Public Health Genomics

Professor Tom Connor, Head of Public Health Genomics Programme

KRIC, June 2023

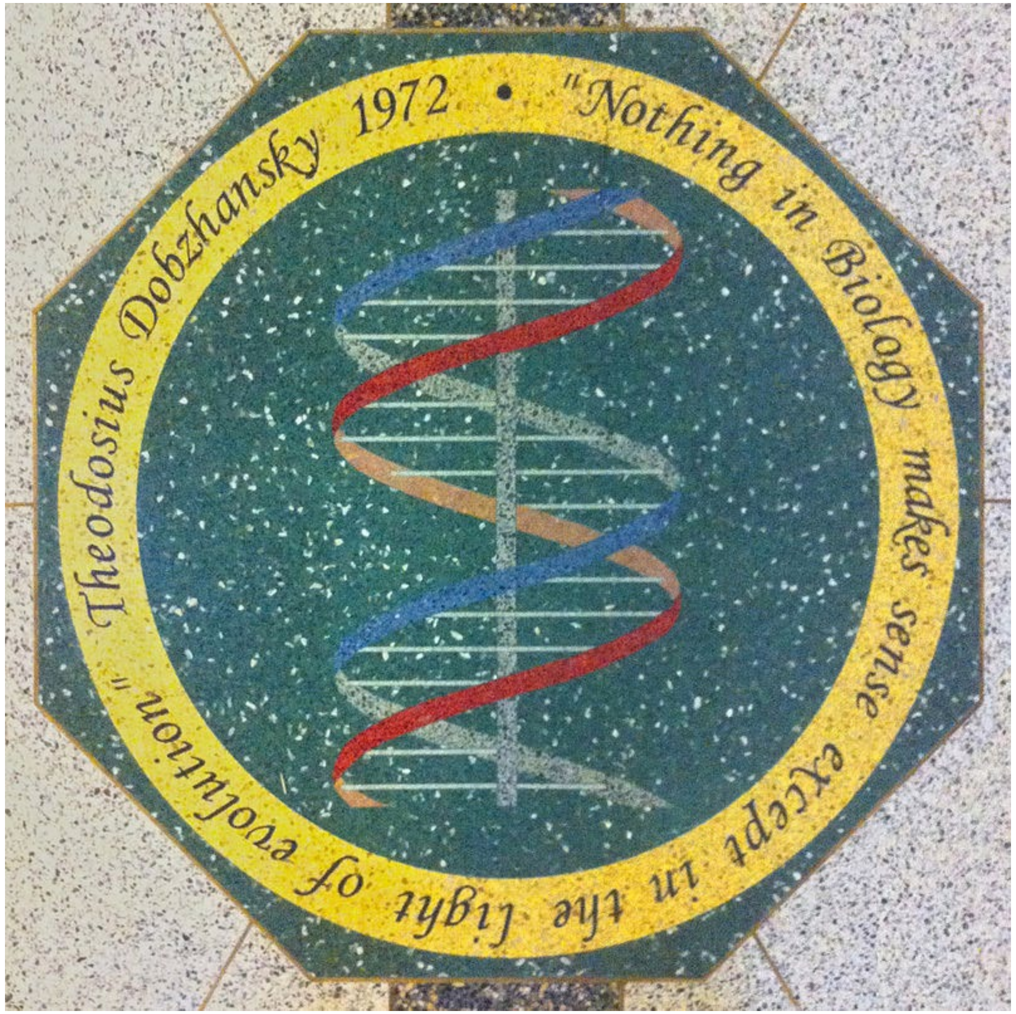
GIG CYMRU NHS WALES | Iechyd Cyhoeddus Cymru
Public Health Wales

Partneriaeth Genomeg Cymru
Genomics Partnership Wales

Llywodraeth Cymru
Welsh Government

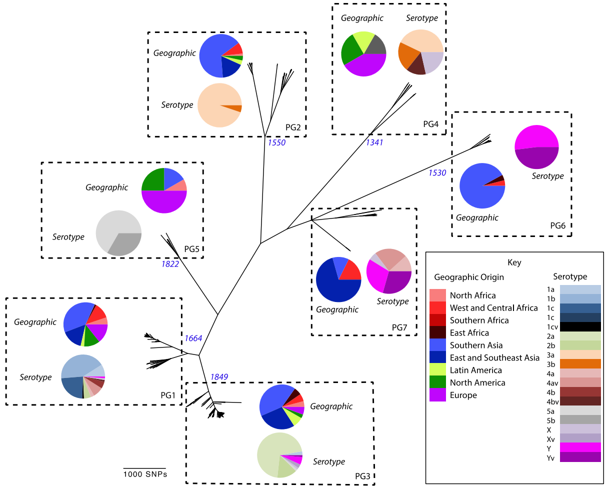
Overview

- Introduce genomics
- Introduce elements of the genomics ecosystem in PHW and more widely in Wales
- Introduce the Public Health Genomics Programme and the potential of genomics
- Introduce our emerging plans and vision

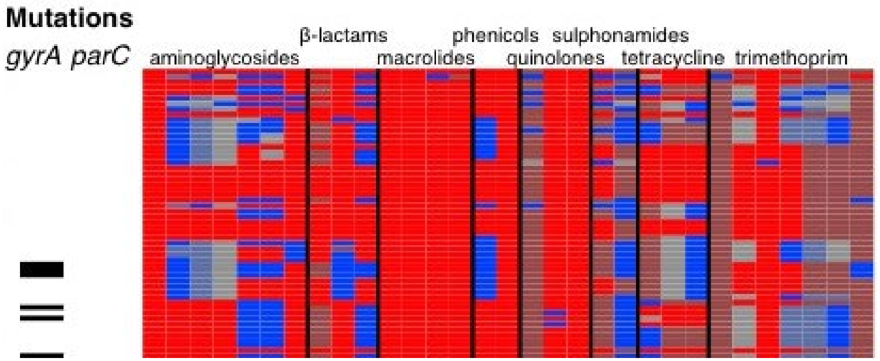


What is genomics?

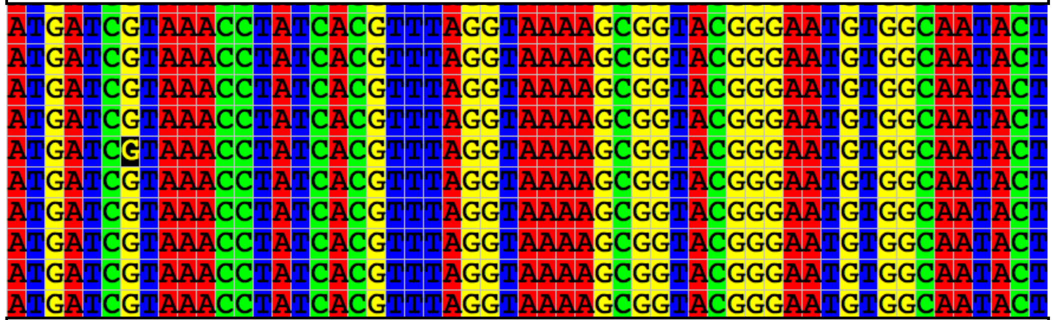
Put simply, genomics is the branch of biology that is concerned with studying genomes



Sequencing instruments read DNA, and by extension, enable us to read the genome of an organism of interest



However, what really defines the field isnt the use of sequencing instruments



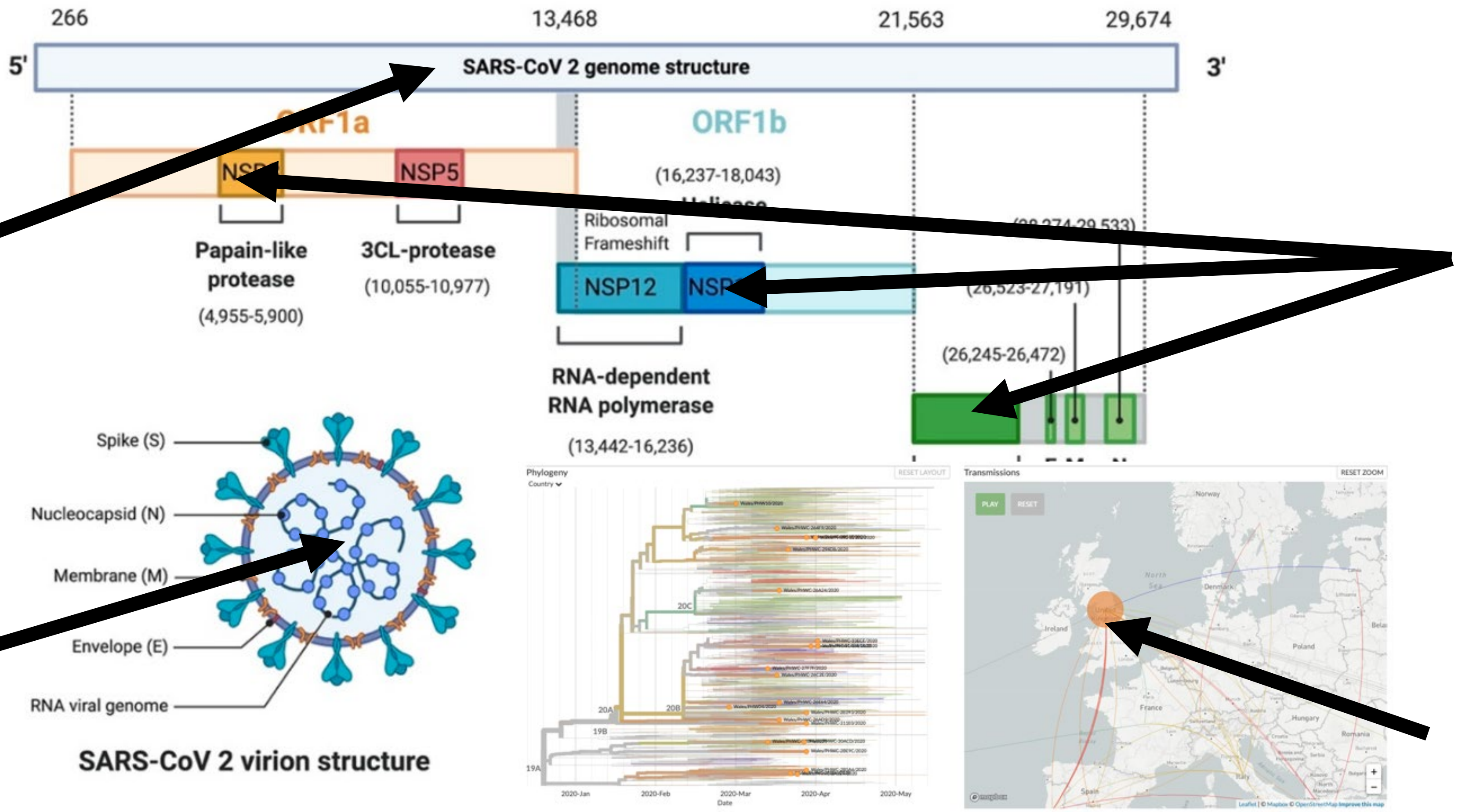
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT
ATGATCGTAAACCTATCACGTTTAGGTAAAAGCGGTACGGGAATGTGGCAATACT

It is the data – and the type of questions we can ask, and answer

What can genomics do?

Study the evolution of the whole organism

Track changes over time



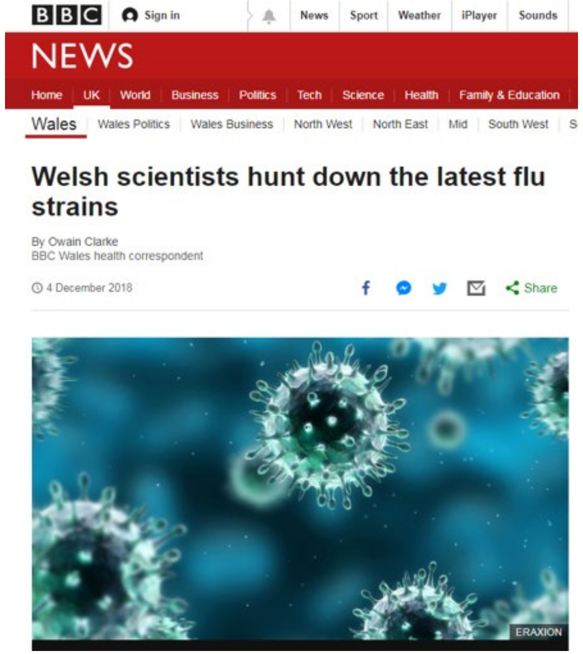
Characterise individual parts that make up an organism

Unpick the spread of the organism in time and space (see data at https://nextstrain.org/)

Image from Alanagreh et al, Pathogens ***2020****, ;* [***https://doi.org/10.3390/pathogens9050331***](https://doi.org/10.3390/pathogens9050331)

Why is genomics important?

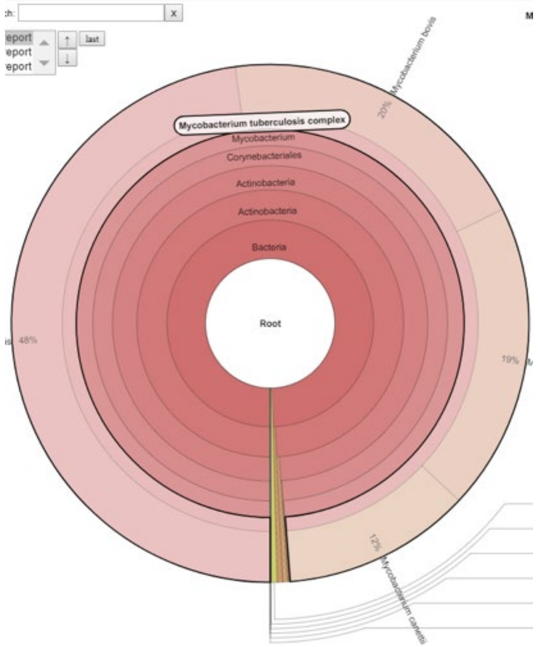
High throughput =
Faster turnaround Times



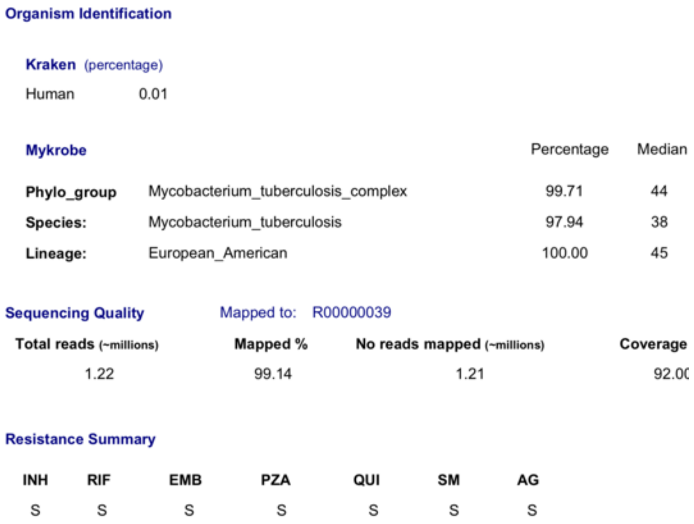
Welsh scientists hunt down the latest flu strains

By Owain Clarke
BBC Wales health correspondent

4 December 2018



Unambiguous
= clearer
diagnostics



Organism Identification

Kraken (percentage)

Human	0.01

Mykrobe

		Percentage	Median
Phylo_group	Mycobacterium_tuberculosis_complex	99.71	44
Species:	Mycobacterium_tuberculosis	97.94	38
Lineage:	European_American	100.00	45

Sequencing Quality Mapped to: R00000039

Total reads (~millions)	Mapped %	No reads mapped (~millions)	Coverage
1.22	99.14	1.21	92.0

Resistance Summary

INH	RIF	EMB	PZA	QUI	SM	AG
S	S	S	S	S	S	S

Whole Genome = More
tests possible
simultaneously

Common data types =
Diagnostic data can be used
for surveillance in real time

Partneriaeth Genomeg Cymru
Genomics Partnership Wales

GIG CYMRU NHS WALES
Iechyd Cyhoeddus Cymru
Public Health Wales

PHWs genomics journey to date

31.03.2016
NGS equipment arrived (3 MiSeqs)

01.10.2016
Generic NGS workflow development complete

12.04.2017
Genomics for Precision Medicine strategy published

14.02.2018
PenGU launched

26.04.2018
GridION installation and training

23.07.2018
Repatriation of HIV service

1.10.2018
Exophialia nanopore analysis

5.10.2018
First flu genome of new surveillance service

21.11.2018
Enterovirus D68 WGS

07.01.2019
Mycobacterial service online

01.08.2019
C. difficile service parallel running

22.01.2020
HIV and TB services accredited

6.3.2020
First Welsh SARS-CoV-2 genome sequenced

31.03.2020
Replacement of 2 MiSeqs with a NextSeq

1.4.2020
COG-UK setup

26.06.2020
SARS-CoV-2 Outbreak support service starts

10.11.2020
C. difficile service accreditation visit

12.2020
20,000th SARS-CoV-2 genome sequenced

1.2021
C. difficile service accredited

15.03.2021
addition of 2 NextSeqs

7.2021
50,000th SARS-CoV-2 genome sequenced

2.2022
SARS-CoV-2 service accredited

3.2022
200,000th SARS-CoV-2 genome sequenced

Partneriaeth Genomeg Cymru
Genomics Partnership Wales

GIG CYMRU
NHS WALES
Iechyd Cyhoeddus Cymru
Public Health Wales

In summary....

How it started

HIV | WCM

Lab | Bioinformatics

7 lab staff, 1x 40% FTE bioinformatician
2,000 samples/year

How its going

HIV | WCM | Influenza

C. difficile | SARS-CoV-2

AMR | Enterovirus | CF/16S

Lab | Bioinformatics | Research

Health Protection | Epidemiology | Surveillance

14 lab staff, 8 bioinformaticians, over 50 people involved across PHW
Capacity of 8,000+ samples/week

The Wider Welsh context – Genomics Partnership Wales

Established in 2017 to deliver the Welsh Government Genomics for Precision Medicine Strategy

Working together to harness the potential of genomics to improve the health, wellbeing and prosperity of the people of Wales

Welsh Government investments

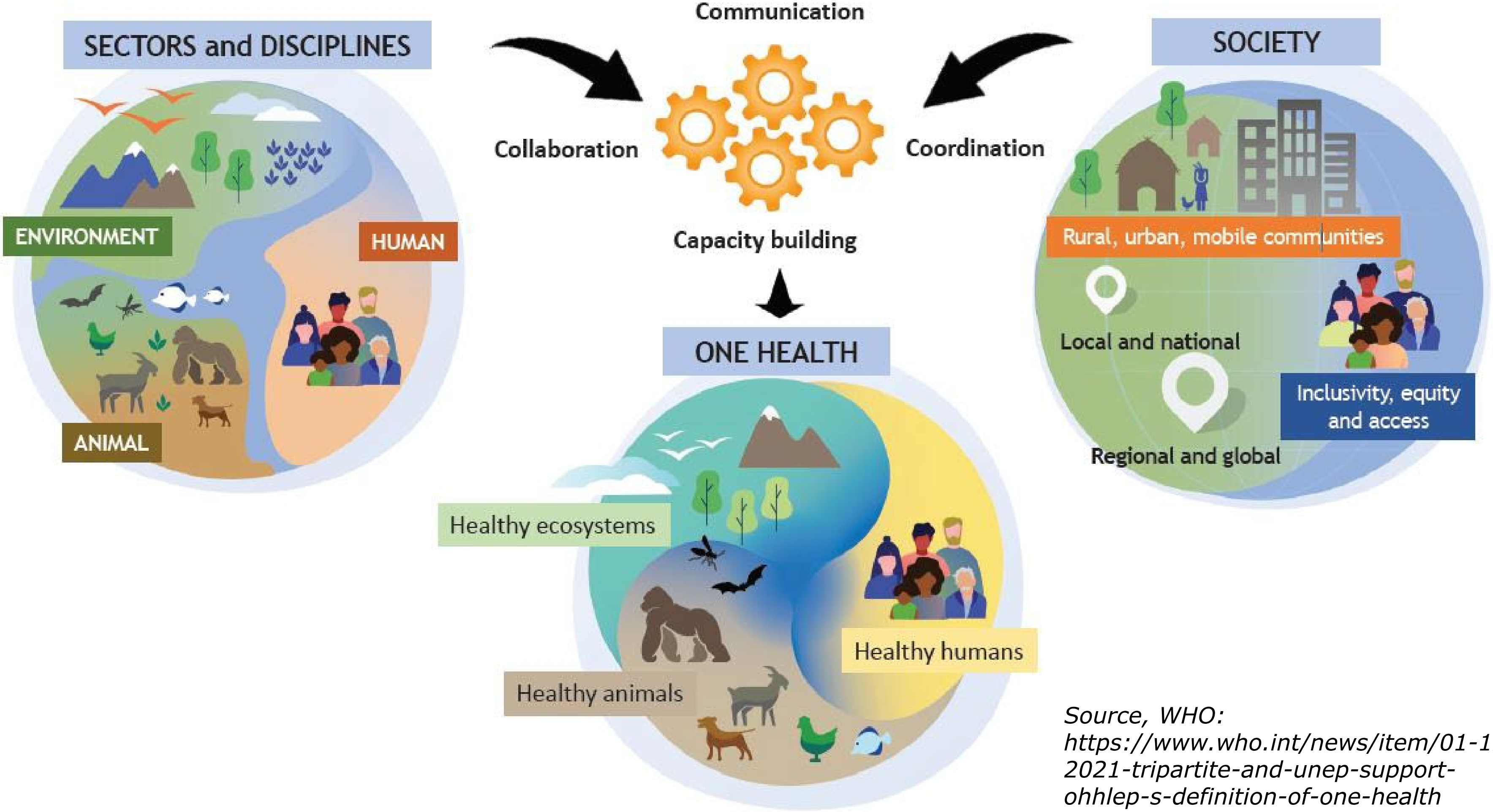
- £16M investment to create a Centre of Excellence for Precision Medicine
- Will see co-location of human and pathogen genomics services, plus Cardiff University
  - Build complete, move due in November
  - Many opportunities

Genomics to enable precision public health

Genomics potentially impacts most areas of public health activity, and we are already working across areas and functions

GPW gives us a structure, and brings together capacity and data

Genomics enables ”the right intervention to the right population at the right time”



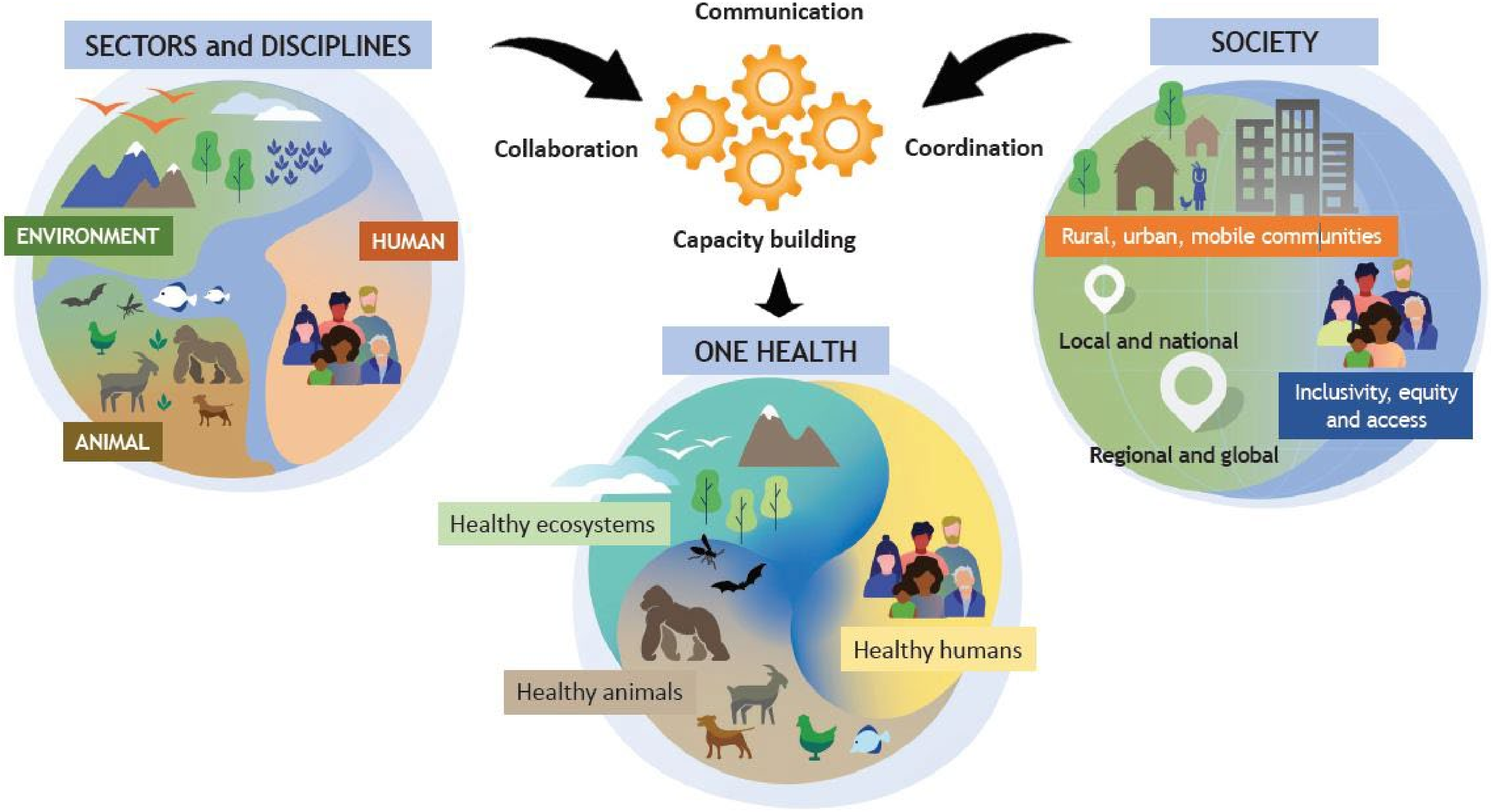
Source, WHO: https://www.who.int/news/item/01-12-2021-tripartite-and-unep-support-ohhlep-s-definition-of-one-health

Genomics to enable precision public health

Everything we want to do is also bound up in the ONE health concept

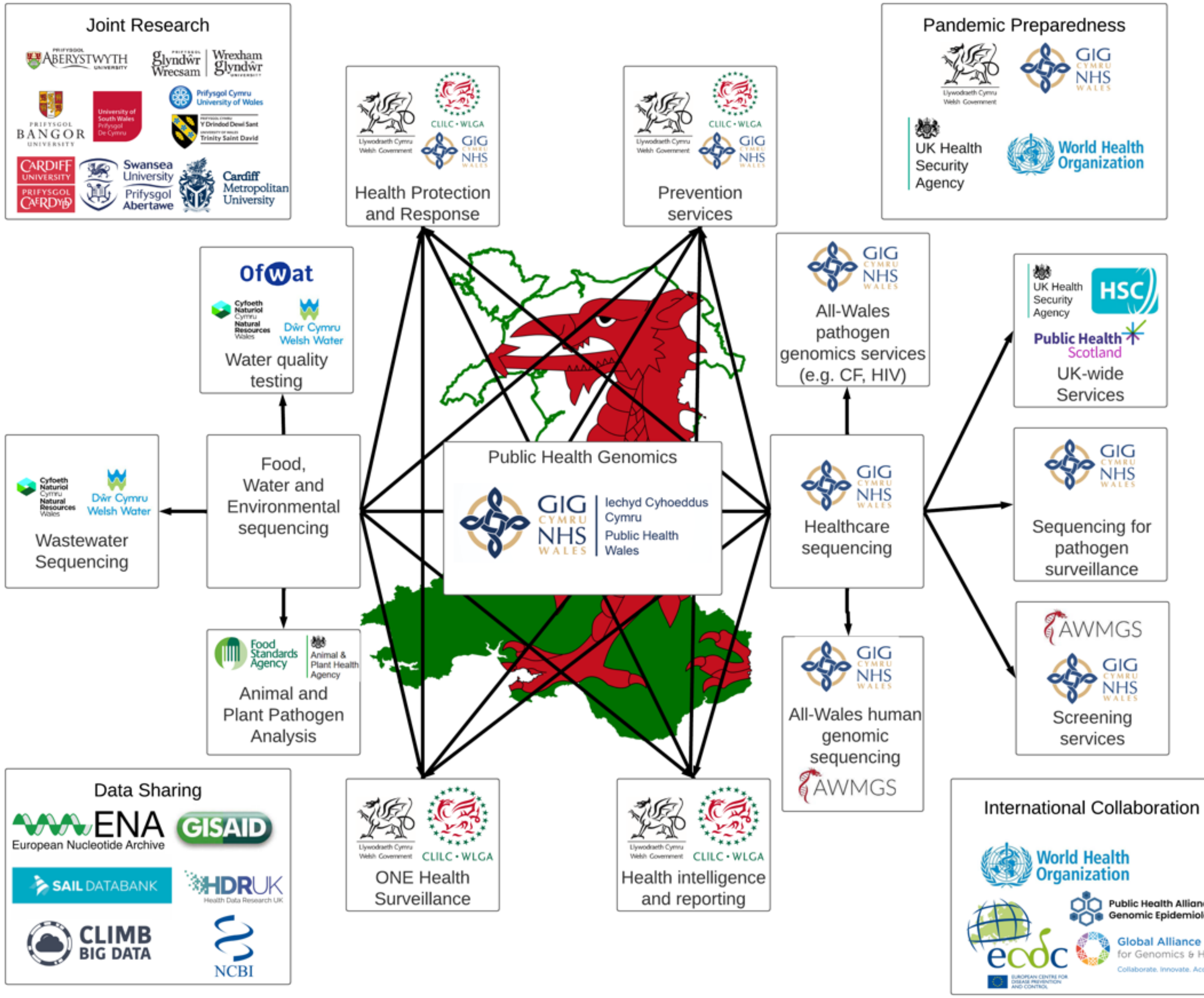
To provide effective public health services, we have to understand the interface of humans, pathogens and the environment

Genomics is a key tool to do this



Building blocks

- We sit at the heart of a maturing genomics ecosystem in Wales.
- Data generation and exploitation and significant infrastructure.
- Moving into a new phase, seeing increasing adoption of genomics across Wales, and increased data generation.



Joint Research

Health Protection and Response

Prevention services

Pandemic Preparedness

Water quality testing

All-Wales pathogen genomics services (e.g. CF, HIV)

UK-wide Services

Wastewater Sequencing

Food, Water and Environmental sequencing

Public Health Genomics

Healthcare sequencing

Sequencing for pathogen surveillance

Animal and Plant Pathogen Analysis

All-Wales human genomic sequencing

Screening services

Data Sharing

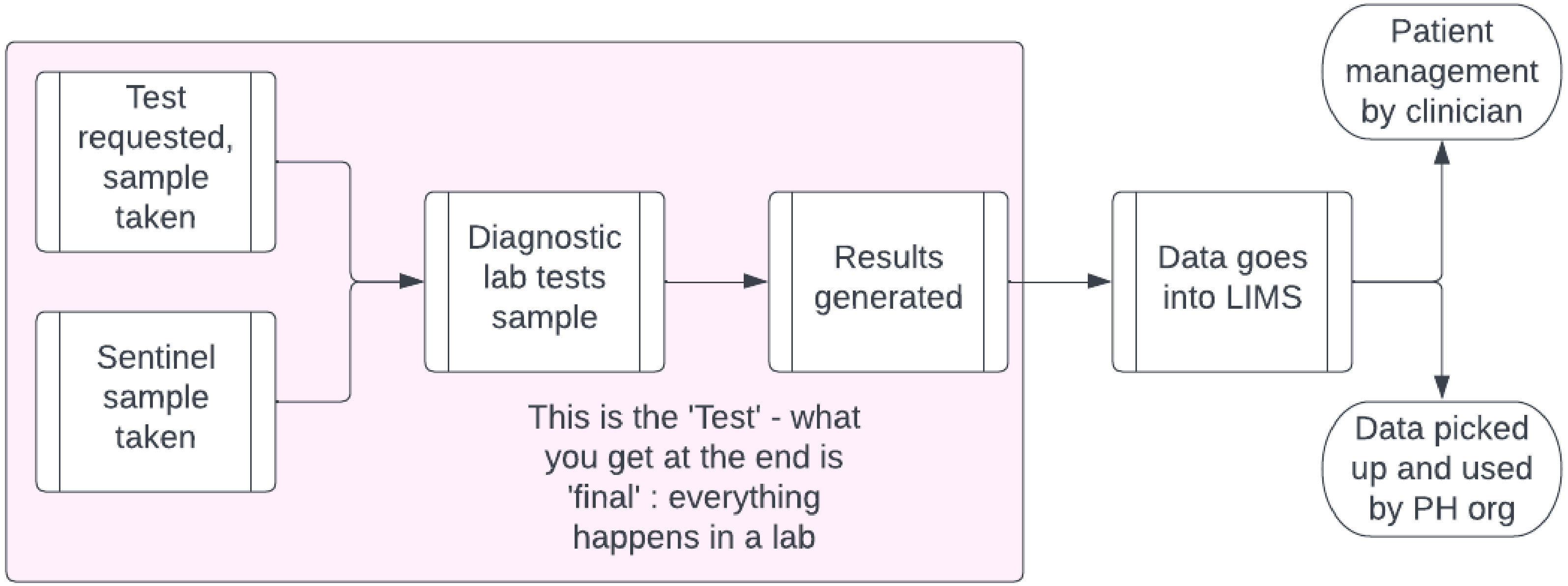
ONE Health Surveillance

Health intelligence and reporting

International Collaboration

Genomics benefits – A traditional view of testing and surveillance

- Surveillance data (and official stats etc) are derived from diagnostic tests
- Some sentinel schemes, but largely depend on the tests that are done in the diagnostic lab(s)



Test requested, sample taken

Sentinel sample taken

Diagnostic lab tests sample

Results generated

This is the 'Test' - what you get at the end is 'final' : everything happens in a lab

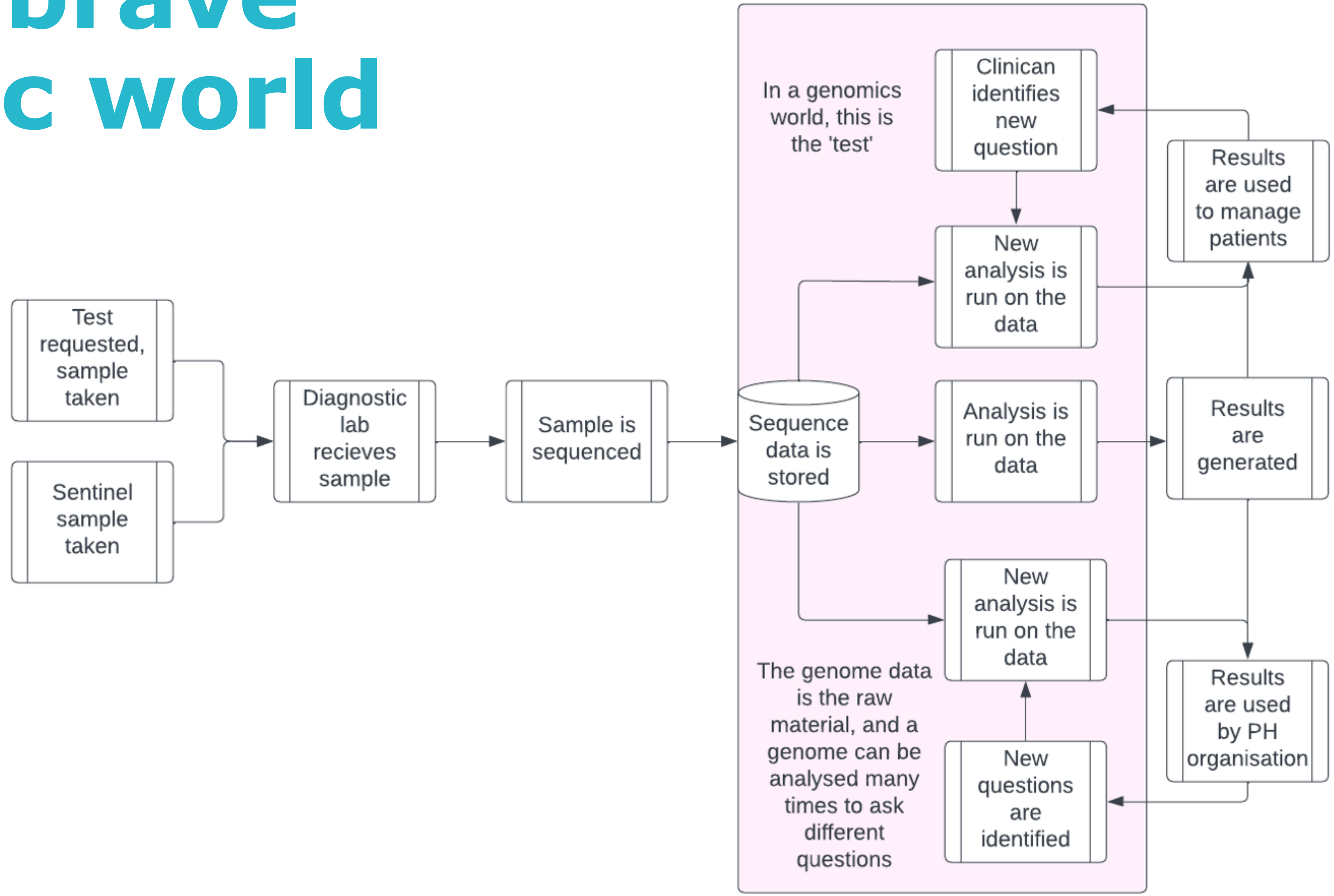
Data goes into LIMS

Patient management by clinician

Data picked up and used by PH org

Benefits – a brave new Genomic world

- In genomics, the tests are digital
- The raw input (genome sequences) are generated once, and can be analysed many times and in different ways



Test requested, sample taken

Sentinel sample taken

Diagnostic lab recieves sample

Sample is sequenced

Sequence data is stored

In a genomics world, this is the 'test'

Clinican identifies new question

New analysis is run on the data

Analysis is run on the data

New analysis is run on the data

New questions are identified

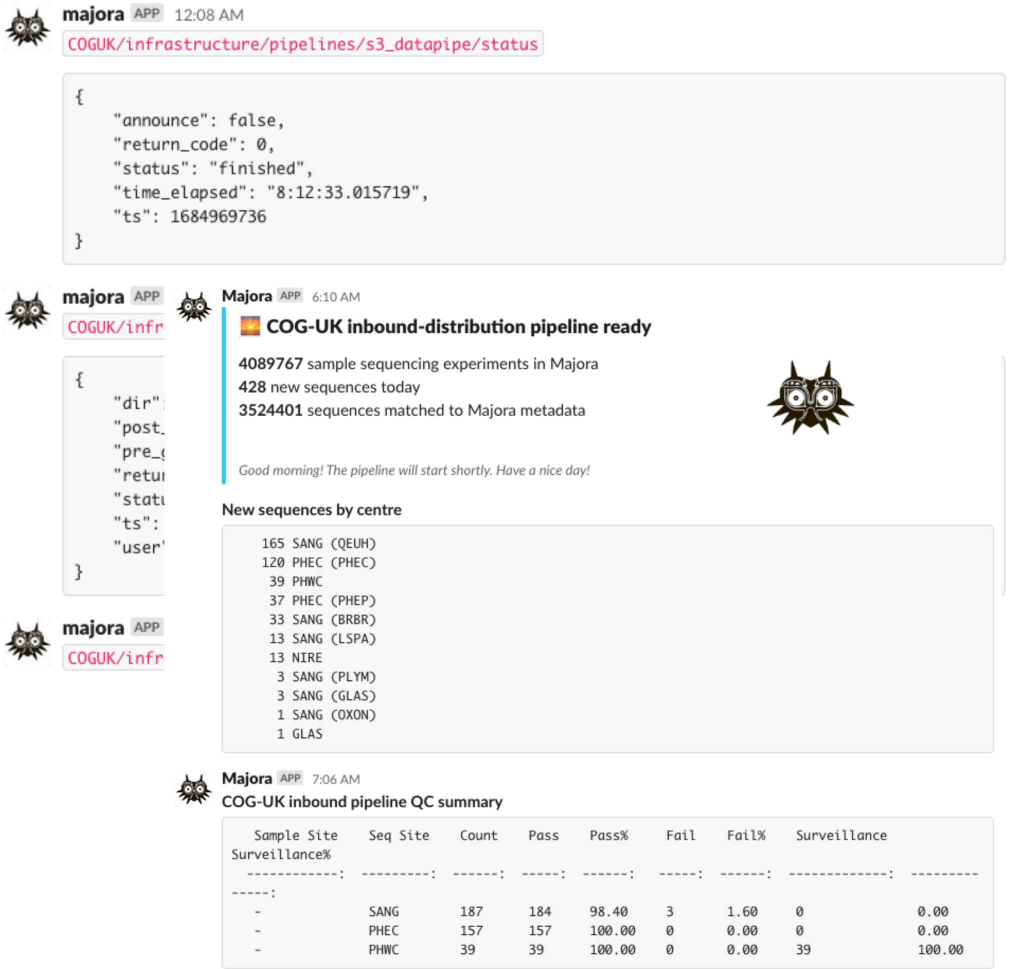
The genome data is the raw material, and a genome can be analysed many times to ask different questions

Results are used to manage patients

Results are generated

Results are used by PH organisation

Designing real-time systems



majora APP 12:08 AM

`COGUK/infrastructure/pipelines/s3_datapipe/status`

```
{
    "announce": false,
    "return_code": 0,
    "status": "finished",
    "time_elapsed": "8:12:33.015719",
    "ts": 1684969736
}
```

majora APP

`COGUK/infr`

```
{
    "dir"
    "post_
    "pre_
    "retu
    "statu
    "ts":
    "user'
}
```

majora APP

`COGUK/infr`

Majora APP 6:10 AM

🌅 COG-UK inbound-distribution pipeline ready

4089767 sample sequencing experiments in Majora
428 new sequences today
3524401 sequences matched to Majora metadata

Good morning! The pipeline will start shortly. Have a nice day!

New sequences by centre

```
  165 SANG (QEUH)
  120 PHEC (PHEC)
   39 PHWC
   37 PHEC (PHEP)
   33 SANG (BRBR)
   13 SANG (LSPA)
   13 NIRE
    3 SANG (PLYM)
    3 SANG (GLAS)
    1 SANG (OXON)
    1 GLAS
```

Majora APP 7:06 AM

COG-UK inbound pipeline QC summary

```
  Sample Site    Seq Site    Count    Pass    Pass%    Fail    Fail%    Surveillance
Surveillance%
  -----------:  ---------:  ------:  -----:  ------:  -----:  ------:  ------------:  ---------
-----:
  -              SANG        187      184     98.40    3       1.60     0               0.00
  -              PHEC        157      157     100.00   0       0.00     0               0.00
  -              PHWC        39       39      100.00   0       0.00     39              100.00
```

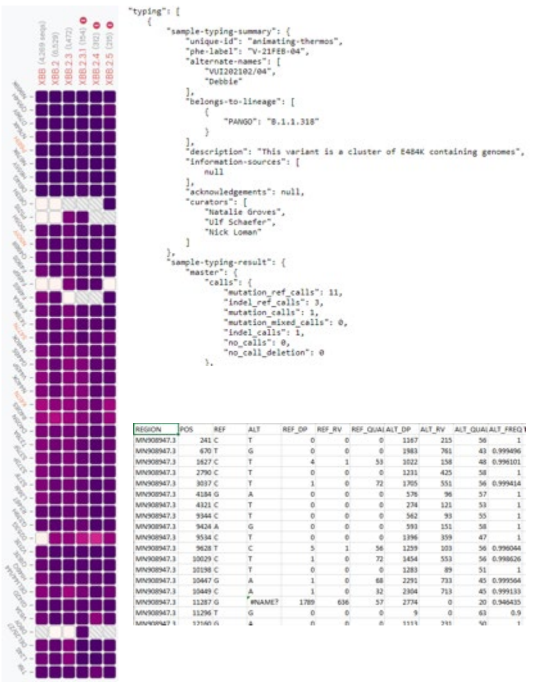
- Genomics data is digital, and so is always generated and used in ‘real-time’
- Means that you can design systems to enable that data to be used in many places in real time
- Opportunities in terms of messaging queues / designing systems that are event-driven

Implications

- The fabric of the genomics enterprise is data and its analysis by domain specialists (bioinformaticians).
- The same genomics data generated for one purpose could be used for many different public health purposes, just by creating a new analysis.
  - Surveillance, official stats etc using genomics data would no longer have to just 'take what they are given', or have limited resolution.
  - We (PHW, researchers, industry) can reuse data and maximise its value, asking deeper, more complex questions that can drive significant change.
- Routine generation of genomics data in the NHS is creating a huge, rich, complex dataset, that can be mined for insight.
  - This is a bioinformatics/data science challenge.
  - This is an IG/data sharing challenge

Potential: from patient to population

Clinicians

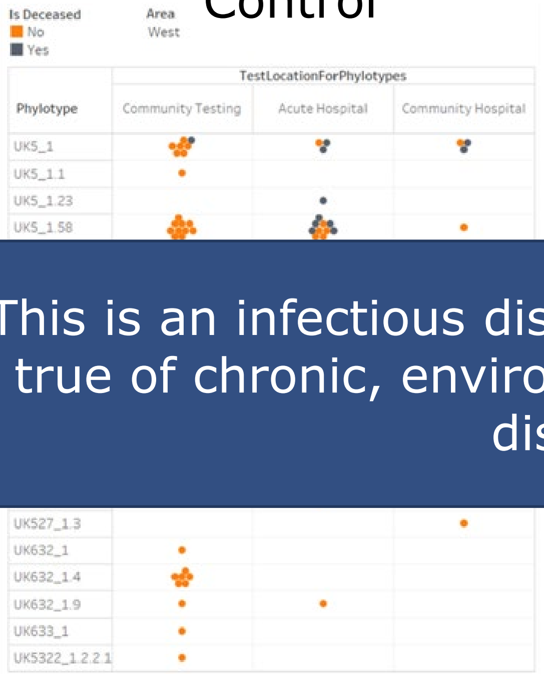


```
"typing": [
    {
        "sample-typing-summary": {
            "unique-id": "animating-thermos",
            "phe-label": "V-21FEB-04",
            "alternate-names": [
                "VUI202102/04",
                "Debbie"
            ],
            "belongs-to-lineage": [
                {
                    "PANGO": "B.1.1.318"
                }
            ],
            "description": "This variant is a cluster of E484K containing genomes",
            "information-sources": [
                null
            ],
            "acknowledgements": null,
            "curators": [
                "Natalie Groves",
                "Ulf Schaefer",
                "Nick Loman"
            ]
        },
        "sample-typing-result": {
            "master": {
                "calls": {
                    "mutation_ref_calls": 11,
                    "indel_ref_calls": 3,
                    "mutation_calls": 1,
                    "mutation_mixed_calls": 0,
                    "indel_calls": 1,
                    "no_calls": 0,
                    "no_call_deletion": 0
                },
```

REGION	POS	REF	ALT	REF_DP	REF_RV	REF_QUAL	ALT_DP	ALT_RV	ALT_QUAL	ALT_FREQ
MN908947.3	241	C	T	0	0	0	1167	215	56	1
MN908947.3	670	T	G	0	0	0	1983	761	43	0.999496
MN908947.3	1627	C	T	4	1	53	1022	158	48	0.996101
MN908947.3	2790	C	T	0	0	0	1231	425	58	1
MN908947.3	3037	C	T	1	0	72	1705	551	56	0.999414
MN908947.3	4184	G	A	0	0	0	576	96	57	1
MN908947.3	4321	C	T	0	0	0	274	121	53	1
MN908947.3	9344	C	T	0	0	0	562	93	55	1
MN908947.3	9424	A	G	0	0	0	593	151	58	1
MN908947.3	9534	C	T	0	0	0	1396	359	47	1
MN908947.3	9628	T	C	5	1	56	1259	103	56	0.996044
MN908947.3	10029	C	T	1	0	72	1454	553	56	0.998626
MN908947.3	10198	C	T	0	0	0	1283	89	51	1
MN908947.3	10447	G	A	1	0	68	2291	733	45	0.999564
MN908947.3	10449	C	A	1	0	32	2304	713	45	0.999133
MN908947.3	11287	G	#NAME?	1789	636	57	2774	0	20	0.946435
MN908947.3	11296	T	G	0	0	0	9	0	63	0.9

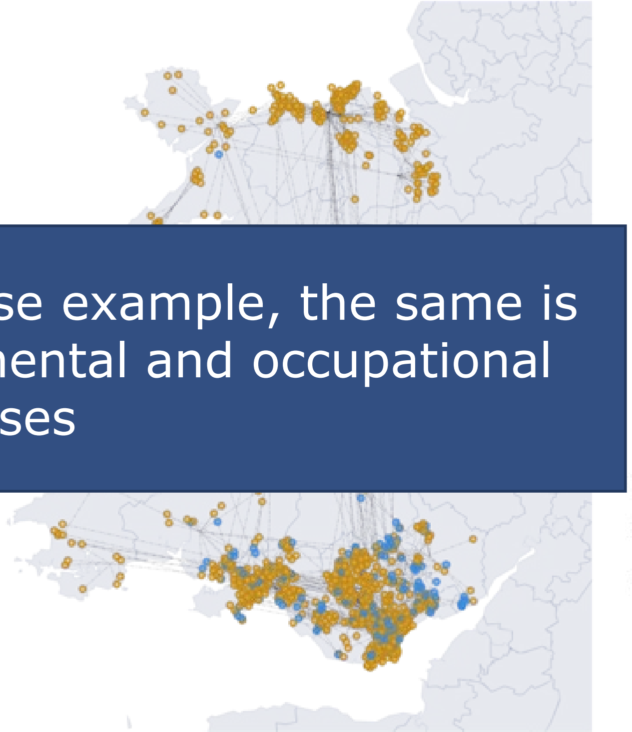
Diagnostics / patient level

Infection Prevention & Control



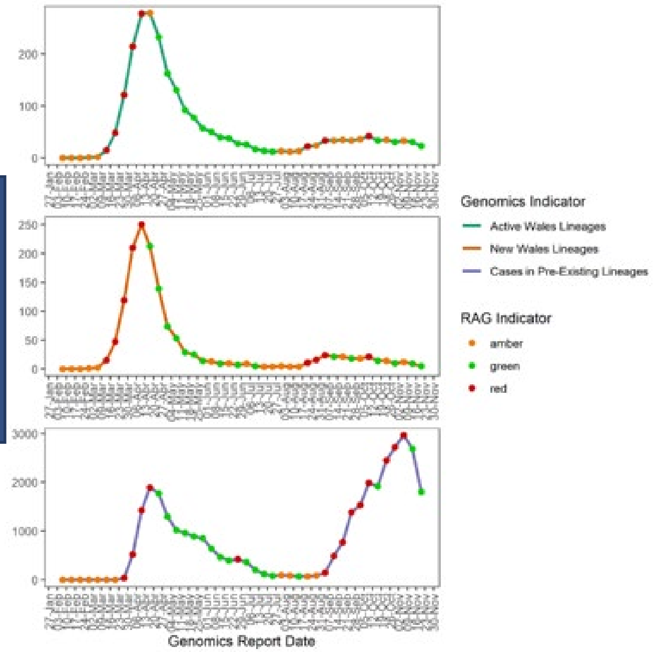
Hospital/local Outbreaks

Surveillance / Health Protection



Regional/national Surveillance/action

Policymakers and Government



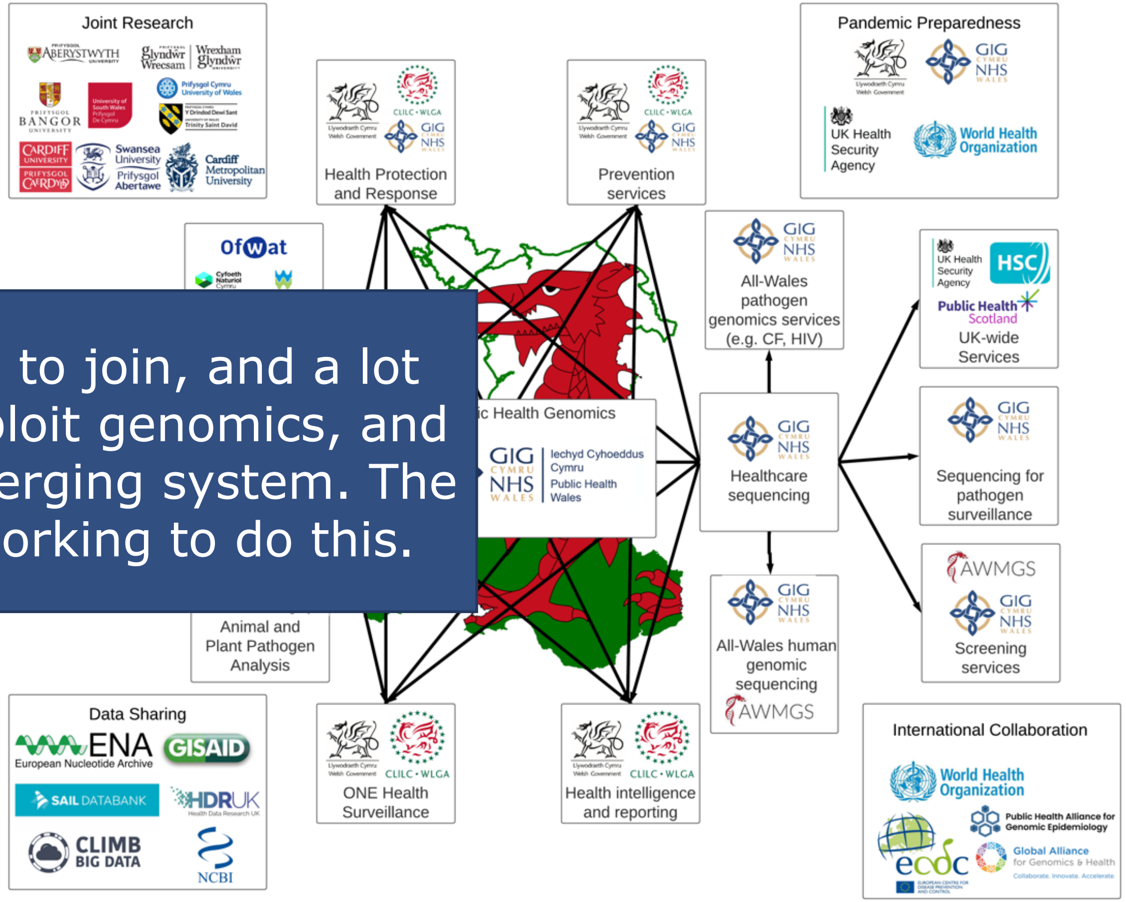
National response and policy

This is an infectious disease example, the same is true of chronic, environmental and occupational diseases

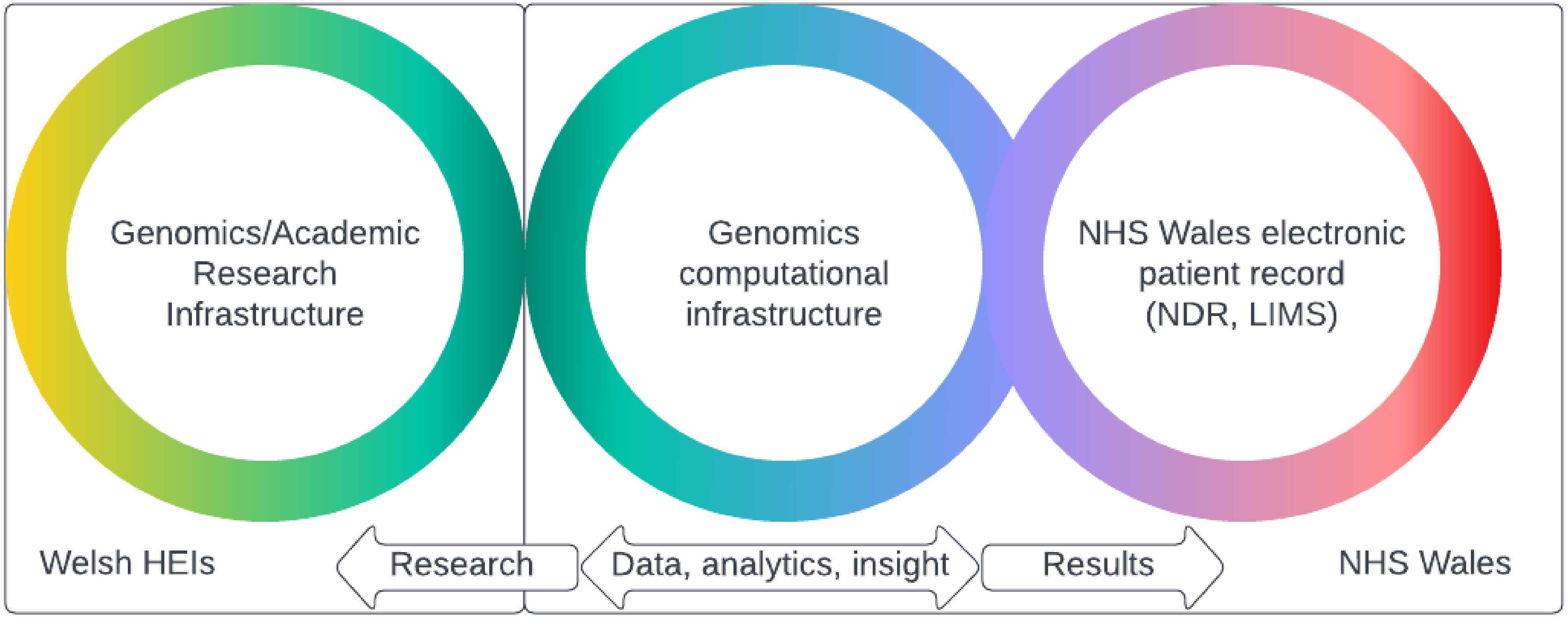
Development of a Public Health Genomics programme

- *Vision: Driving the development of genomics services from patient to population, to protect and improve health and well-being and reduce health inequali*
- *Role: A system*
  stakeholders f
  UK to identify,
  implementation of genomics practices to prevent and control infectious, (and) chronic, environmental, and occupational diseases in Wales.

Bottom line: there are a lot of dots to join, and a lot we need to do to evolve PHW to exploit genomics, and ensure we are a core part of the emerging system. The PH Genomics Programme will be working to do this.



Key activities: genomics data



Genomics/Academic Research Infrastructure

Genomics computational infrastructure

NHS Wales electronic patient record (NDR, LIMS)

Welsh HEIs

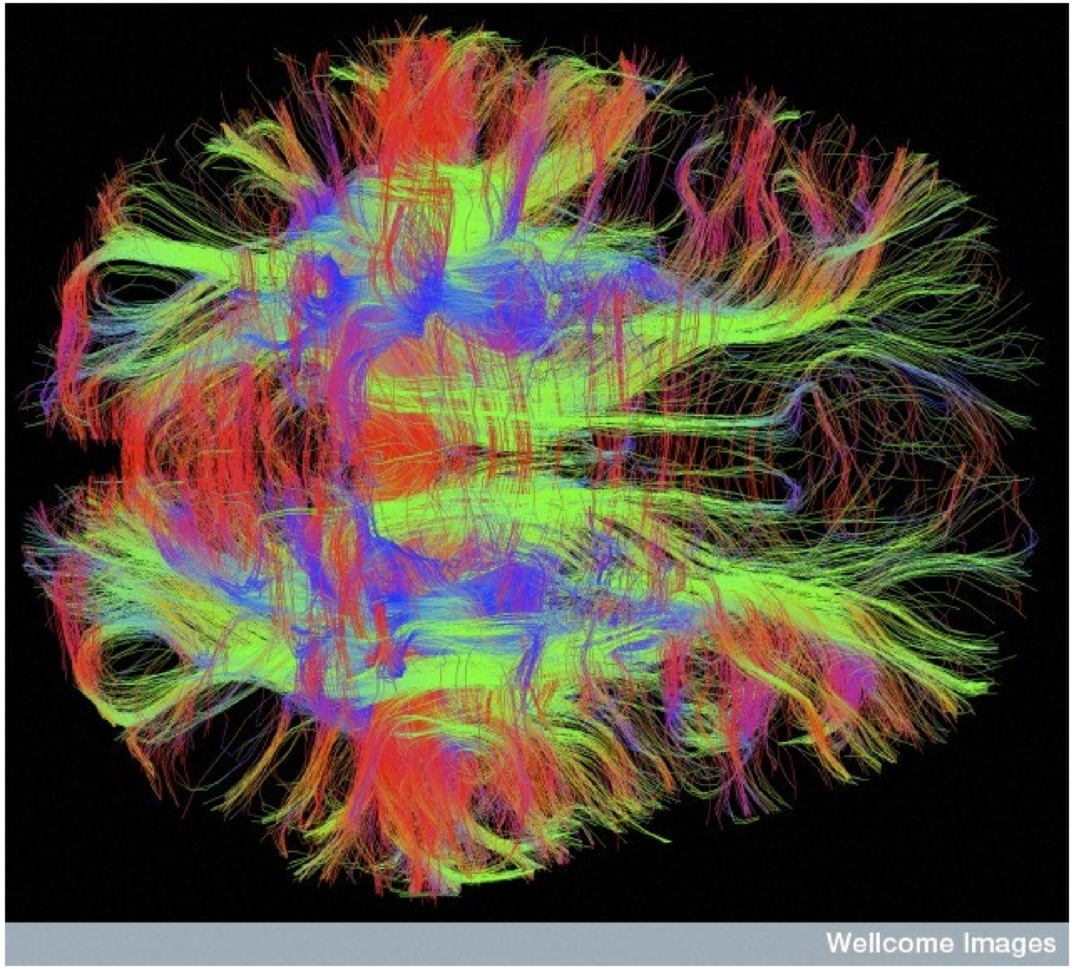
Research

Data, analytics, insight

Results

NHS Wales

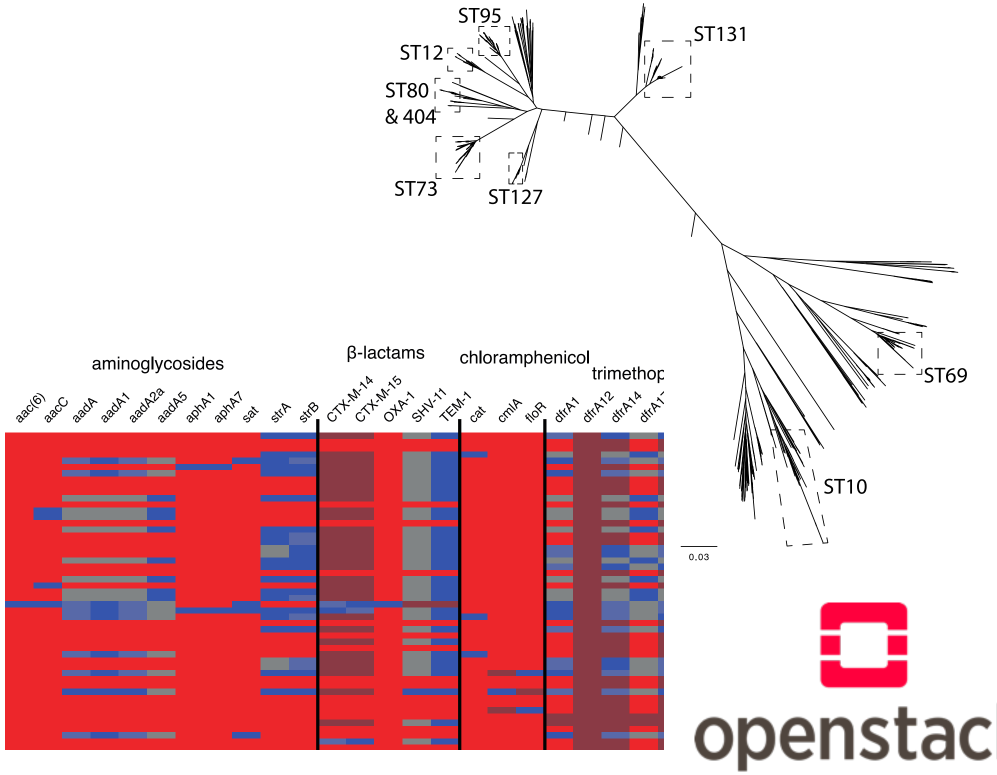
Key activities: infrastructure


Wellcome Images

- Analysing genomics data requires high and high throughput performance computing resources
- We have infrastructure for this, both within PHW/GPW and Wales
- How we specify, access, use and share that infrastructure has PHW and NHS-wide implications
- And without it – we have no service

Key activities: data analysis

- Bioinformatics ; develop the pipelines and interpret results
- Aligned to data science, many common skills, and undertake common activities
- Working to build this community across PHW
- Working to develop career models for these staff
- Recognise that without them there is no service



jupyter

openstack

ceph

DELL EMC

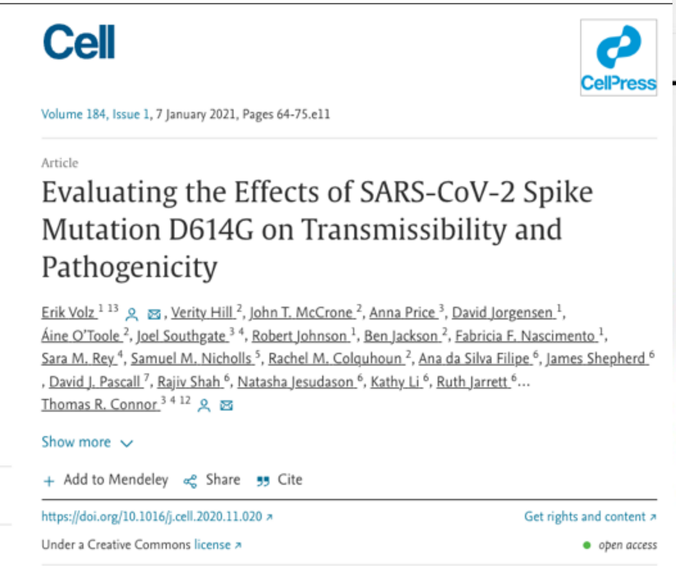
ANSIBLE

slurm
workload manager

nextflow

redhat

Key activities: research



Cell CellPress

Volume 184, Issue 1, 7 January 2021, Pages 64-75.e11

Article

Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity

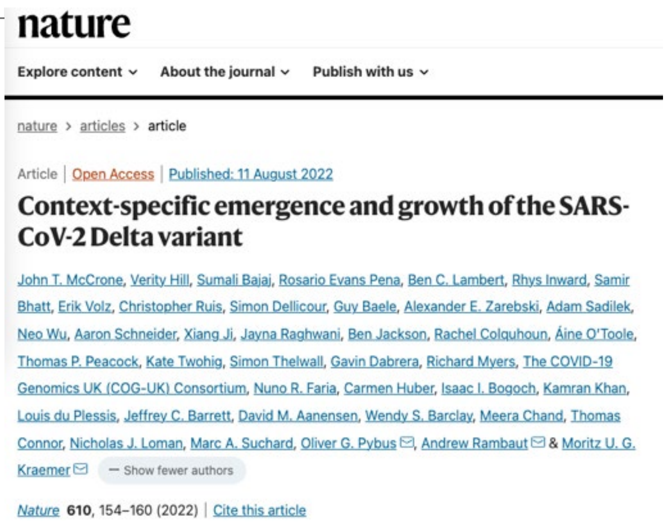
Erik Volz[1,13], Verity Hill[2], John T. McCrone[2], Anna Price[3], David Jorgensen[1], Áine O'Toole[2], Joel Southgate[3,4], Robert Johnson[1], Ben Jackson[2], Fabricia F. Nascimento[1], Sara M. Rey[4], Samuel M. Nicholls[5], Rachel M. Colquhoun[2], Ana da Silva Filipe[6], James Shepherd[6], David J. Pascall[7], Rajiv Shah[6], Natasha Jesudason[6], Kathy Li[6], Ruth Jarrett[6]… Thomas R. Connor[3,4,12]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.cell.2020.11.020 Get rights and content

Under a Creative Commons license open access



nature

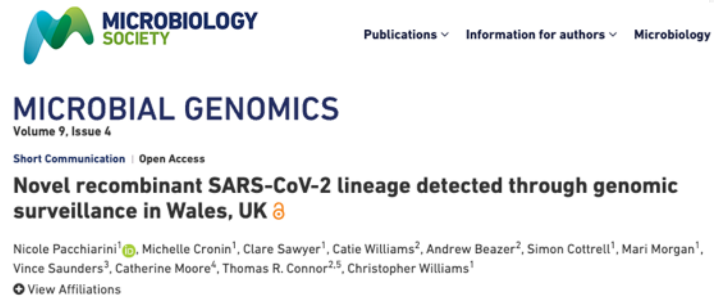
nature > articles > article

Article | Open Access | Published: 11 August 2022

Context-specific emergence and growth of the SARS-CoV-2 Delta variant

John T. McCrone, Verity Hill, Sumali Bajaj, Rosario Evans Pena, Ben C. Lambert, Rhys Inward, Samir Bhatt, Erik Volz, Christopher Ruis, Simon Dellicour, Guy Baele, Alexander E. Zarebski, Adam Sadilek, Neo Wu, Aaron Schneider, Xiang Ji, Jayna Raghwani, Ben Jackson, Rachel Colquhoun, Áine O'Toole, Thomas P. Peacock, Kate Twohig, Simon Thelwall, Gavin Dabrera, Richard Myers, The COVID-19 Genomics UK (COG-UK) Consortium, Nuno R. Faria, Carmen Huber, Isaac I. Bogoch, Kamran Khan, Louis du Plessis, Jeffrey C. Barrett, David M. Aanensen, Wendy S. Barclay, Meera Chand, Thomas Connor, Nicholas J. Loman, Marc A. Suchard, Oliver G. Pybus, Andrew Rambaut & Moritz U. G. Kraemer — Show fewer authors

Nature 610, 154–160 (2022) | Cite this article



MICROBIOLOGY SOCIETY

MICROBIAL GENOMICS
Volume 9, Issue 4

Short Communication | Open Access

Novel recombinant SARS-CoV-2 lineage detected through genomic surveillance in Wales, UK

Nicole Pacchiarini[1], Michelle Cronin[1], Clare Sawyer[1], Catie Williams[2], Andrew Beazer[2], Simon Cottrell[1], Mari Morgan[1], Vince Saunders[3], Catherine Moore[4], Thomas R. Connor[2,5], Christopher Williams[1]

View Affiliations

nature reviews genetics

nature > nature reviews genetics > review articles > article

Review Article | Published: 22 April 2022

Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic

Stephen W. Attwood, Sarah C. Hill, David M. Aanensen, Thomas R. Connor & Oliver G. Pybus

Nature Reviews Genetics 23, 547–562 (2022) | Cite this article

THE JOURNAL OF Hospital Infection — Healthcare Infection Society

RESEARCH ARTICLE | VOLUME 131, P23-33, JANUARY 2023 Download Full Issue

Detailed analysis of in-hospital transmission of SARS-CoV-2 using whole genome sequencing

L. O'Connell † • H. Asad † • G. Hall • … M. Morgan • T.R. Connor • B. Healy • Show all authors • Show footnotes

Published: October 11, 2022 • DOI: https://doi.org/10.1016/j.jhin.2022.09.023 • Check for updates

Eurosurveillance Europe's journal on infectious disease surveillance, epidemiol

Home / Eurosurveillance / Volume 25, Issue 32, 13/Aug/2020 / Article

Rapid communication — Open Access

Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020

Check for updates

Erik Alm[1], Eeva K Broberg[1], Thomas Connor[2,3], Emma B Hodcroft[4], Andrey B Komissarov[5], Sebastian Maurer-Stroh[6,7,8,9], Angeliki Melidou[1], Richard A Neher[4], Áine O'Toole[10], Dmitriy Pereyaslov[11], The WHO European Region sequencing laboratories and GISAID EpiCoV group[12]

- Genomics data is highly valuable for research
- Multiple publications over the course of the COVID-19 pandemic
- COVID-19 data we shared enabled even more papers (over 150 to date)
- Need to build policies and processes, to enable research, share data,, and ensure credit to staff

Partneriaeth Genomeg Cymru
Genomics Partnership Wales

GIG CYMRU NHS WALES
Iechyd Cyhoeddus Cymru
Public Health Wales

Key activities: strategic relationships

- UK-wide genomics landscape
- Welsh relationships, within GPW
- International relationships (PHA4GE, GA4GH, WHO, ECDC)
- New area for us, new area for them
- Lots of challenges

More immediate future: Planning and growth

- Welsh Government Genomics Delivery plan launched in 2022
- Pathogen Genomics Delivery Plan under development now
- Activities relating to non-infectious disease Public Health under way
- Sets the scene for the next 3 years

Llywodraeth Cymru
Welsh Government

GIG CYMRU NHS WALES

Partneriaeth Genomeg Cymru
Genomics Partnership Wales

Genomics Delivery Plan for Wales

2022 – 2025

Delivery plan elements

- Wider plan includes:
  - Pathogen Genomics
  - Human genomics
  - Research
  - Enablers / data
  - Enablers / digital infrastructure
- Implications across PHW: PH Genomics Programme is key point of contact and coordination

Llywodraeth Cymru
Welsh Government

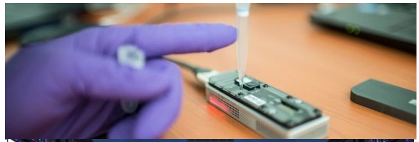
GIG CYMRU
NHS WALES

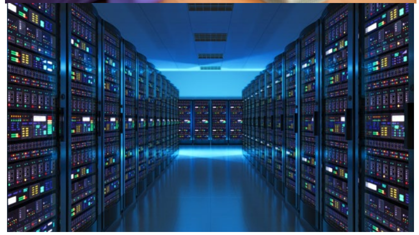
Partneriaeth Genomeg Cymru
Genomics Partnership Wales

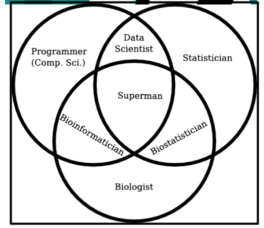
Genomics Delivery Plan for Wales

2022 – 2025

Welsh Public Health Genomics, 2035







Testing:

- Close to patient
- Could be carried out by PHW or others
- Data flows to centralised data analysis/storage
- Genomics-aware workforce
- Bioinformatics and Data Science service

Input

Areas of activity/ use

- Prevention
- Precision Medicine
- Precision Public Health
- Real-time surveillance
- Patient access to data
- ONE Health

Output

Practice

Established Processes

- SOPs for information to action
- Defined best practice
- Data Sharing Agreements
- Threat response processes
- Horizon scanning
- Genomic risk assessment
- Sequencing allocation guidelines

Impact

Improvement:

- Reduced Inequality
- Better outcomes
- Better designed services
- New research
- Patients more engaged/ involved

Image Credit: Alex Cagan

GIG CYMRU NHS WALES | Iechyd Cyhoeddus Cymru Public Health Wales