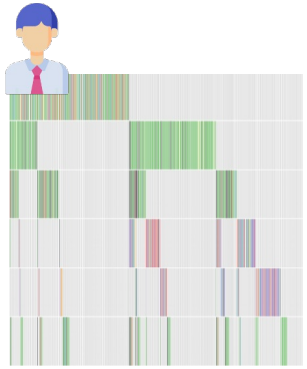
GRAPHEN
An Artificial Intelligence Company

◆**Why Graphen => Evaluation of each variant are processed through integrating its multi-aspect evidence, network relation of citations, and personal condition, making precise scoring results.**

- Using the newest variant annotation databases with information of clinical experiment details, population frequency and statistic index, also using the citation network of corresponding variants to obtain detail variant information.
- **Input**: Gene Profiles (ATOMS/WGS)
- **Out**: Risk scores of diseases, corresponding evidence, and the interpretations of the relations between variants and diseases
- **Target**: Costumer
- **Why use it** => Using Graphen's variant scoring system can provide **Quantified Disease Risks and Interpretable Evidence**

Perform the **Graphen complex variant risk evaluation system** on personal NGS sequencing data (including both WGS and WES data), **quantified evaluation results can be presented to each in-risk disease** through **different clinical information and other risk factors**, such as **the genetic frequency divergence in different areas and races**. Further cooperating with **Ardi understanding and the literature reasoning tools**, the relation between variant and the disease can be interpreted and explained. Quantifying the inherit risk of diseases and other medical issues thus could be possible.

Personal Massive Variants Risk Likelihood Quantification Service



WGS/ATOMS data

+

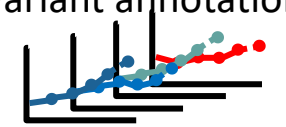
Condition Factors

- Age
- Gender
- ATOMight
- Family History
- Race
- Living area

+

Clinical Annotations

Variant annotation



Allele frequency

ATOM Evaluating System

Quantified Disease Risks

Potential Disease	Category	Known Risky Alleles	Your Matched Alleles	Allele Match Ratio	Risk Score Total	Risk Score African
esophageal carcinoma	Tumor	14	6	43%	95.00	95.00
b-cell acute lymphoblastic leukemia	Tumor	27	11	41%	23.72	40.52
upper aerodigestive tract neoplasm	Tumor	4	2	50%	50.73	37.10
non-small cell lung carcinoma	Tumor	21	8	38%	22.72	25.53
lung adenocarcinoma	Tumor	47	8	17%	19.31	21.46
nephroblastoma	Tumor	6	4	67%	34.33	20.61

+

Interpretable Evidence

Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations.

We conducted a joint (pooled) analysis of three genome-wide association studies (GWAS) of esophageal squamous cell carcinoma (ESCC) in individuals of Chinese ancestry (5,337 ESCC cases and 5,787 controls) with 9,654 ESCC cases and 10,058 controls for follow-up. In a logistic regression model adjusted for age, sex, study and two eigenvectors, two new loci achieved genome-wide significance, marked by rs7447927 at 5q31.2 (per-allele odds ratio (OR) = 0.85, 95% confidence interval (CI) = 0.82-0.88; P = 7.72 10(-20)) and rs1642764 at 17p13.1 (per-allele OR = 0.88, 95% CI = 0.85-0.91; P = 3.10 10(-13)). rs7447927 is a synonymous SNP in TMEM173, and rs1642764 is an intronic SNP in ATP1B2, near TP53. Furthermore, a locus in the HLA class II region at 6p21.32 (rs35597309) achieved genome-wide significance in the two populations at highest risk for ESSC (OR = 1.33, 95% CI = 1.22-1.46; P = 1.99 10(-10)). Our joint analysis identifies new ESCC susceptibility loci overall as well as a new locus unique to the population in the Taihang Mountain region at high risk of ESCC.

Mutation Intelligence

Progress Prediction

Personalized disease risk and drug sensitivity testing services

GRAPHEN
An Artificial Intelligence Company

- **TA**
  - Hospitals/health centers with disease risk prevention and diagnosis development needs
  - Hospitals/health centers with the need for drug sensitivity and diagnosis
- **Advantage**
  - Graphen uses NLP algorithm to automatically read high-quality medical literature, look for genetic variation (SNP) and disease correlation, establish a knowledge graph of genetic variation and give disease correlation scores to provide quantitative information on disease risk and preventive health care.
- **Application**
  - Digital health platform: for enterprise employee health check to provide this service, not only traditional blood biochemical testing, but more comprehensive understanding of potential disease risks, to provide customer disease prevention and health information

Potential Disease	Category	Known Risky Alleles	Your Matched Alleles	Allele Match Ratio	Risk Score Total	Risk Score African
esophageal carcinoma	Tumor	14	6	43%	95.00	95.00
b-cell acute lymphoblastic leukemia	Tumor	27	11	41%	23.72	40.52
upper aerodigestive tract neoplasm	Tumor	4	2	50%	50.73	37.10
non-small cell lung carcinoma	Tumor	21	8	38%	22.72	25.53
lung adenocarcinoma	Tumor	47	8	17%	19.31	21.46
nephroblastoma	Tumor	6	4	67%	34.33	20.61

GRAPHEN

www.graphen.ai

中文： www.graphen.ai/tw

日本語： www.graphen.ai/jp

Looking for Global Business and R&D Collaborations.
Contact: atom@graphen.ai

Advancing State-of-the-Art AI Technologies for The Mankind