UKRI Medical Research Council

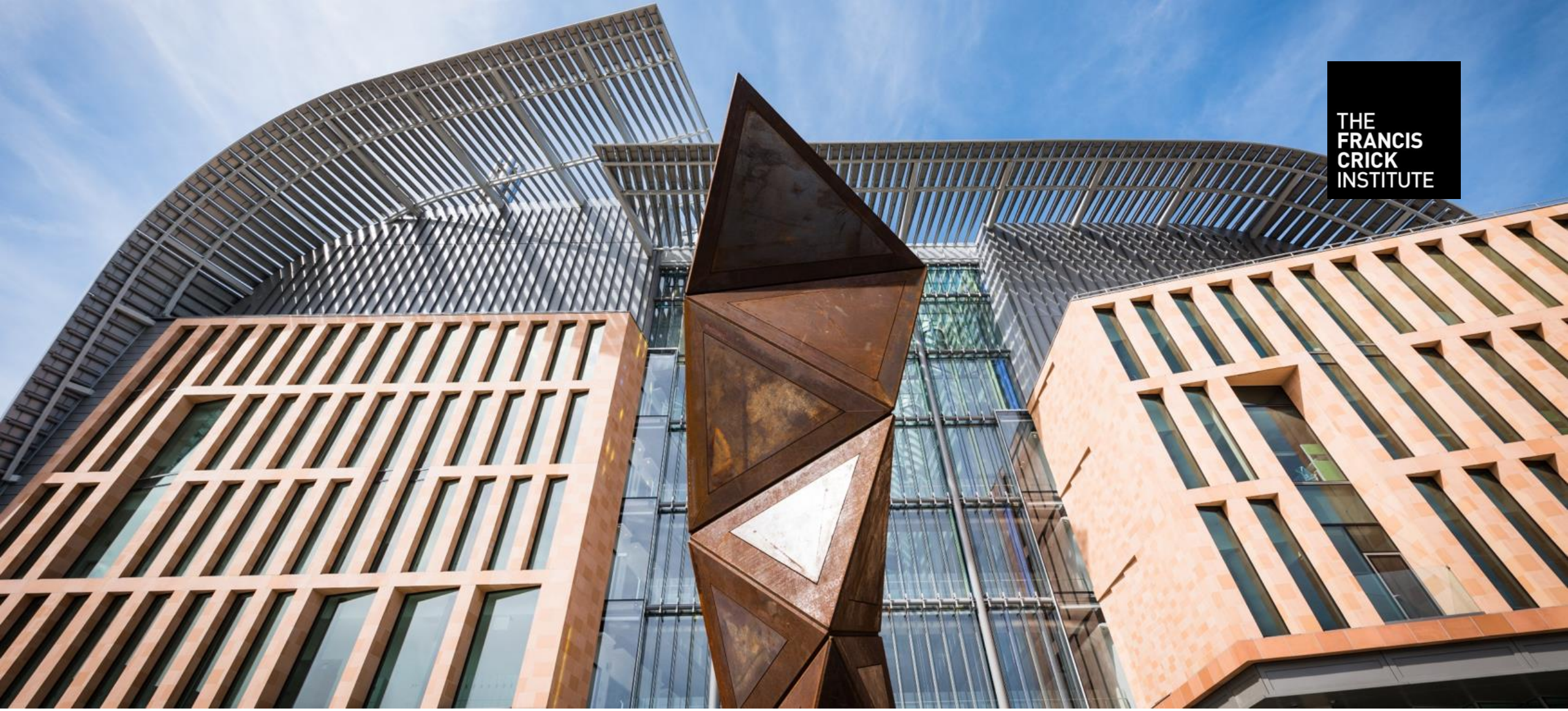
UCL

Imperial College London

King's College London

Wellcome

Cancer Research UK



UKRI Medical Research Council

UCL

Imperial College London

King's College London

Wellcome

Cancer Research UK

Heterogeneous Computing at the Crick

Steve Hindmarsh
Head of Scientific Computing, The Francis Crick Institute
CIUK 9th December 2021

Steve.Hindmarsh@crick.ac.uk

THE FRANCIS CRICK INSTITUTE

The Francis Crick Institute

- A biomedical discovery institute dedicated to ***understanding the fundamental biology underlying health and disease***.
- Founded in 2015 with the merger of two London research institutes from the Medical Research Council and Cancer Research UK into the Crick
- Supported by our founding partners:

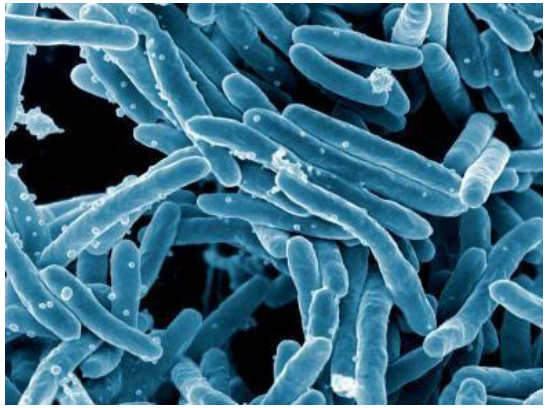
THE FRANCIS CRICK INSTITUTE

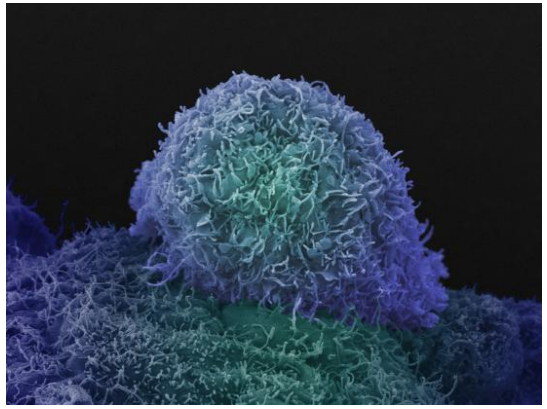
“Discovery without boundaries”

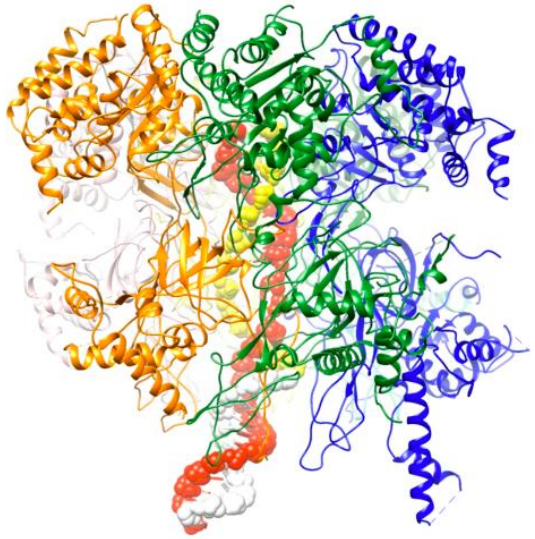
- Core research areas:

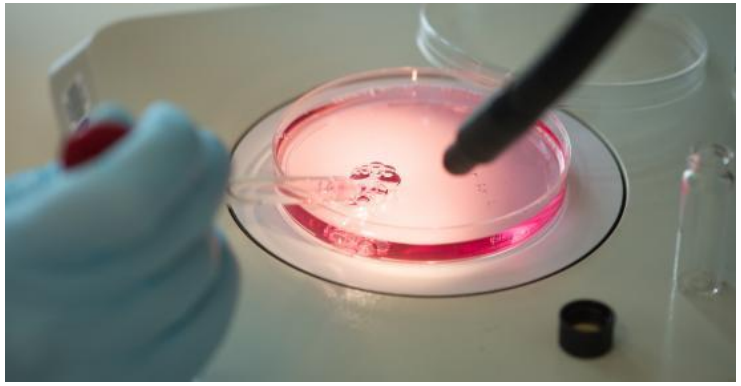
- Growth and development
- Health and ageing
- Human biology
- Cancer
- Immune system
- Infectious disease (including COVID-19!)
- Neuroscience

Multi-disciplinary approach: biology, physics, chemistry, bioinformatics, maths, engineering…









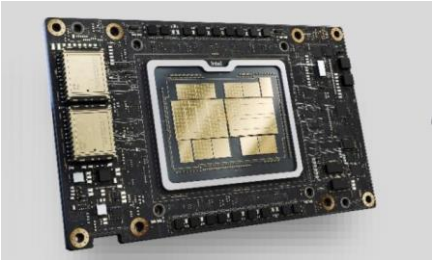
Spoiler alert: Not just HPC!

THE FRANCIS CRICK INSTITUTE

HPC

+ ? + ?

Compute - mainstream research trends:



Past	Present	Future
Gigaflops	Petaflops	Exaflops
CPU	CPU + GPU	CPU + GPU + FPGA + custom ASICs + xPU + Quantum?

↑ Processor heterogeneity – **workload** specific
↑ Parallelism (# processor cores and # nodes)
↑ Memory capacity & bandwidth (but ↓ per core)
↑ Power and cooling requirements

Cloud:

Past	Present	Future
On-site cluster	On-site clusters + Private Cloud/ VMs + **niche** Public Cloud	On-site/ Public Cloud **hybrid**

- **Workload** dependent – ‘data gravity’, network, new processor types
- Cost? (Storage vs Compute), CAPEX → OPEX Funding
- Not everything will be cloud!

Ease of access:

Past	Present	Future
UNIX/Linux Command line Interface (CLI)	CLI + basic GUI	CLI + GUIs + embedded in applications

↑ Access by researchers, clinicians, industry partners etc.

↓ Barrier to HPC benefits

CAMP: Crick Data Analysis and Management Platform

THE FRANCIS CRICK INSTITUTE

OCF

DDN STORAGE

Mellanox TECHNOLOGIES

CAMP

Lenovo

DELL Technologies

intel

NVIDIA

IBM Spectrum Scale

Crick CPU cluster (2016-)

THE FRANCIS CRICK INSTITUTE

~3000 physical cores, (6000 virtual cores with hyperthreading)

- 194 Regular CPU nodes: 2 x 8-core Intel Haswell, 256 GB RAM
- 4 High RAM CPU nodes: 4 x 12-core Intel Haswell, 1.5 TB RAM
- 8 interactive CPU nodes: 2 x 12-core Intel Skylake, 384 GB RAM
- InfiniBand FDR + 40G Ethernet

OCF

Lenovo

intel

Mellanox
TECHNOLOGIES

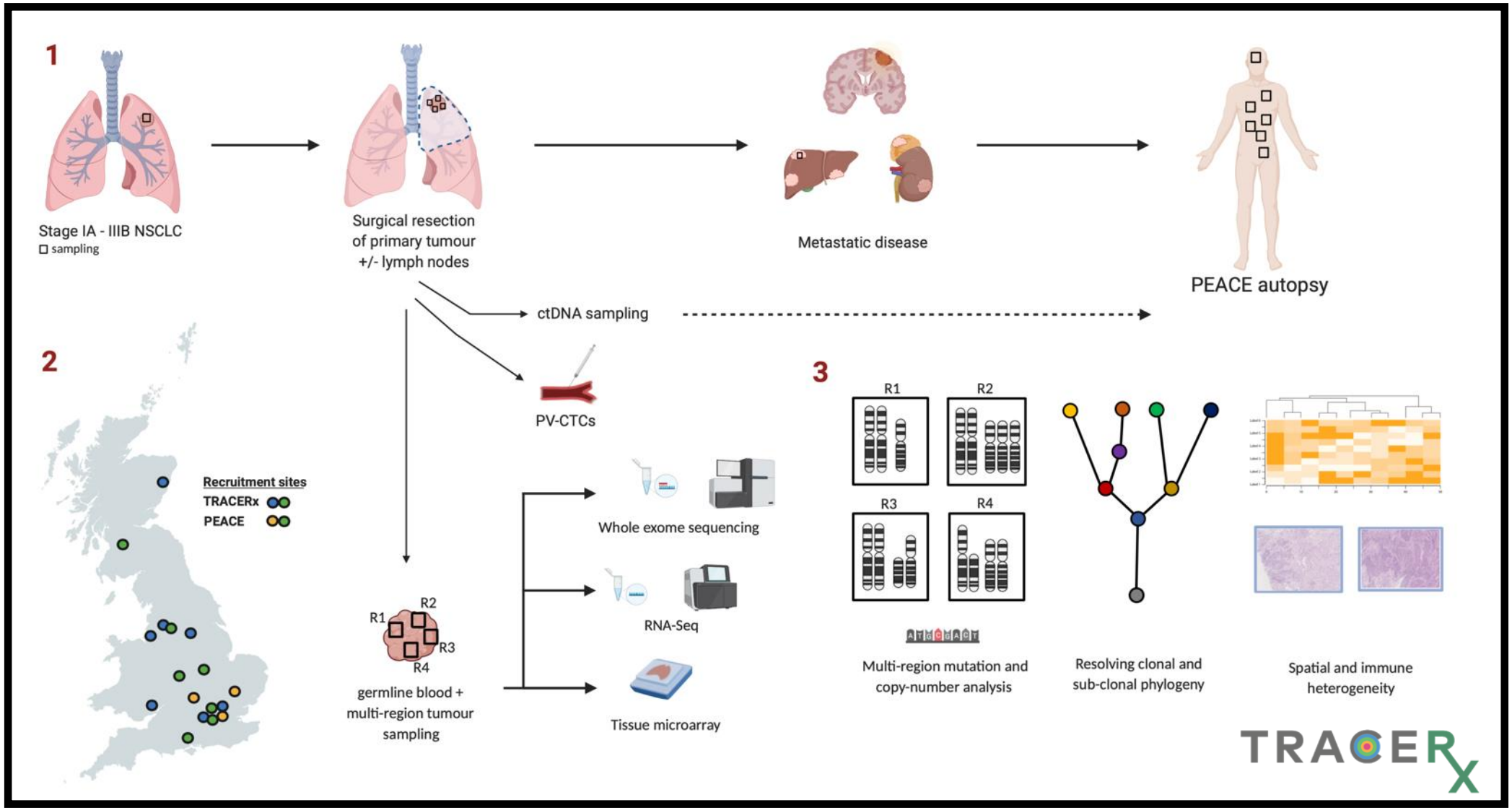
- Workloads: Genomics, data analysis, molecular modelling

CPU Research Applications

THE FRANCIS CRICK INSTITUTE

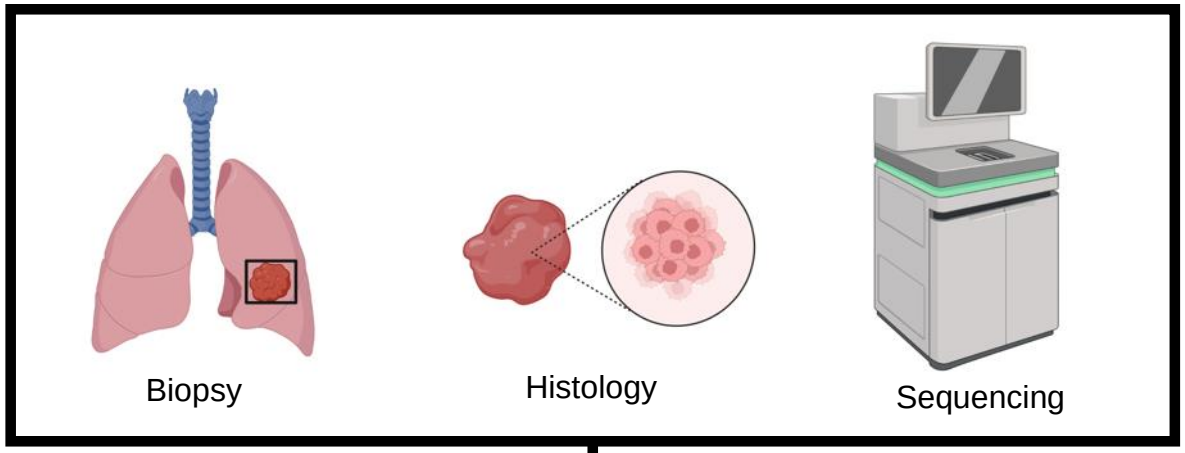
- Cancer evolution genomics
- COVID-19 variant tracking

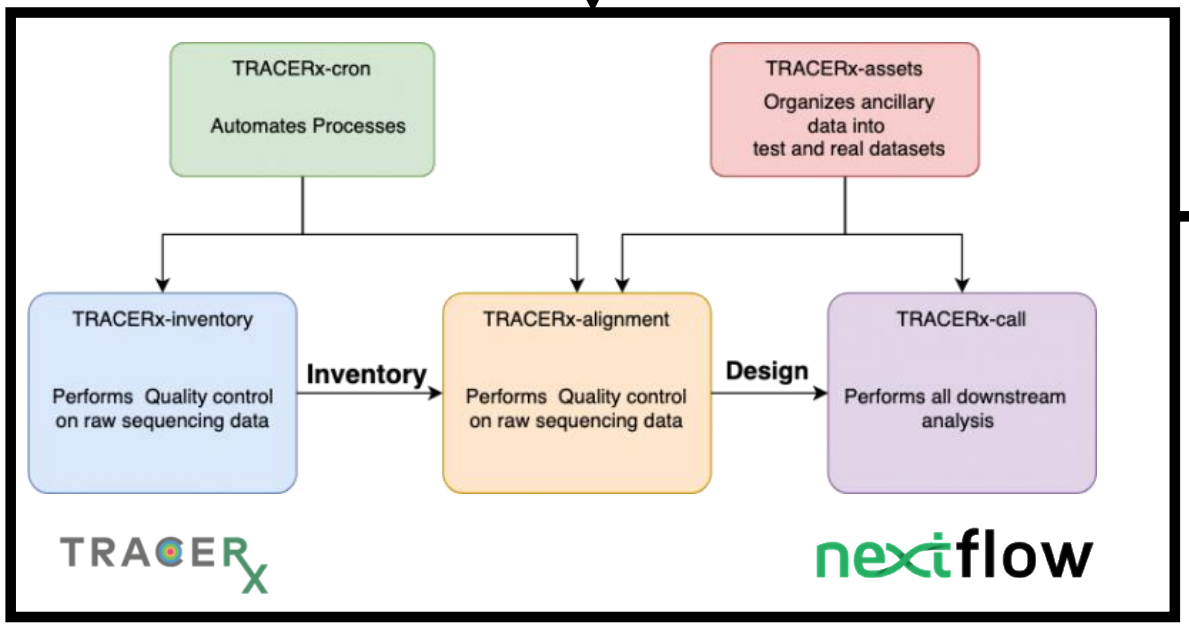
The TRACERx study aims to profile the evolutionary history of 842 patients with non-small cell lung cancer



Bailey et al 2021

The Crick's HPC facility has ensured the processing of whole exome sequencing data





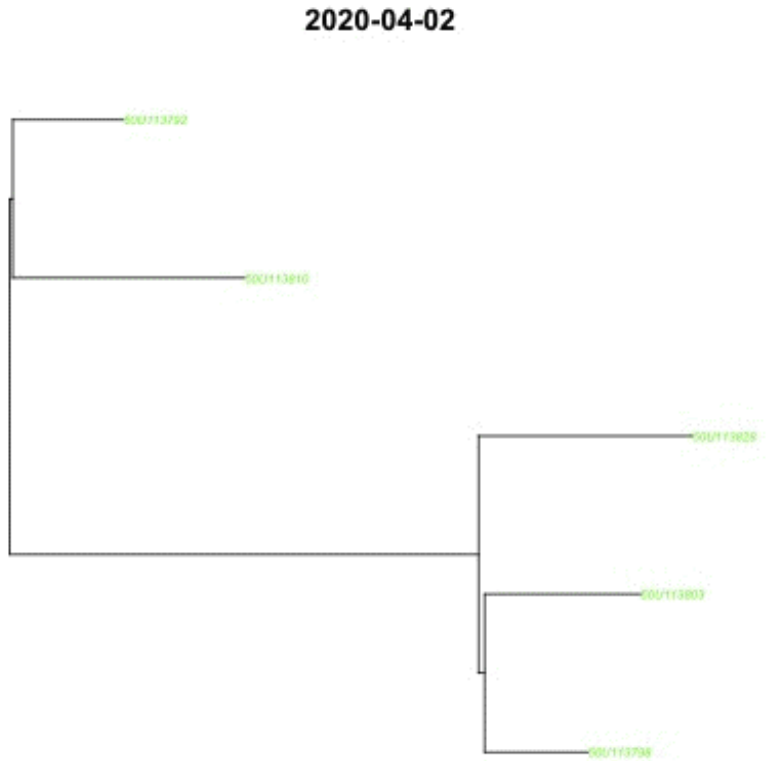
The exome makes up 1-2% of the genome
The TRACERx study has now expanded to whole genome sequencing with 357 samples processing

Jamal-Hanjani et. al 2017
Rosenthal et al 2019
Biswas et al 2019
Lopez et al 2020

COVID-19 sequencing (part of COG-UK variant tracking)

THE FRANCIS CRICK INSTITUTE

- All PCR+ tests processed by Crick were sequenced
- Sequence data processed on CPU cluster
- Animation shows new variants over time
- Data fed into COG-UK



COG-UK COVID-19 GENOMICS UK CONSORTIUM

Crick GPU clusters (2019-)

Main GPU cluster:

- 40 nodes: 4 x Nvidia V100 32 GB NVLink, 2 x 20-core Intel Skylake, 768 GB RAM

Structural Biology (cryo-EM) GPU cluster:

- 11 nodes: 4 x Nvidia RTX5000 16 GB, 2 x 20-core Intel Skylake, 348 GB RAM
- (Replaced local GPU workstations for cryo-EM)

Interactive GPU cluster:

- 5 nodes: 4 x Nvidia RTX5000 16 GB, 2 x 20-core Intel Skylake, 348 GB RAM

InfiniBand FDR + 40G Ethernet

- Workloads: cryo-EM, image processing, AI/ML

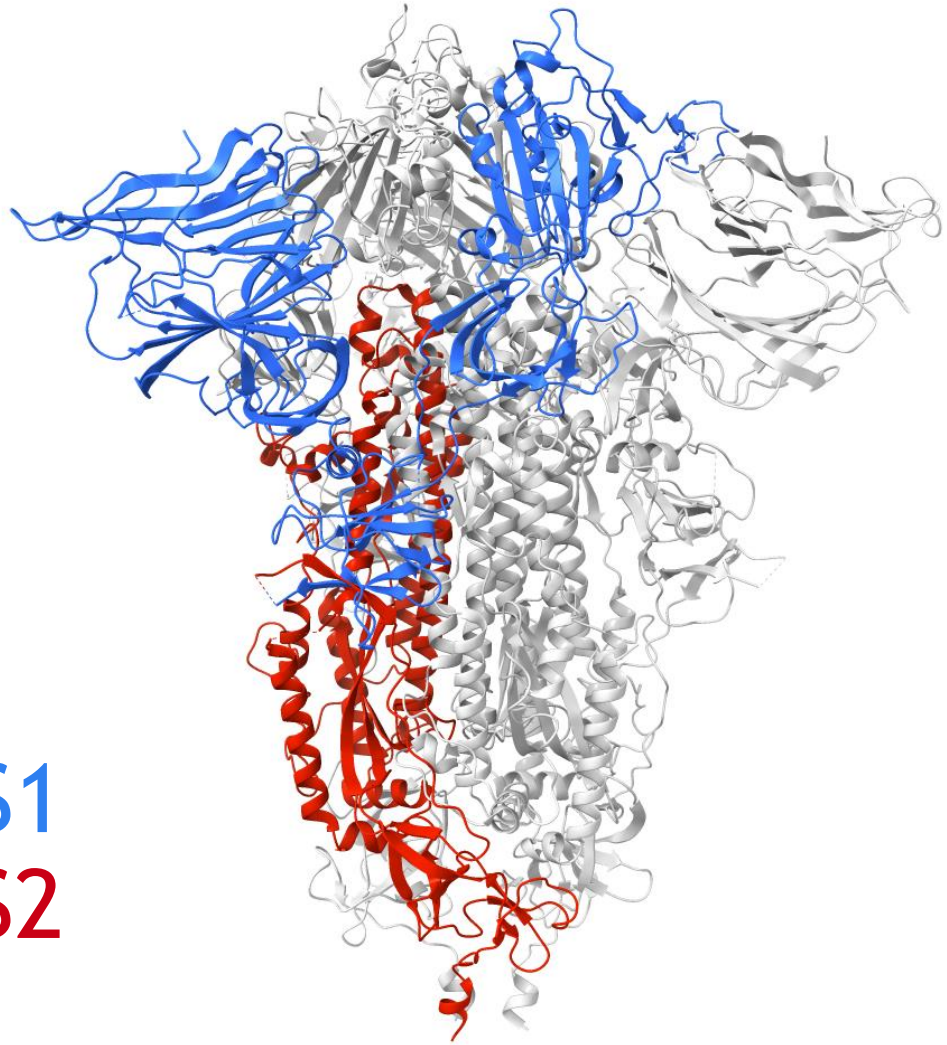
GPU Research Applications

THE FRANCIS CRICK INSTITUTE

- SARS-CoV-2 structure
- Dynamic protein structures
- Cell organelle image segmentation

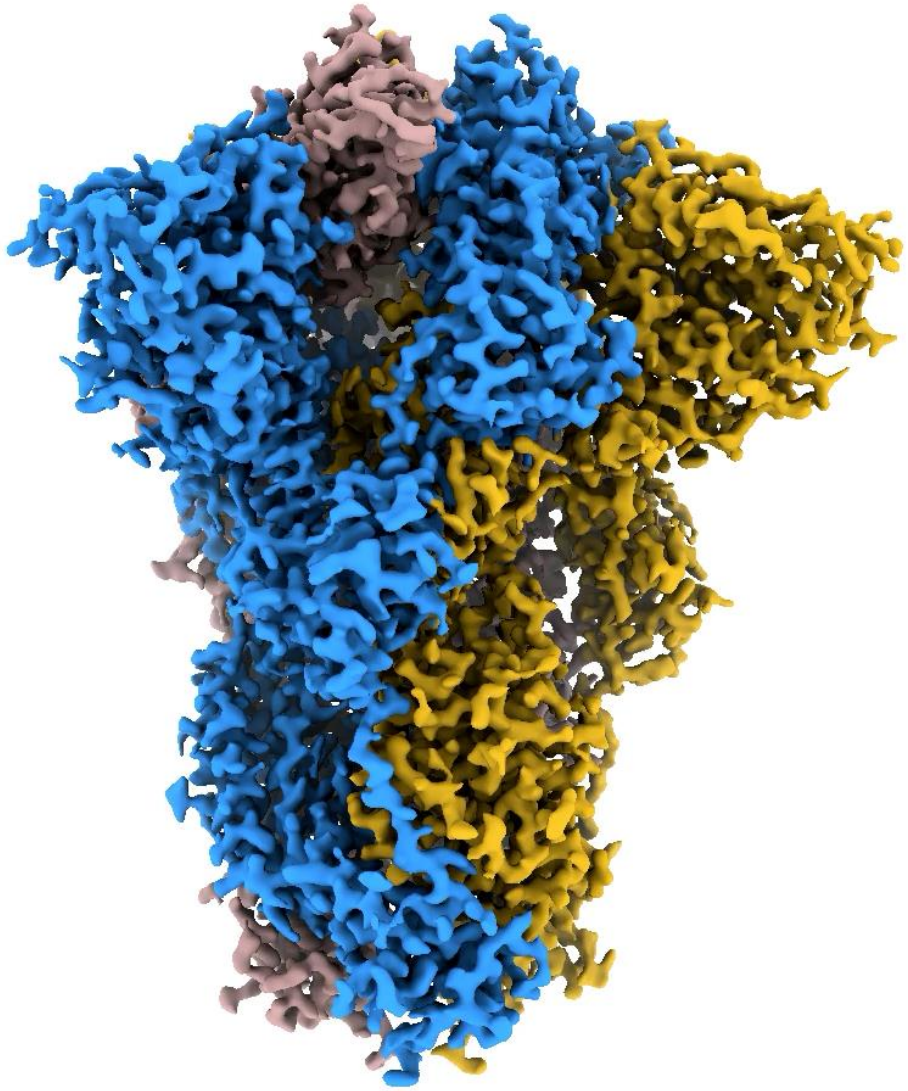
SARS-CoV-2 Spike Structure at 2.6Å

THE FRANCIS CRICK INSTITUTE



S1

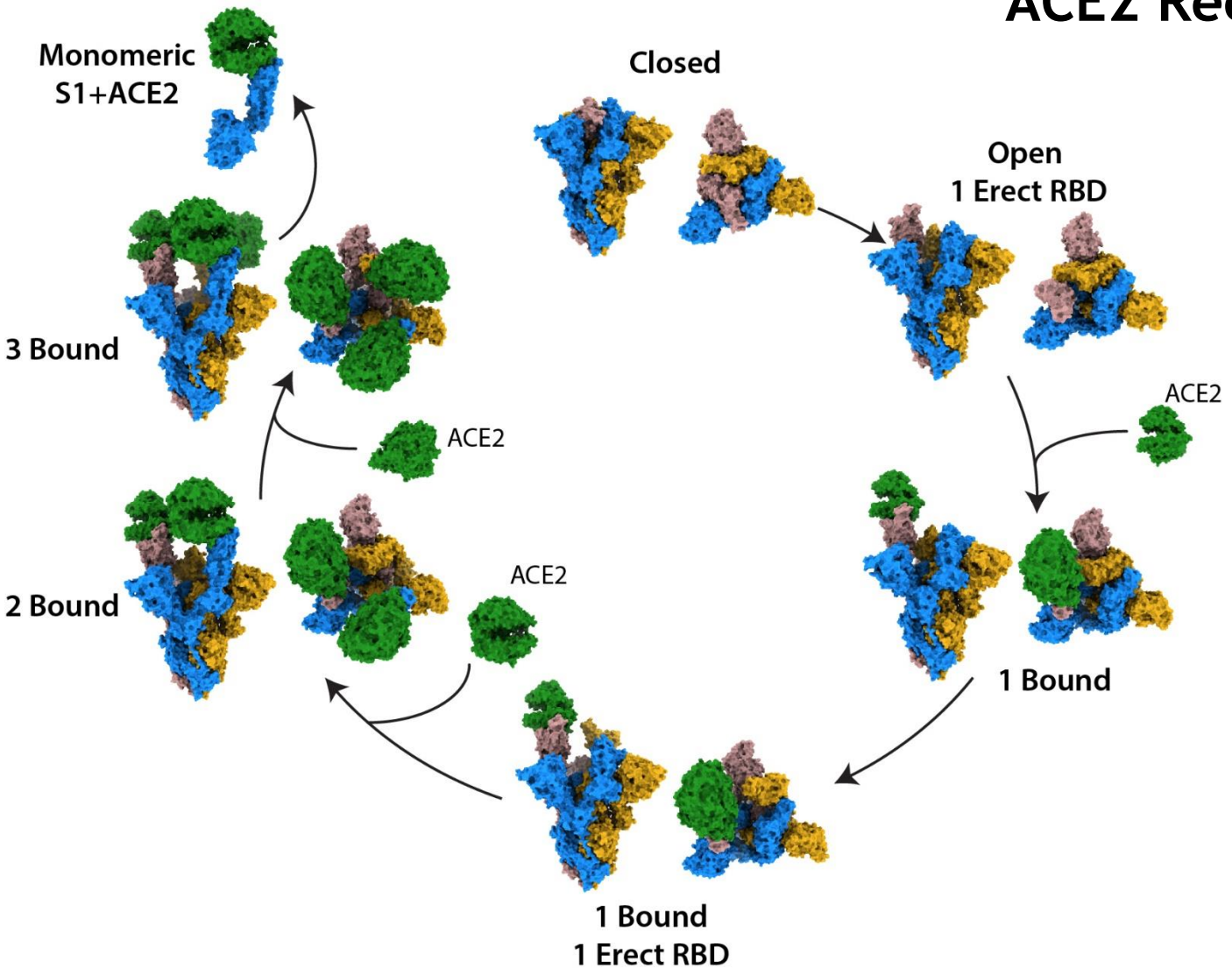
S2

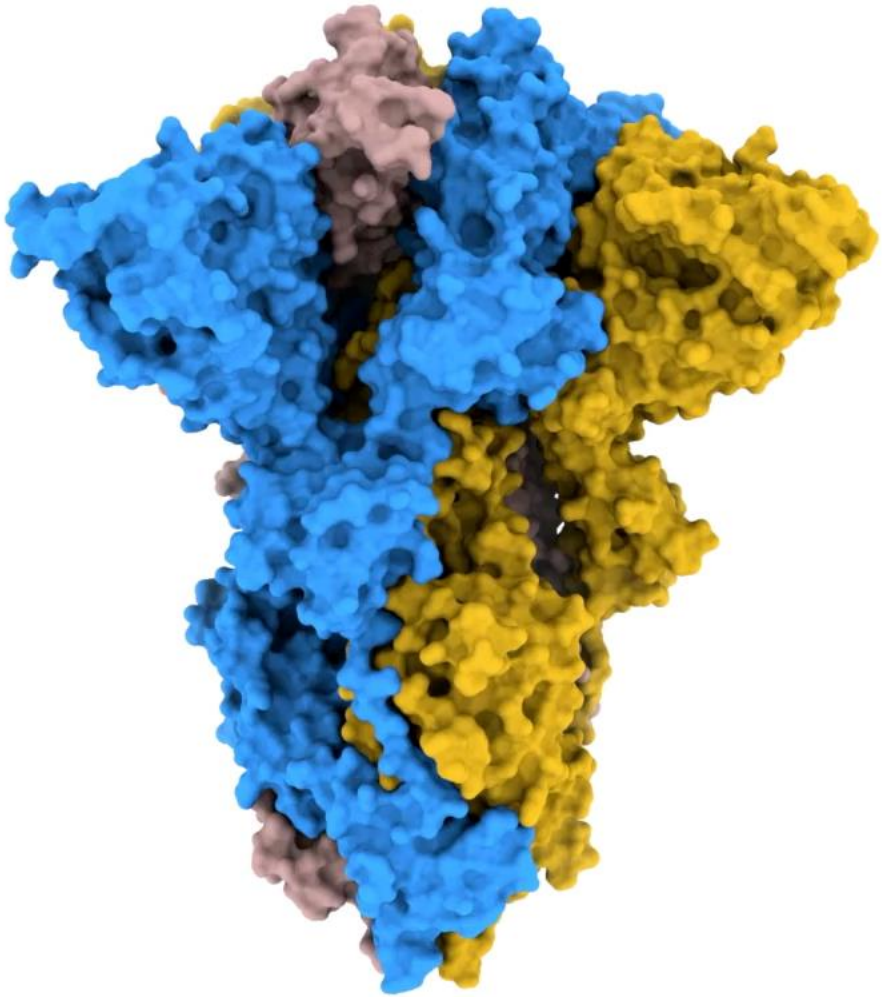


EMPIAR-10469

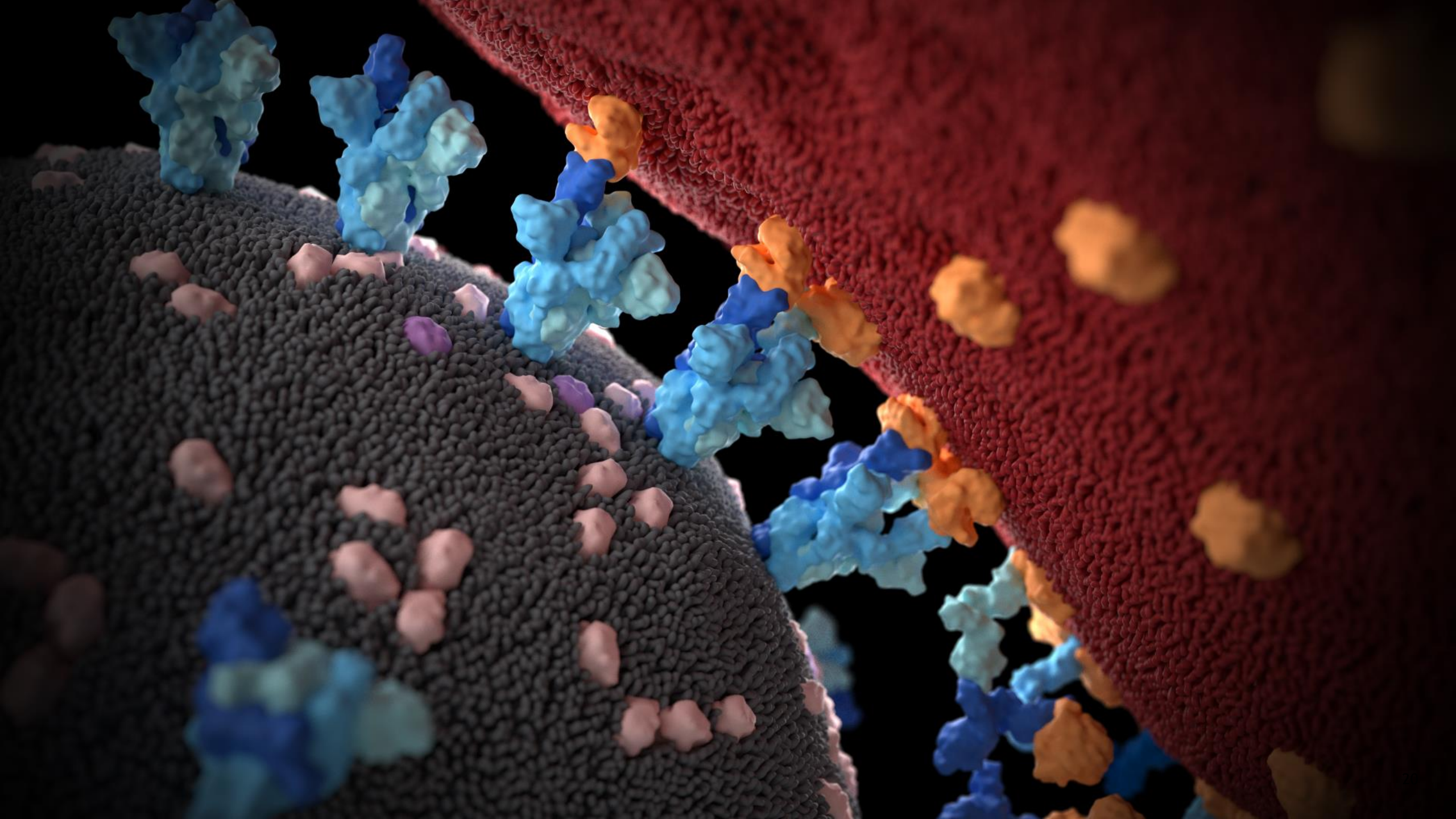
Wrobel*, Benton* et al., NSMB, 2020

ACE2 Receptor Binding to SARS-CoV-2 Spike





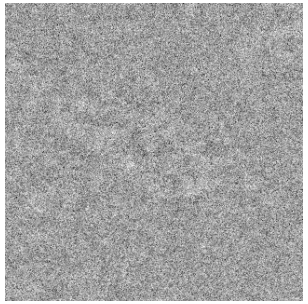
Benton et al., Nature, 2020



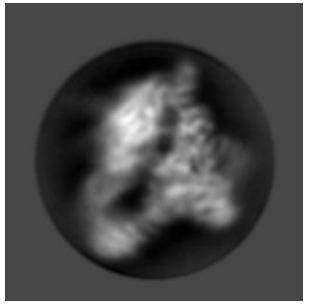
Dynamic protein structures from cryo-EM

Software solutions
cryoDRGN and others

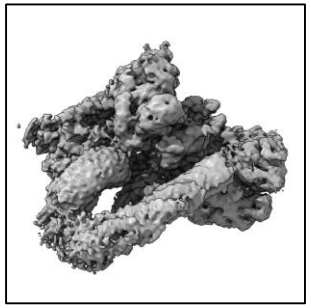
Cryo-EM data



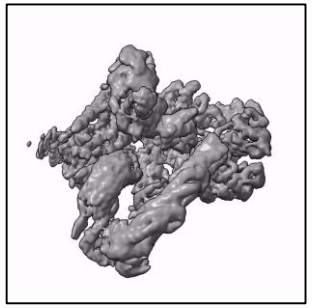
2D alignment



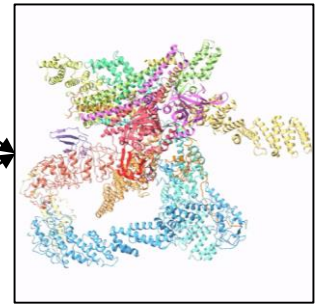
3D alignment



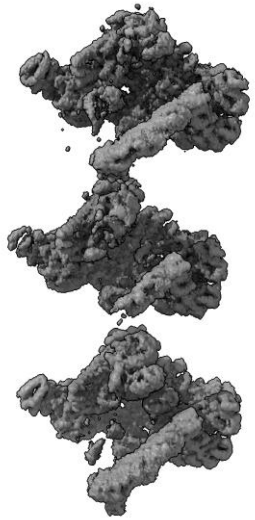
Dynamics



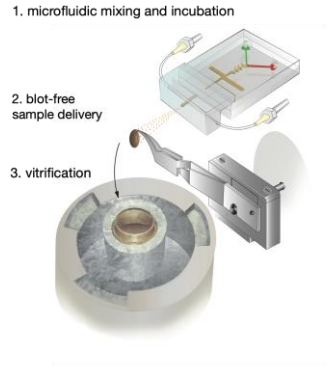
Understanding
function

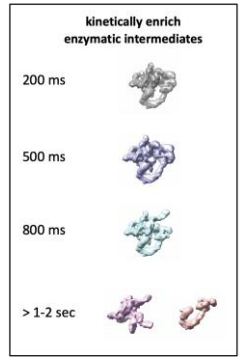


Different classes



Hardware solutions
time-resolved sample preparation





Cryo-em reconstruction
Märt-Erik Mäeots

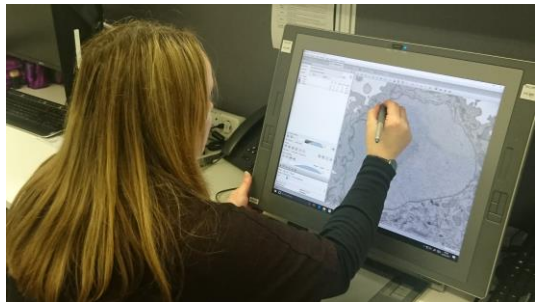
Cell image segmentation using Deep Learning

THE FRANCIS CRICK INSTITUTE

Collaboration with the Electron Microscopy core facility and the Zooniverse citizen science team – called Etch-a-cell.

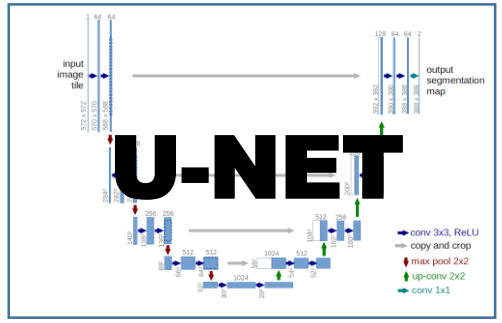
- Organelle segmentation starting with Nuclear Envelope and moving on to Mitochondria and Endoplasmic Reticulum
- Using crowd sourced annotations to train deep learning models

Expert data



Crowd-sourcing

Machine Learning

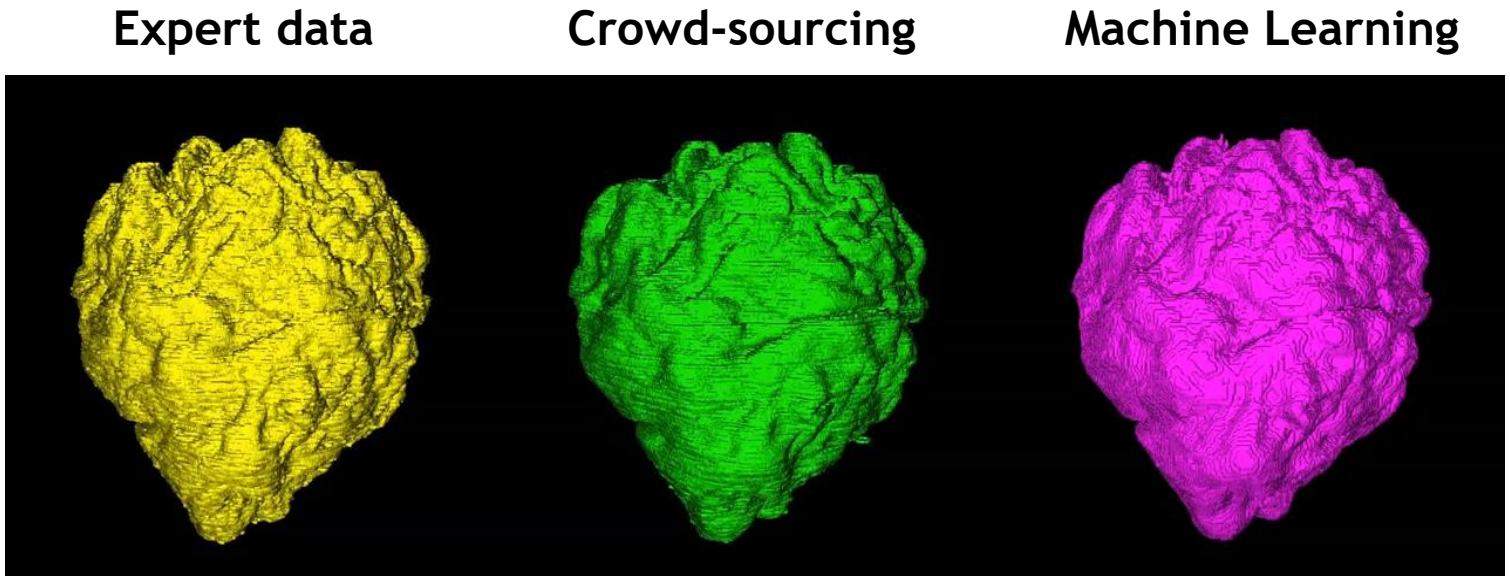


https://github.com/FrancisCrickInstitute/Etch-a-Cell-Nuclear-Envelope | https://www.zooniverse.org/

Segmentation results and next steps

THE FRANCIS CRICK INSTITUTE

- Dice score >0.95 on unseen images



Next Steps

- Future goal to improve generalization across cell and microscope types
- Strategies include training on a greater variety of training data, normalization schemes and adopting alternate model formulations

Crick research data storage 2016-

THE FRANCIS CRICK INSTITUTE

- 11 PiB DDN GRIDScaler: 2 x SFA12K + 2 x GS7K
- MEDIAScaler – NFS/SMB presentation to Mac/Win/Linux clients
- IBM GPFS/Spectrum Scale v4 > v5

OCF

DDN STORAGE

IBM Spectrum Scale

- Backbone of our research capability
- Special thanks to DDN for their excellent support and IBM for v5 licence transition

(Join the Spectrum Scale User Group tomorrow for more details!)

New Crick research data storage 2022-

THE FRANCIS CRICK INSTITUTE

- Lenovo DSS-G
- 15 PiB HDD + 1.2 PiB NVMe
- CES protocol nodes
- IBM Spectrum Scale v5
- Expandable to 30 PiB just by adding 1 PiB disk shelves
- ~9 PiB data migration using Atempo Miria

OCF

Lenovo

IBM Spectrum Scale

Other compute resources

THE FRANCIS CRICK INSTITUTE

- Cloud – AWS and GCP pilots
- Virtual GPU desktops (VMware) – e.g. iterative ML model development

The future of Crick compute is heterogeneous!

THE FRANCIS CRICK INSTITUTE

- Procurement planned in 2022 to replace CPU cluster (EOS June 2023)
- CPU + GPU? + custom ASICs?
- Driven by workloads
- Easy access to cloud (software and networking)
- ‘Spectrum of compute’ for researchers
- Containerisation/virtualisation support
- Software developers/engineers needed!
- GPU cluster EOS 2024
- Quantum?

One size does not fit all… workload specific

Not just HPC - the whole picture

THE FRANCIS CRICK INSTITUTE

HPC

People, Support & Expertise

Relationships

Scientific Computing core facility

Providing Crick researchers with advanced scientific computing platforms, support and skills to deliver discoveries to change lives

We provide a broad range of support for scientific computing across 3 teams, total 20 staff:

- Research Data Services / Database Team
- Software Development & Machine Learning Team
- Research Computing Platforms/HPC

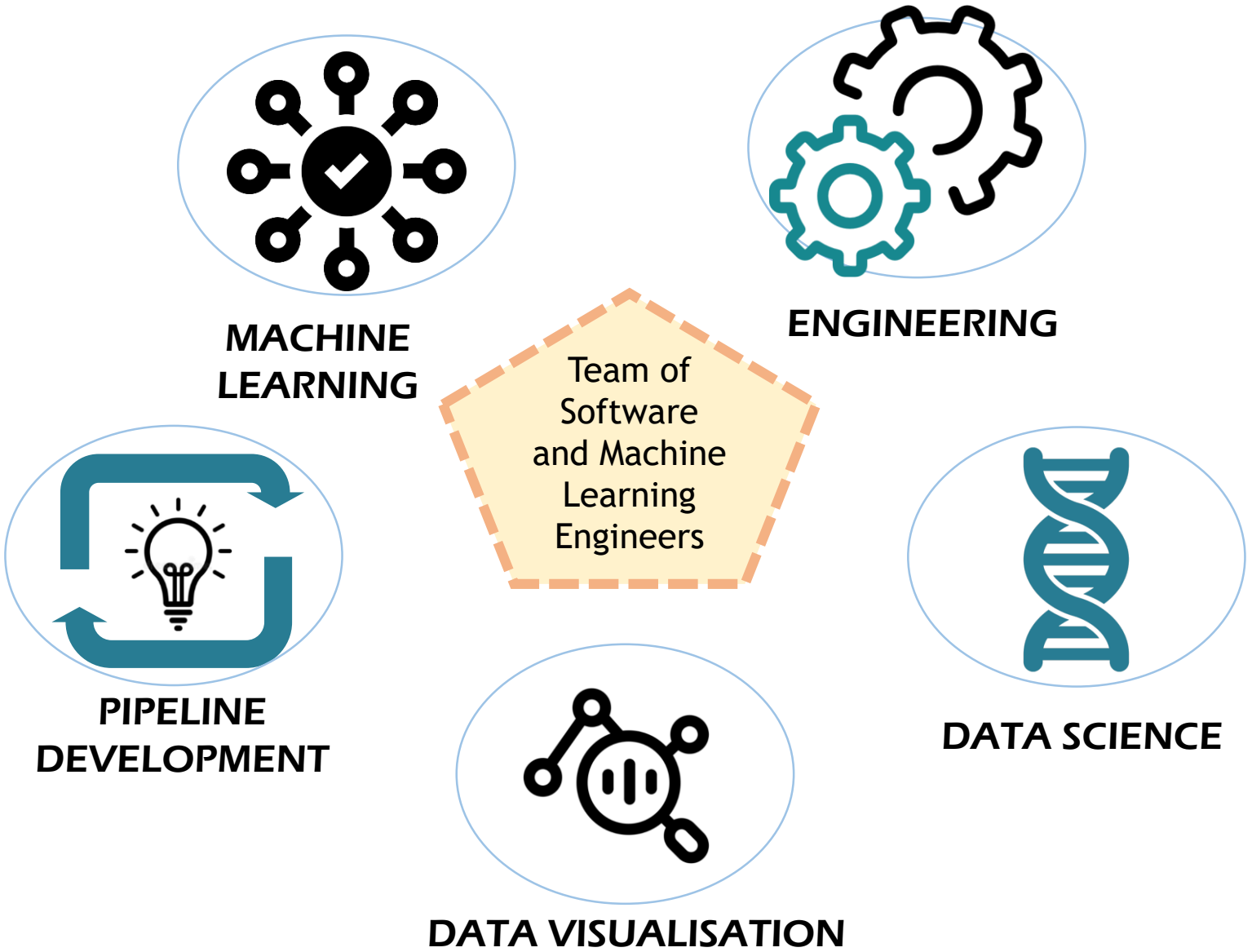
Karen Ambrose
Research Data Services

Amy Strange
Software Development & Machine Learning

Wei Xing
Research Computing Platforms/HPC

Software Development and Machine Learning

THE FRANCIS CRICK INSTITUTE



My own next move…

NBI Partnership

Healthy Plants, Healthy People, Healthy Planet

Hiring in London and Norwich…

Crick HPC & Research Data Systems Engineer

https://www.crick.ac.uk/careers-study/vacancies/2021-10-27-hpc-research-data-systems-engineer

NBI Research Computing Junior Sys Admin

https://jobs.nbi.ac.uk/Details.asp?vacancyID=16662

Talk to me/get in touch:

Steve.Hindmarsh@crick.ac.uk

crick.ac.uk