Outbreak data analysis platform

Outbreak data analysis platform

Created for Covid-19 research, the ISARIC4C data analysis platform combines the data safeguards of an NHS trusted research environment, with >£100M of exabyte-scale computational capacity on the UK national supercomputer. This creates a unique opportunity to combine clinical and genomics research to identify the right treatments, for the right patients.

Data held

The outbreak analysis platform was developed by ISARIC4C to *encourgage and facilitate research* by collating, linking and curating clinical and research data, enabling deep integrative analyses of multi-omic disease profiling, stratified by viral variant, clinical phenotype and outcome.

This platform now serves as a hub for a coordinated UK national research response to COVID-19. Data are included from:

- ISARIC4C tier 0: (unconsented) prospective clinical data from 206069 cases
- ISARIC4C tiers 1 and 2: serial multiomic assays from research samples of blood, respiratory secretions, urine, and stool from 2309 cases
- COG-UK: (unconsented) summary variant data from COG-UK viral sequencing study is already included for matched patients
- GenOMICC study complete data: microarray and whole genome sequence data from 12477 cases
- PHOSP complete data: follow-up clinical and biological data generated by the Post-Hospitalisation for COVID-19 follow-up study (1075 cases)
- UK-CIC: deep immunological phenotyping data from across the UK Coronavirus Immunology Consortium, using ISARIC4C samples and local collections.

Research data within the analysis platform is already linked to:

- NHS Scotland primary, secondary care and death records
- NHS Digital health records data

In future, plans are in place to transfer data to link with:

- ICNARC and SICSAG critical care audit databases
- NIMS National Immunisation Dataset
- Pillar 1 testing
- Pillar 2 testing
- ONS

Research outputs

The ISARIC Coronavirus Clinical Characterisation Consortium (4C) is the largest observational study of hospitalised patients with COVID-19 anyhwere in the world. By generating, integrating and analysing clinical, biological, genetic and virological data on patients with Covid-19 in UK hospitals (Figure 2), ISARIC4C has provided essential weekly updates to SAGE that guide the public health response, and enabled understanding of the clinical features,[1] prognostication,[2] disease biology[3] and host genetics.[4]

Outbreak data analysis platform

EMBARGO AREA

Data accessible to all contributors under strict publication embargo

NON-EMBARGO AREA

Data accessible to approved bona fide researchers for academic or commercial use

FLEXIBLE COMPUTE SPACE

Flexible high-performance compute area for pseudonymised data with explicit consent

RESEARCH DATA:

ISARIC4C TIER 0/1/2 hospitalised (2,307):

host whole genome seq

serial: multiomics

serology

viral load

PHOSP clinical follow-up: (1,075)

host microarray genotype

serial: multiomics

serology

COG-UK Viral whole genome seq

GenOMICC Host whole genome (12,516)

UK-CIC immunology

LINKED DATA:

Summary (non-disclosive) clinical data

All data released by investigators at or before publication

Supervised exports of non-disclosive data

Supervised exports of non-disclosive data

TRUSTED RESEARCH ENVIRONMENT

Safe haven research space for analysis of pseudonymised data without consent

RESEARCH DATA:

ISARIC4C TIER 0 (CO-CIN) (206,378)

ISARIC4C TIER 1/2 Multi-omics (2,307)

PHOSP clinical follow-up (1,075)

PHOSP TIER 1/2 multiomics

COG-UK Viral variant data (21,230)

GenOMICC Host variant data (12,516)

UK-CIC immunology

LINKED DATA (available):

NHS Digital England &Wales In/outpatient

NHS Scotland (SMR00/01) In/outpatient

Death records (Scotland)

Critical care audit (SICSAG)

LINKED DATA (planned):

Critical care audit (ICNARC)

NHS England&Wales Inpatient/Outpatient

National immunisation database (NIMS)

Pillar 1/2 testing

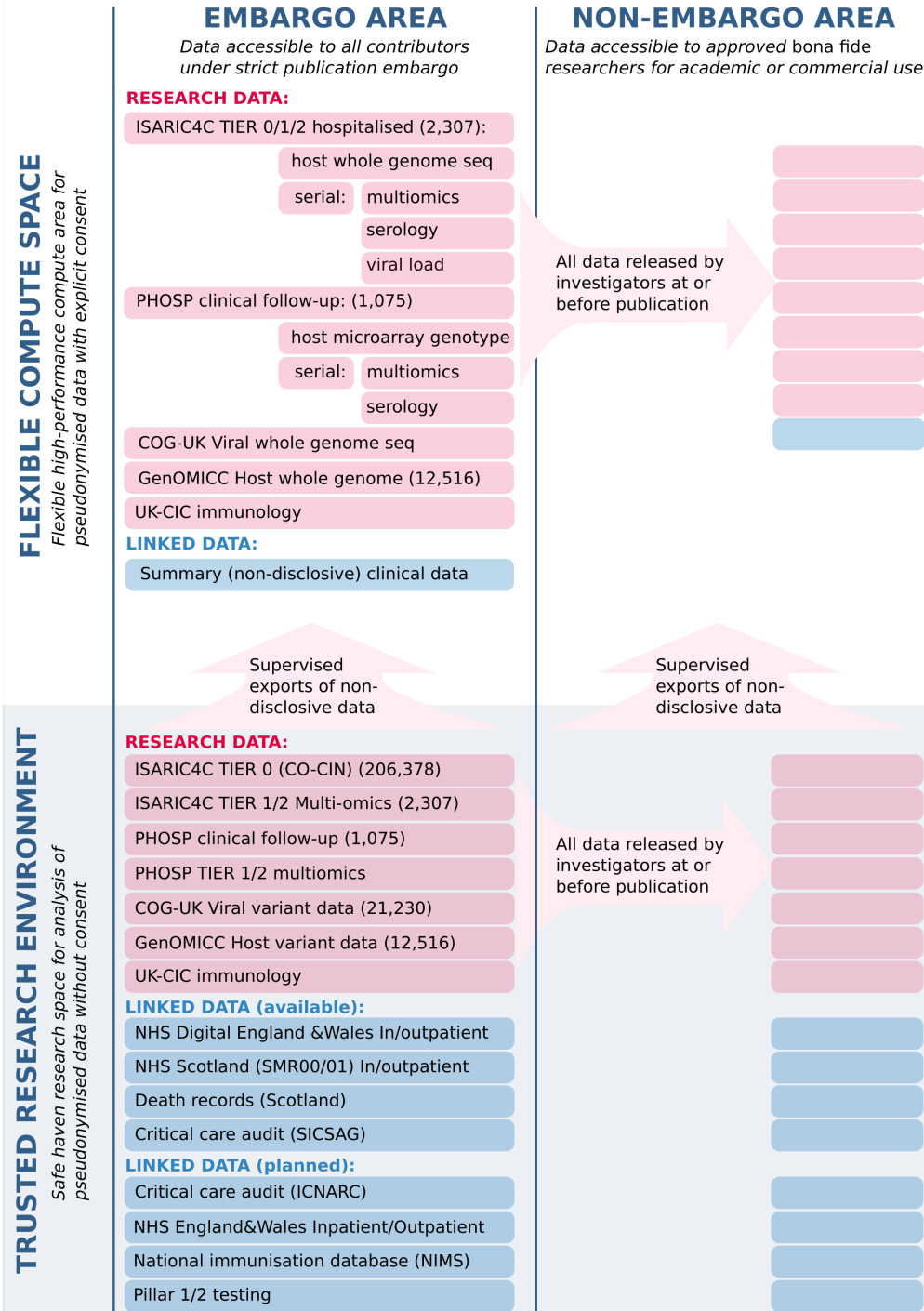All data released by investigators at or before publication

Figure 1: Structure of the ISARIC4C Analysis Platform

Analysis platform structure

There are two routes of access to the analysis platform (Figure 1): 1. NHS Trusted Research Environment (Safe Haven) for access to personal clinical data and data collected without explicit consent. 2. Rapid-access flexible compute for access to non-disclosive research data collected with explicit consent.

Within both of these environments there is an additional division in the data: 1. Publishable "open access" data which any user can use and report as they wish, according to data protection and privacy rules; 2. Embargoed active research data, shared by academic investigators and available for linked analysis but not for publication without agreement from all contributors.

This design is intended to build trust in order to encourage immediate contributions of research data from academic collaborators.
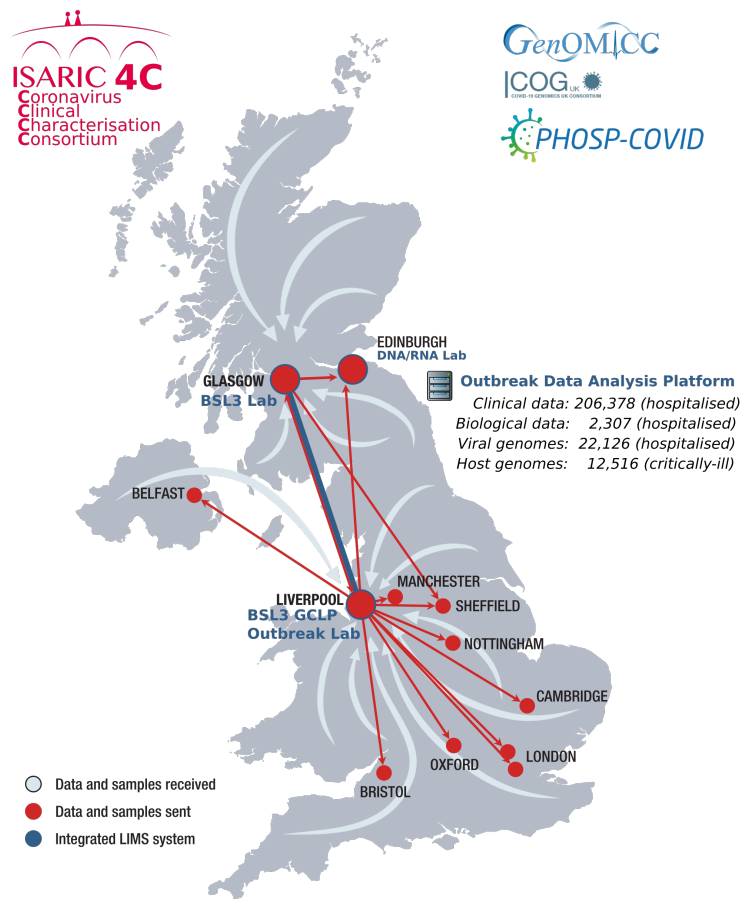

Figure 2: ISARIC4C study and data analysis platform

Future plans

Rapid addition of viral sequence data from the COG-UK platform will enable real-time detection of the clinical impact of new viral strains, in-depth biological study of reinfection, and host:pathogen interactions at a genetic and mechanistic level.

References

1. Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., Merson, L., Lee, J., Plotkin, D., Sigfrid, L., Halpin, S., Jackson, C., Gamble, C., Horby, P.W., Nguyen-Van-Tam, J.S., Ho, A., Russell, C.D., Dunning, J., Openshaw, P.J., Baillie, J.K. & Semple, M.G. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. *BMJ* **369**, (2020).

2. Knight, S.R., Ho, A., Pius, R., Buchan, I., Carson, G., Drake, T.M., Dunning, J., Fairfield, C.J., Gamble, C., Green, C.A., Gupta, R., Halpin, S., Hardwick, H.E., Holden, K.A., Horby, P.W., Jackson, C., Mclean, K.A., Merson, L., Nguyen-Van-Tam, J.S., Norman, L., Noursadeghi, M., Olliaro, P.L., Pritchard, M.G., Russell, C.D., Shaw, C.A., Sheikh, A., Solomon, T., Sudlow, C., Swann, O.V., Turtle, L.C., Openshaw, P.J., Baillie, J.K., Semple, M.G., Docherty, A.B. & Harrison, E.M. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: Development and validation of the 4C mortality score. *BMJ (Clinical research ed.)* **370**, m3339(2020).

3. Thwaites, R.S., Uruchurtu, A.S.S., Siggins, M.K., Liew, F., Russell, C.D., Moore, S.C., Fairfield, C., Carter, E., Abrams, S., Short, C.-E., Thaventhiran, T., Bergstrom, E., Gardener, Z., Ascough, S., Chiu, C., Docherty, A.B., Hunt, D., Crow, Y.J., Solomon, T., Taylor, G.P., Turtle, L., Harrison, E.M., Dunning, J., Semple, M.G., Baillie, J.K., Openshaw, P.J. & Investigators**, on behalf of the I. Inflammatory profiles across the spectrum of disease reveal a distinct role for GM-CSF in severe COVID-19. *Science Immunology* **6**, (2021).

4. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., Furniss, J., Richmond, A., Gountouna, E., Wrobel, N., Harrison, D., Wang, B., Wu, Y., Meynert, A., Griffiths, F., Oosthuyzen, W., Kousathanas, A., Moutsianas, L., Yang, Z., Zhai, R., Zheng, C., Grimes, G., Beale, R., Millar, J., Shih, B., Keating, S., Zechner, M., Haley, C., Porteous, D.J., Hayward, C., Yang, J., Knight, J., Summers, C., Shankar-Hari, M., Klenerman, P., Turtle, L., Ho, A., Moore, S.C., Hinds, C., Horby, P., Nichol, A., Maslove, D., Ling, L., McAuley, D., Montgomery, H., Walsh, T., Pereira, A.C., Renieri, A., Shen, X., Ponting, C.P., Fawkes, A., Tenesa, A., Caulfield, M., Scott, R., Rowan, K., Murphy, L., Openshaw, P.J.M., Semple, M.G., Law, A., Vitart, V., Wilson, J.F. & Baillie, J.K. Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98(2021).