NIHR | Health Protection Research Unit
in Genomics and Enabling Data
at University of Warwick

Health Protection Research Unit in Genomics and Enabling Data

Xavier Didelot

University of Warwick

30th January 2025

UK Health Security Agency

Centre for Genomic Pathogen Surveillance

THE UNIVERSITY OF WARWICK

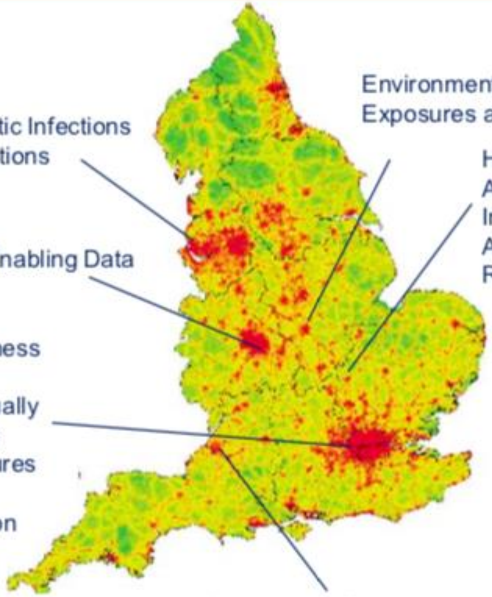
UNIVERSITY OF CAMBRIDGE

NIHR HPRUs - 1 April 2020 - 31 March 2025

14 Units and 1 Development Award funded;

- Environmental Research Units; Infections Research Units and Cross Cutting Units
- 3 new Units and 3 existing Units with new Directors
- New area 'Genomics and Enabling Data'

The aim of the NIHR **HPRU scheme is to support PHE** in **delivering its objectives and functions** for public health protection, including building an evidence base for public health policy and practise;



- Emerging and Zoonotic Infections
- Gastrointestinal Infections

Genomics and Enabling Data

- Emergency Preparedness and Response
- Blood Borne and Sexually Transmitted Infections
- Environmental Exposures and Health
- Chemical and Radiation Threats and Hazards
- Respiratory Infections
- Modelling and Health Economics
- Environmental Change and Health
- Vaccines and Immunisation

Environmental Exposures and Health

Healthcare Associated Infections and Antimicrobial Resistance

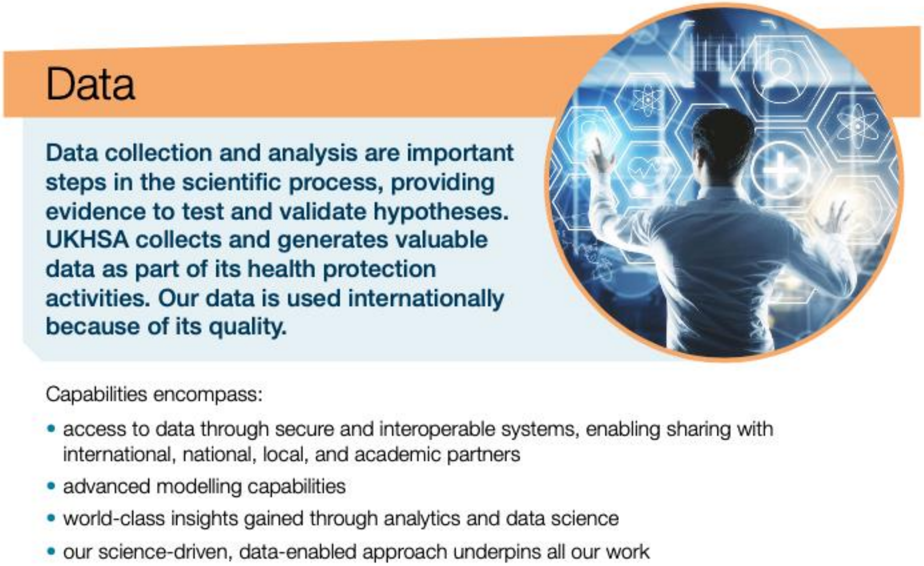
Behavioural Science and Evaluation

NIHR | National Institute for Health Research

UK Health
Security
Agency

UKHSA Science Strategy 2023 to 2033

Securing health and prosperity

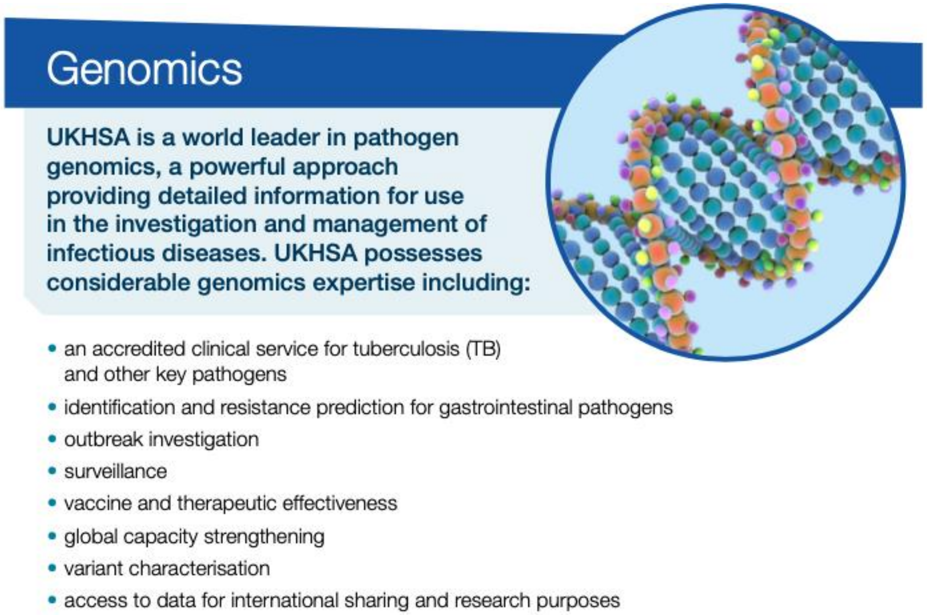


Data

Data collection and analysis are important steps in the scientific process, providing evidence to test and validate hypotheses. UKHSA collects and generates valuable data as part of its health protection activities. Our data is used internationally because of its quality.

Capabilities encompass:

- access to data through secure and interoperable systems, enabling sharing with international, national, local, and academic partners
- advanced modelling capabilities
- world-class insights gained through analytics and data science
- our science-driven, data-enabled approach underpins all our work



Genomics

UKHSA is a world leader in pathogen genomics, a powerful approach providing detailed information for use in the investigation and management of infectious diseases. UKHSA possesses considerable genomics expertise including:

- an accredited clinical service for tuberculosis (TB) and other key pathogens
- identification and resistance prediction for gastrointestinal pathogens
- outbreak investigation
- surveillance
- vaccine and therapeutic effectiveness
- global capacity strengthening
- variant characterisation
- access to data for international sharing and research purposes

NIHR | Health Protection Research Unit in Genomics and Enabling Data at University of Warwick

https://warwick.ac.uk/hpruged/

- NIHR funding £4m over 2020-2025
- Collaboration between UKHSA and three academic institutions
- Mission: to provide the methodological backbone required to improve national public health using genomic and epidemiological data
- 4 research themes, 12 projects
- Academic Career Development
- Knowledge Mobilisation
- Patient and Public Involvement and Engagement

UK Health Security Agency

THE UNIVERSITY OF WARWICK

Centre for Genomic Pathogen Surveillance

UNIVERSITY OF CAMBRIDGE

Structure

	Theme 1 Outbreak analysis	**Theme 2** Integrating genomics into epidemiology	**Theme 3** Interface and implementation	**Theme 4** Evolutionary dynamics
Theme leads	Didelot (Warwick)	Keeling (Warwick)	Aanensen (CGPS)	Parkhill (Cambridge)
UKHSA leads	Myers	Ribeca	Misra	Chattaway
Projects	1 Outbreak detection 2 Transmission analysis 3 Informing outbreak control	4 Hypothesis testing using genomic data 5 Population structure analysis 6 Population size dynamics	7 Interface tools 8 Databases for genomic and enabling data 9 Collection of enabling data	10 Evolution in genomic data 11 Evolution in metagenomic data 12 Prediction of evolutionary trends

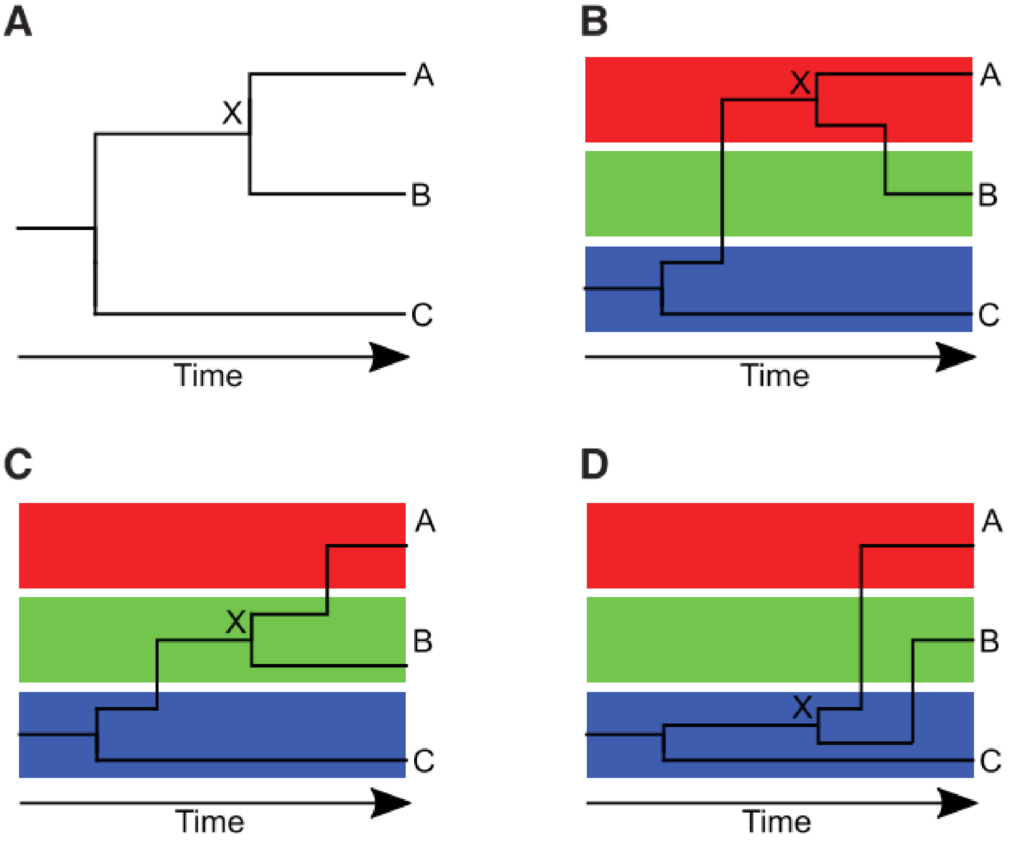
Challenges identified by PHE

- **Develop improved methods and algorithms for rapid and effective detection of genetic variants, and analysis of the impact of microbial genome sequencing on outbreak detection, investigation and monitoring, in particular, tracking real-time progression of the outbreak.**
- What synthetic evidence methods and machine learning take account of the complex transmission dynamics of infectious diseases, using data from surveillance systems, population-based surveys, genetic sequencing, epidemiological data from contact tracing and behavioural data?
- Develop tools and methods for rapid collation, visualisation and analysis of microbial genomic data for robust characterisation and prediction of resistance, and linking data to reservoirs and sources of transmission to identify threats.
- How can more granular data be achieved to enhance understanding of infections and drug resistance, devising policies and monitoring outcomes without significantly increasing data collection burdens?

Theme 1: outbreak analysis

Xavier Didelot (Warwick) and Richard Myers (UKHSA)

- Outbreak detection
- Transmission analysis
- Informing outbreak control



Challenges identified by PHE

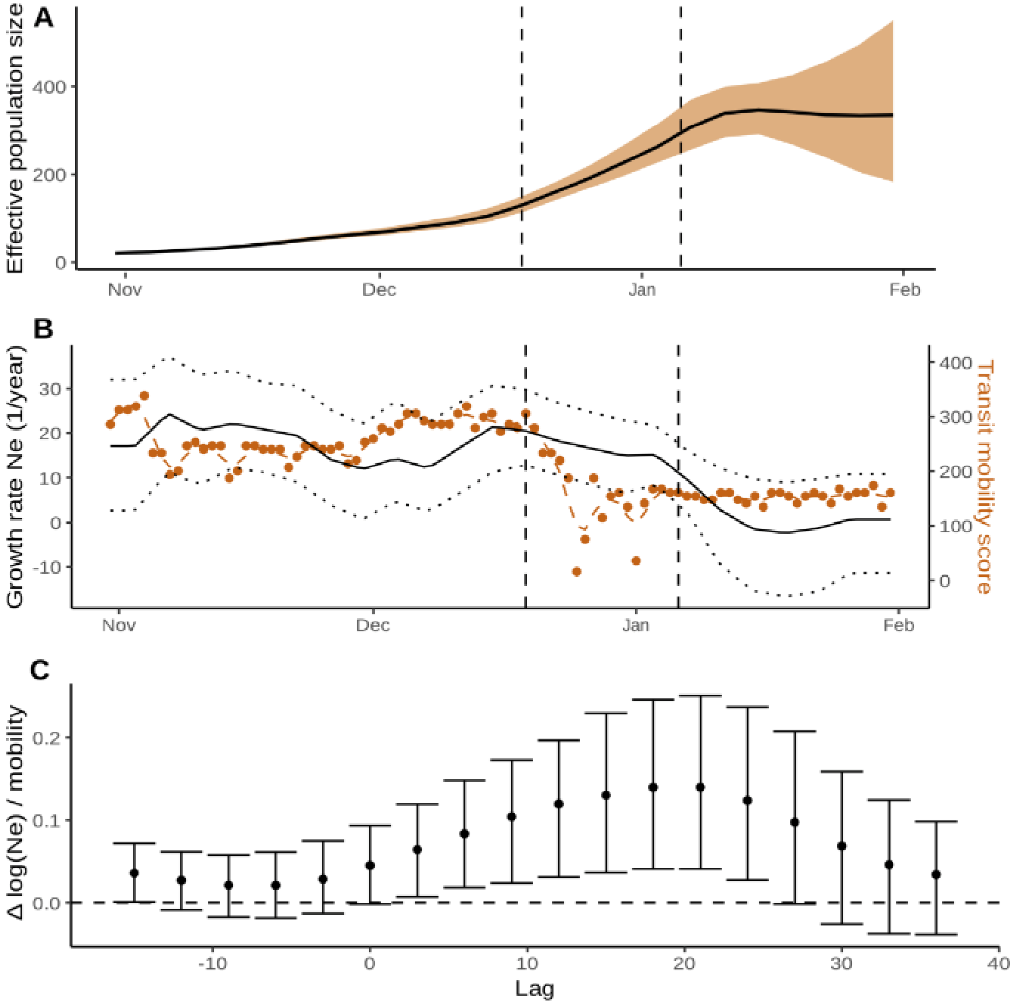
- Develop improved methods and algorithms for rapid and effective detection of genetic variants, and analysis of the impact of microbial genome sequencing on outbreak detection, investigation and monitoring, in particular, tracking real-time progression of the outbreak.
- **What synthetic evidence methods and machine learning take account of the complex transmission dynamics of infectious diseases, using data from surveillance systems, population-based surveys, genetic sequencing, epidemiological data from contact tracing and behavioural data?**
- Develop tools and methods for rapid collation, visualisation and analysis of microbial genomic data for robust characterisation and prediction of resistance, and linking data to reservoirs and sources of transmission to identify threats.
- How can more granular data be achieved to enhance understanding of infections and drug resistance, devising policies and monitoring outcomes without significantly increasing data collection burdens?

NIHR | Health Protection Research Unit in Genomics and Enabling Data at University of Warwick

Theme 2: Integrate genomics into epidemiology

Matt Keeling (Warwick) and Paolo Ribeca (UKHSA)

- Hypothesis testing using genomics
- Population structure analysis
- Population size dynamics



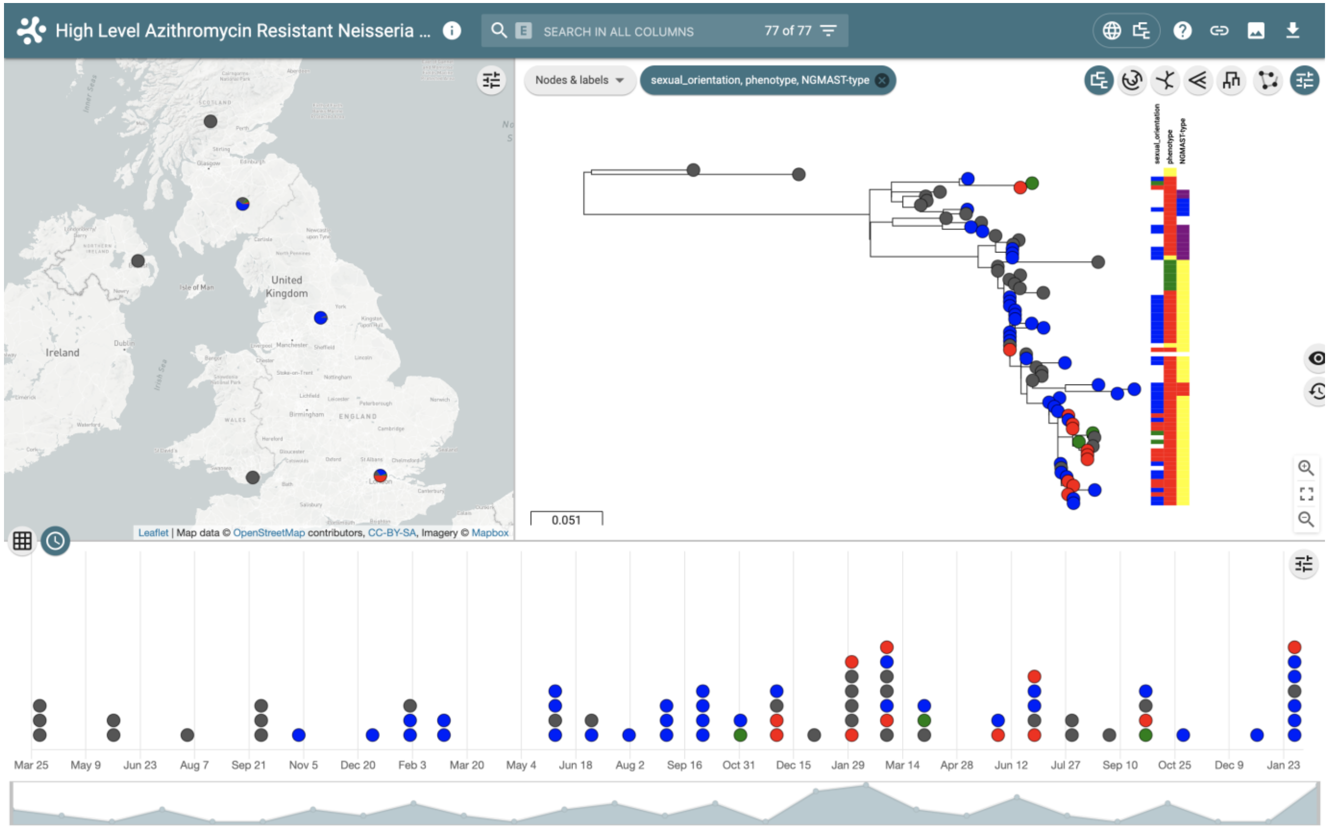
Challenges identified by PHE

- Develop improved methods and algorithms for rapid and effective detection of genetic variants, and analysis of the impact of microbial genome sequencing on outbreak detection, investigation and monitoring, in particular, tracking real-time progression of the outbreak.
- What synthetic evidence methods and machine learning take account of the complex transmission dynamics of infectious diseases, using data from surveillance systems, population-based surveys, genetic sequencing, epidemiological data from contact tracing and behavioural data?
- **Develop tools and methods for rapid collation, visualisation and analysis of microbial genomic data for robust characterisation and prediction of resistance, and linking data to reservoirs and sources of transmission to identify threats.**
- How can more granular data be achieved to enhance understanding of infections and drug resistance, devising policies and monitoring outcomes without significantly increasing data collection burdens?

Theme 3: Interface and implementation

David Aanensen (CGPS) and Raju Misra (UKHSA)

- Interface tools
- Databases for genomic and enabling data
- Collection of enabling data



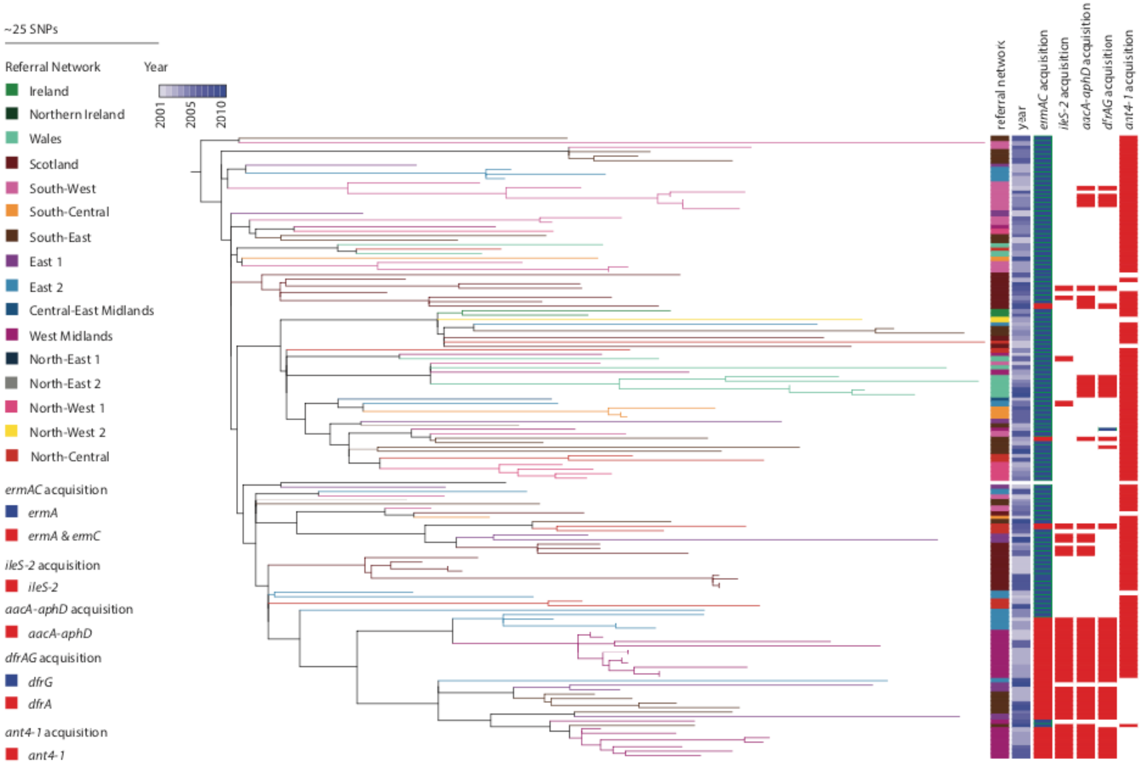
Challenges identified by PHE

- Develop improved methods and algorithms for rapid and effective detection of genetic variants, and analysis of the impact of microbial genome sequencing on outbreak detection, investigation and monitoring, in particular, tracking real-time progression of the outbreak.
- What synthetic evidence methods and machine learning take account of the complex transmission dynamics of infectious diseases, using data from surveillance systems, population-based surveys, genetic sequencing, epidemiological data from contact tracing and behavioural data?
- Develop tools and methods for rapid collation, visualisation and analysis of microbial genomic data for robust characterisation and prediction of resistance, and linking data to reservoirs and sources of transmission to identify threats.
- **How can more granular data be achieved to enhance understanding of infections and drug resistance, devising policies and monitoring outcomes without significantly increasing data collection burdens?**

Theme 4: evolutionary dynamics

Julian Parkhill (Cambridge) and Marie Chattaway (UKHSA)

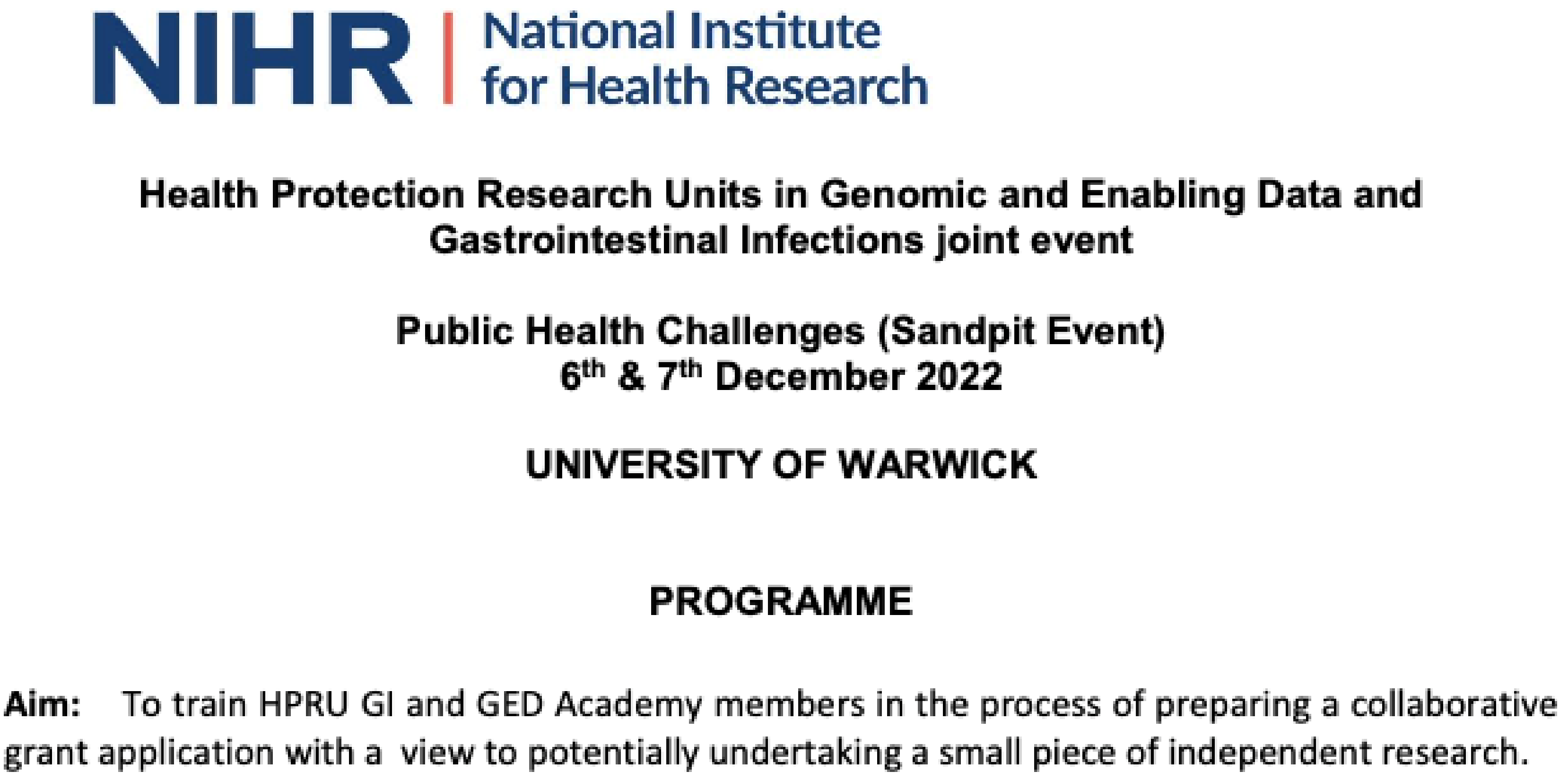
- Evolution in genomic data
- Evolution in metagenomic data
- Prediction of evolutionary trends



Collaboration with other HPRUs

- Methodology with broad relevance
- Links with disease-specific areas to retain applicability
- Open source software releases
- Collaborative work including:
  - HPRU in GI (Liverpool/Warwick)
  - HPRU in BBVSTI (UCL)
  - HPRU in MHE (Imperial)
  - HPRU in HCAI & AMR (Imperial, Oxford)
  - HPRU in RI (Imperial)

Training example: joint sandpit event



NIHR | National Institute for Health Research

Health Protection Research Units in Genomic and Enabling Data and Gastrointestinal Infections joint event

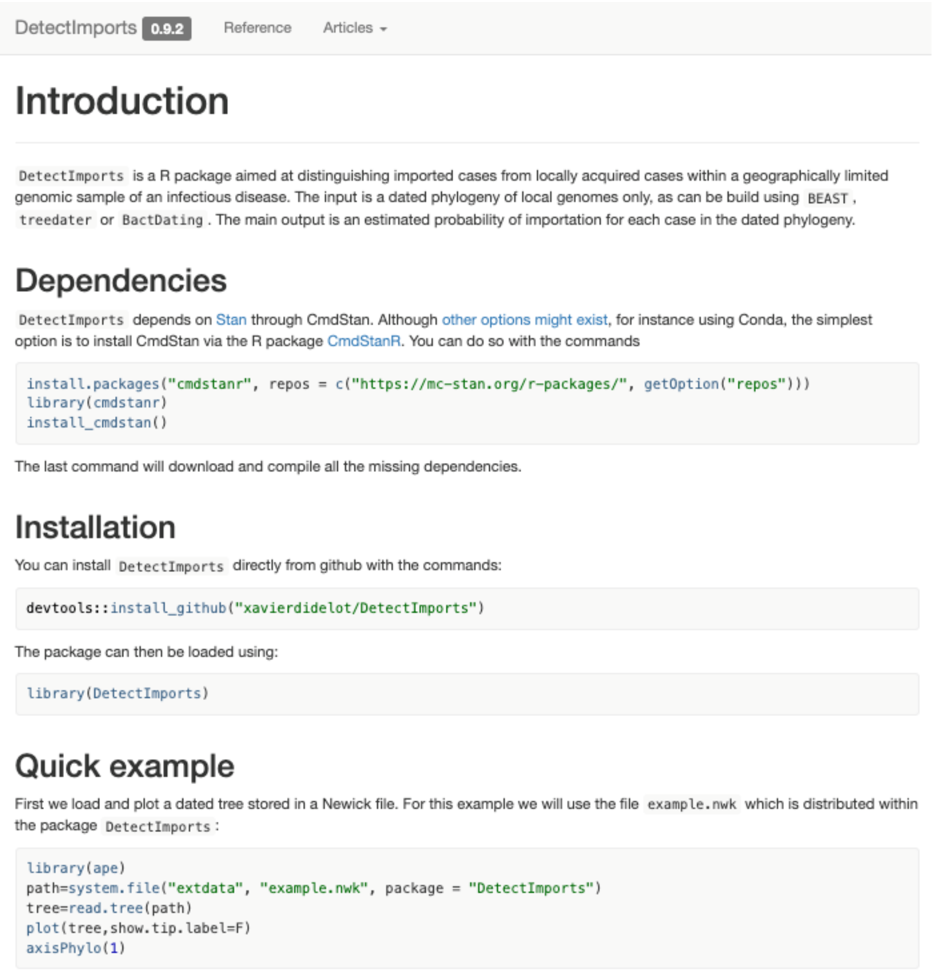
Public Health Challenges (Sandpit Event)
6th & 7th December 2022

UNIVERSITY OF WARWICK

PROGRAMME

Aim: To train HPRU GI and GED Academy members in the process of preparing a collaborative grant application with a view to potentially undertaking a small piece of independent research.

Knowledge mobilization examples



Introduction

`DetectImports` is a R package aimed at distinguishing imported cases from locally acquired cases within a geographically limited genomic sample of an infectious disease. The input is a dated phylogeny of local genomes only, as can be build using `BEAST`, `treedater` or `BactDating`. The main output is an estimated probability of importation for each case in the dated phylogeny.

Dependencies

`DetectImports` depends on Stan through CmdStan. Although other options might exist, for instance using Conda, the simplest option is to install CmdStan via the R package CmdStanR. You can do so with the commands

```
install.packages("cmdstanr", repos = c("https://mc-stan.org/r-packages/", getOption("repos")))
library(cmdstanr)
install_cmdstan()
```

The last command will download and compile all the missing dependencies.

Installation

You can install `DetectImports` directly from github with the commands:

```
devtools::install_github("xavierdidelot/DetectImports")
```

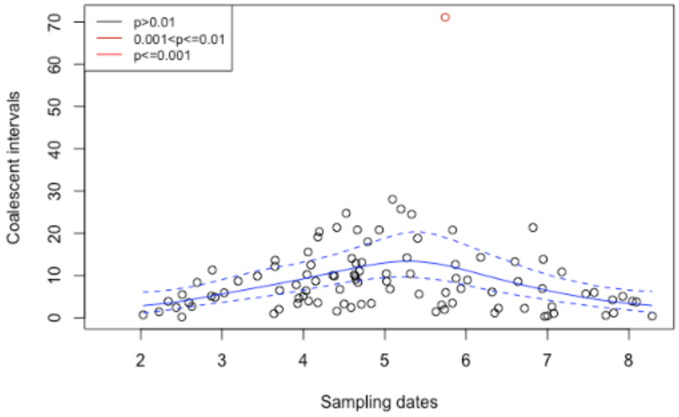
The package can then be loaded using:

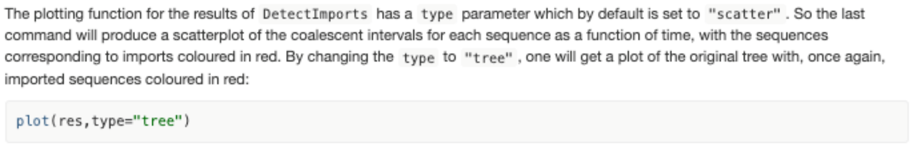
```
library(DetectImports)
```

Quick example

First we load and plot a dated tree stored in a Newick file. For this example we will use the file `example.nwk` which is distributed within the package `DetectImports`:

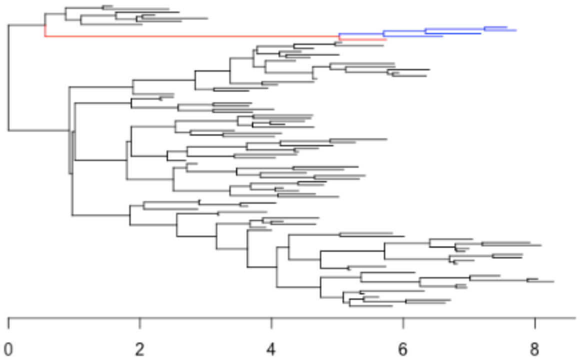
```
library(ape)
path=system.file("extdata", "example.nwk", package = "DetectImports")
tree=read.tree(path)
plot(tree,show.tip.label=F)
axisPhylo(1)
```





The plotting function for the results of `DetectImports` has a `type` parameter which by default is set to `"scatter"`. So the last command will produce a scatterplot of the coalescent intervals for each sequence as a function of time, with the sequences corresponding to imports coloured in red. By changing the `type` to `"tree"`, one will get a plot of the original tree with, once again, imported sequences coloured in red:

```
plot(res,type="tree")
```



NIHR | Health Protection Research Unit
in Genomics and Enabling Data
at University of Warwick

Programme of the Annual Scientific Conference
Monday 22nd and Tuesday 23rd April 2024
UKHSA, COLINDALE

DAY 1 ***Monday 22nd April 2024***

12:00 pm *Registration and Lunch*

1:00 pm **Welcome – Marie Chattaway & Paolo Ribeca, UK Health Security Agency**

SESSION 1, Chaired by Xavier Didelot, University of Warwick

1:10 pm Microbiome response to antibiotics across multiple scales
Chris Quince (Earlham)

1:30 pm Diversity and distribution of mobile genetic elements in foodborne pathogens
Clare Barker (UKHSA)

1:50 pm A phylogenetic approach for estimating rates of gene gain/loss and selection
Caitlin Collins (Cambridge)

2:10 pm Classifying microbial species in contaminated disease-outbreak isolates using kmer spectra-based multidimensional clustering
Ryan Morrison (UKHSA)

2:30 pm Knowledge mobilization
Noel McCarthy (Trinity College Dublin)

2:50 pm From research to routine - building the public health genomics programme in Wales
Tom Connor (Cardiff)

3:10 pm **Coffee break/poster viewing**

SESSION 2, Chaired by Noel McCarthy, Trinity College Dublin

3:40 pm **KEYNOTE SPEAKER**
Genomic platforms for surveillance of enteric fever and AMR
Kat Holt (LSHTM)

4:10 pm LineageCapture: Phylogenetic identification of bacterial clade members excluded by SNP clustering
Matt Moore (Warwick)

4:30 pm Projections for COVID-19, a tale of two problems
Matt Keeling (Warwick)

4:50 pm Microevolution during the emergence of pandemic monophasic Salmonella Typhimurium ST34
Robert Kingsley (Quadram)

5:10 pm KPop: An assembly-free and scalable method for the comparative analysis of microbial genomes
Paolo Ribeca (UKHSA)

NIHR | Health Protection Research Unit
in Genomics and Enabling Data
at University of Warwick

DAY 2 ***Tuesday 23rd April 2024***

09:00 am **Registration, Coffee and Networking**

SESSION 3, Chaired by Marie Chattaway, UK Health Security Agency

10:00 am **KEYNOTE SPEAKER**
Applying innovative genomic approaches to the prevention and control of infectious diseases
Deborah Williamson (UKHSA)

10:30 am EnteroBase in 2024
Sascha Ott (Warwick)

10:50 am Fresh perspectives on bacterial variation: at genome scale and within patients
Gemma Langridge (Quadram)

11:10 am Detection and characterisation of Salmonella enterica serovar Infantis (eBG31) harbouring blaCTX-M-1 causing clinical disease in humans in England.
Matt Bird (UKHSA)

11:30 am Distinguishing imported cases from locally acquired cases within a geographically limited genomic sample of an infectious disease
Xavier Didelot (Warwick)

11:50 am An integrated approach for academic training and professional registration, addressing inequalities for healthcare scientists
Marie Chattaway (UKHSA)

12:00 pm **Group Photo, followed by Poster Viewing and Lunch**

SESSION 4, Chaired by Paolo Ribeca, UK Health Security Agency

1:00 pm Harnessing Genomics for a One Health Approach: Insights from Salmonella Pathogen Lead
Marie Chattaway (UKHSA)

1:20 pm Inference of infectious disease transmission through a relaxed bottleneck using multiple genomes per host
Jake Carson (Warwick)

1:40 pm Signatures of Pathogen Emergence
Julian Parkhill (Cambridge)

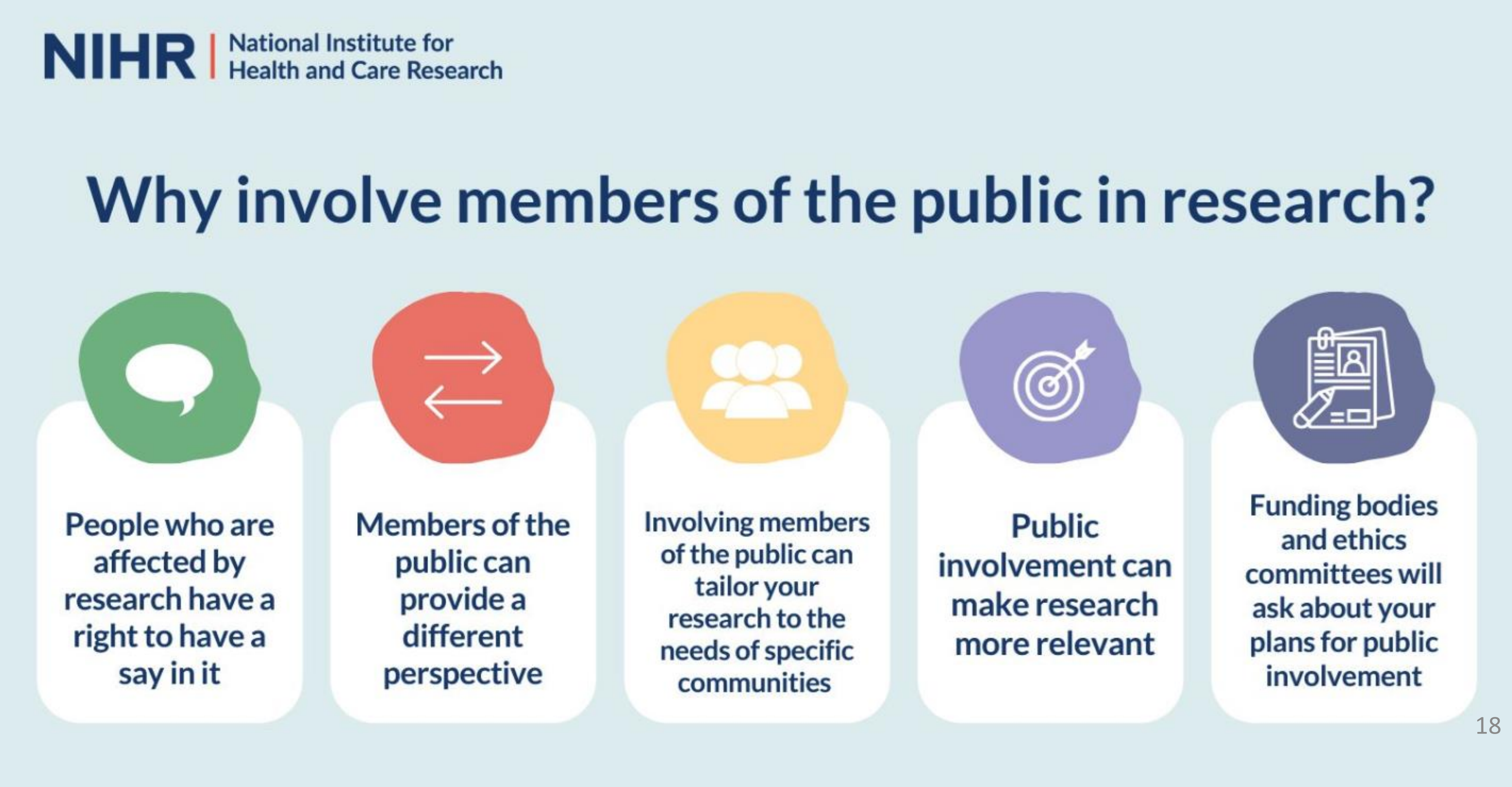
2:00 pm Cóimeáil: A Nanopore-based bioinformatics pipeline for the rapid typing and characterisation of gastrointestinal pathogens
David Greig (UKHSA)

2:20 pm Exploring Black Perspectives on the Impacts of Patient and Public Involvement and its Evaluation (The ELEVATE Study)
Jade Jordan (Warwick)

2:40 pm Deducing clonal complex population structure from gene content
Emily Fotopoulou (UKHSA)

3:00 pm **Closing Remarks - Xavier Didelot**

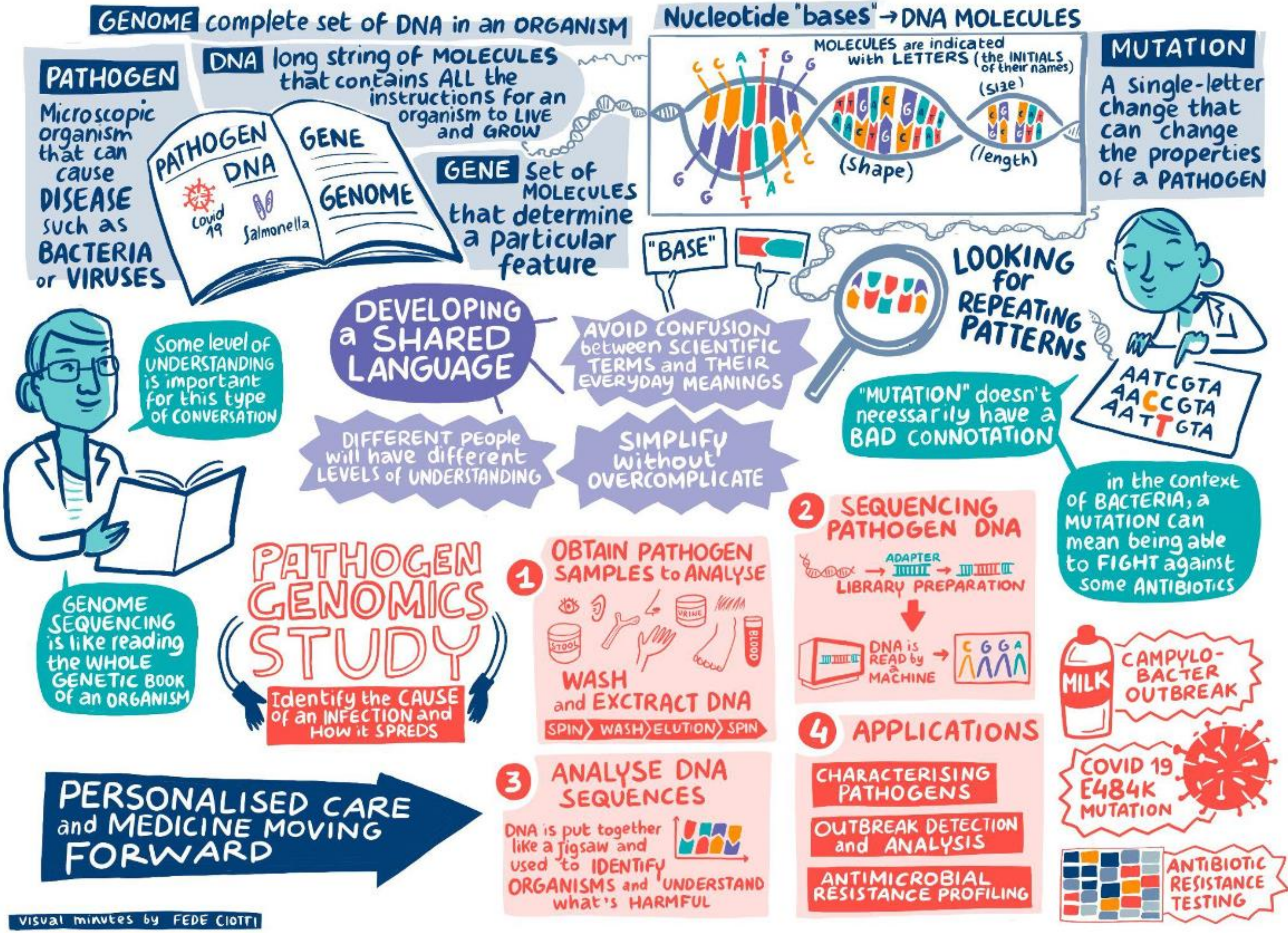
Patient and Public Involvement and Engagement



NIHR | National Institute for Health and Care Research

Why involve members of the public in research?

- People who are affected by research have a right to have a say in it
- Members of the public can provide a different perspective
- Involving members of the public can tailor your research to the needs of specific communities
- Public involvement can make research more relevant
- Funding bodies and ethics committees will ask about your plans for public involvement



GENOME complete set of DNA in an ORGANISM

PATHOGEN Microscopic organism that can cause DISEASE such as BACTERIA or VIRUSES

DNA long string of MOLECULES that contains ALL the instructions for an organism to LIVE and GROW

GENE Set of MOLECULES that determine a particular feature

Nucleotide "bases" → DNA MOLECULES

MOLECULES are indicated with LETTERS (the INITIALS of their names)

MUTATION A single-letter change that can change the properties of a PATHOGEN

DEVELOPING a SHARED LANGUAGE

AVOID CONFUSION between SCIENTIFIC TERMS and THEIR EVERYDAY MEANINGS

DIFFERENT People will have different LEVELS of UNDERSTANDING

SIMPLIFY without OVERCOMPLICATE

LOOKING for REPEATING PATTERNS

"MUTATION" doesn't necessarily have a BAD CONNOTATION

in the context of BACTERIA, a MUTATION can mean being able to FIGHT against some ANTIBIOTICS

Some level of UNDERSTANDING is important for this type of CONVERSATION

GENOME SEQUENCING is like reading the WHOLE GENETIC BOOK of an ORGANISM

PATHOGEN GENOMICS STUDY

Identify the CAUSE of an INFECTION and How it SPREDS

1 OBTAIN PATHOGEN SAMPLES to ANALYSE

WASH and EXCTRACT DNA

SPIN › WASH › ELUTION › SPIN

2 SEQUENCING PATHOGEN DNA

LIBRARY PREPARATION

DNA is READ by a MACHINE

3 ANALYSE DNA SEQUENCES

DNA is put together like a jigsaw and used to IDENTIFY ORGANISMS and UNDERSTAND what's HARMFUL

4 APPLICATIONS

- CHARACTERISING PATHOGENS
- OUTBREAK DETECTION and ANALYSIS
- ANTIMICROBIAL RESISTANCE PROFILING

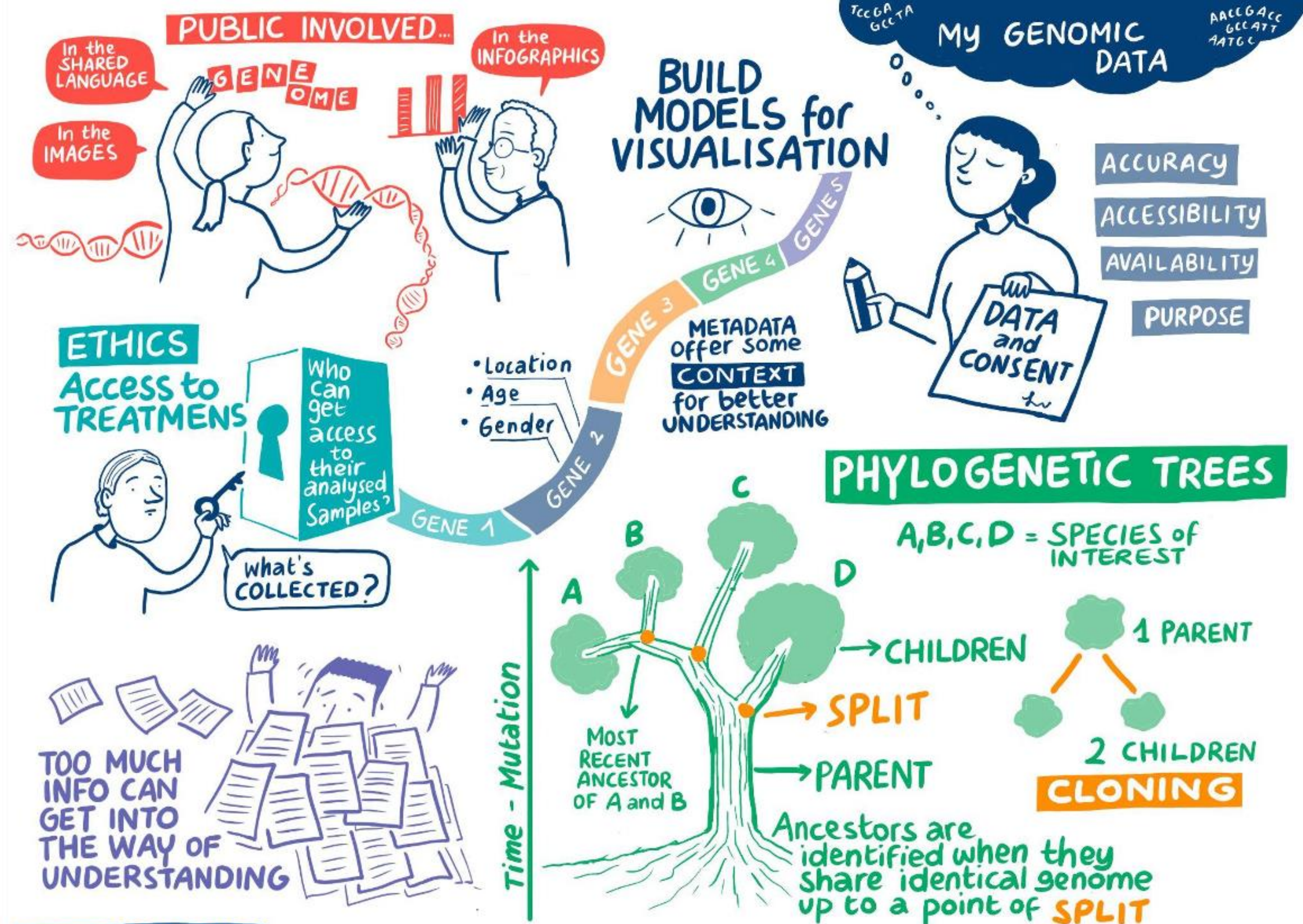
CAMPYLOBACTER OUTBREAK

COVID 19 E484K MUTATION

ANTIBIOTIC RESISTANCE TESTING

PERSONALISED CARE and MEDICINE MOVING FORWARD

visual minutes by FEDE CIOTTI



visual minutes by FEDE CIOTTI

INFOGRAPHICS as MEANS of COMMUNICATION

IT'S ABOUT the STORIES we want to tell...

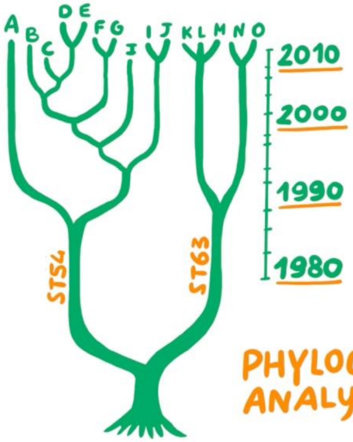
PATHOGEN SEQUENCING TO BE DEVELOPED WITH THE PUBLIC

BE AWARE OF DIVERSITY!

HOW TO MAKE ACCESSIBLE REPRESENTATIONS OF GENOMIC DATA

BETTER USE OF PICTORIAL ELEMENTS

Difficile TRANSMISSION WITHIN HOSPITALS



PHYLOGENETIC ANALYSIS

TREES TELL US THE STORIES OF WHAT WE ARE TRYING TO INVESTIGATE

CLEAR and CONSISTENT USE of COLOURS

RIGHT VALUES for RIGHT PARAMETERS

HOWEVER, LOOKING AT TREES IS NOT ENOUGH SOMETIMES

MAPS and LOCATION DATA

PUBLIC can be involved in TRANSMISSION TREES

↓

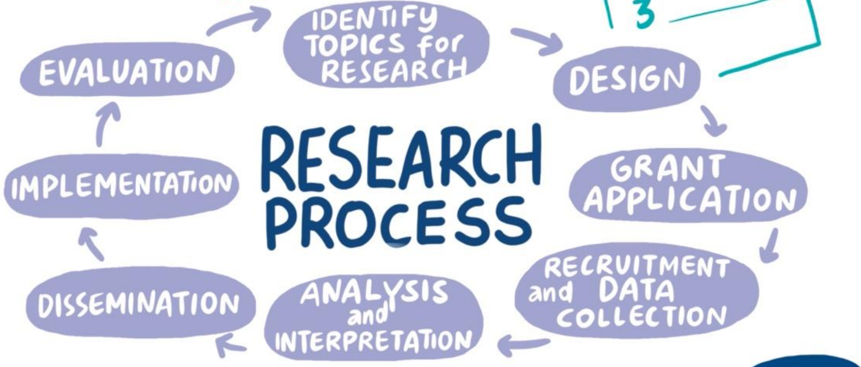
VALIDATE TREES' INTERPRETATION

PLOTS AND GRAPHS

PATIENTS, CARERS and the PUBLIC to be INVOLVED at EVERY STAGE

What should be made a PRIORITY?

NEW AREAS of RESEARCH


IDENTIFY TOPICS for RESEARCH
DESIGN
GRANT APPLICATION
RECRUITMENT and DATA COLLECTION
ANALYSIS and INTERPRETATION
DISSEMINATION
IMPLEMENTATION
EVALUATION
RESEARCH PROCESS

VALUE our TIME

LIVED EXPERIENCE is HARD to FIND

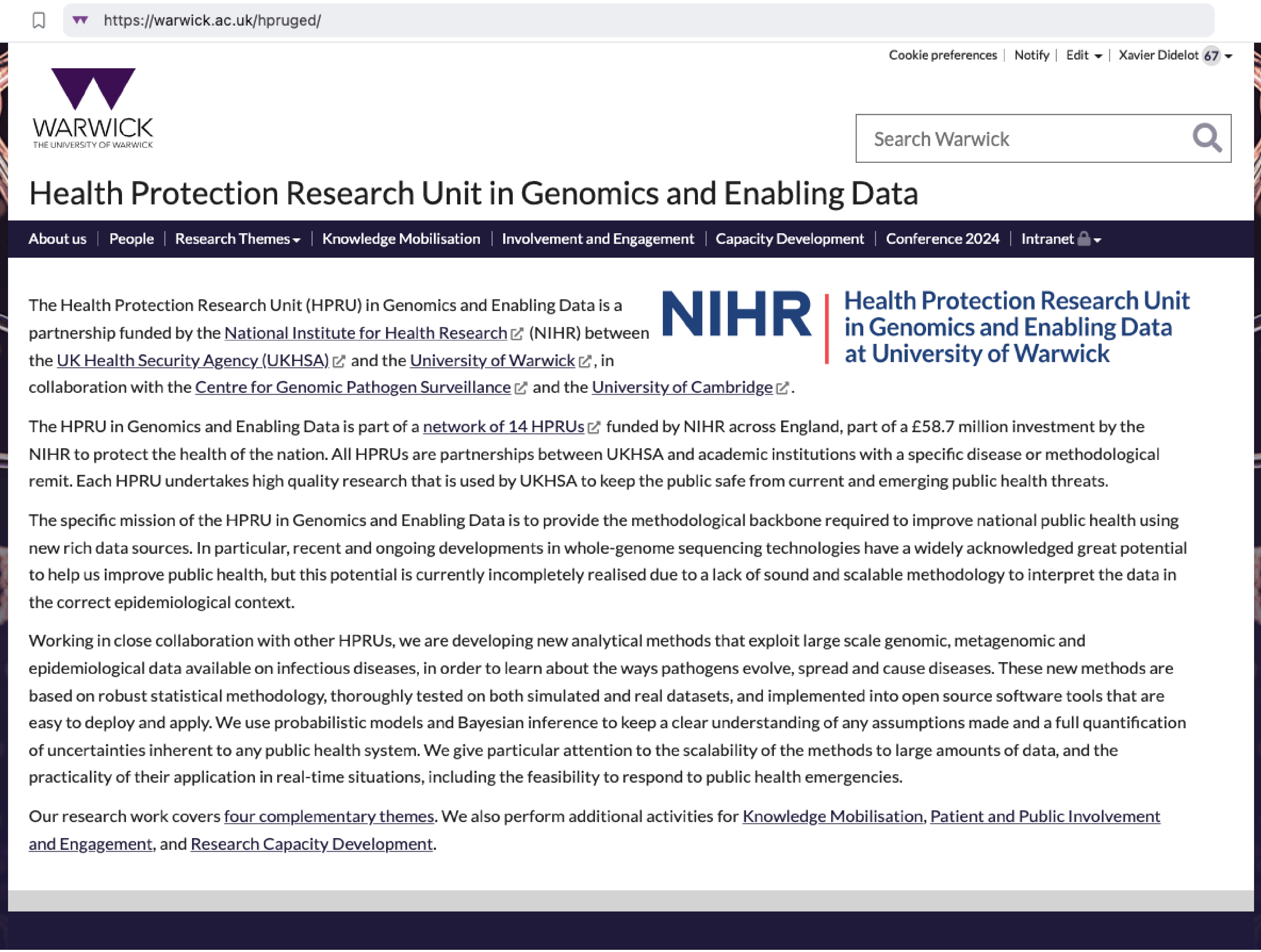
DIFFERENT PEOPLE with DIFFERENT SKILLS coming TOGETHER!

OUTCOMES: are we measuring the RIGHT THINGS in the RIGHT WAY?

Can PUBLIC have a say on CRITERIA around "INCLUSION" and "EXCLUSION"?

don't pursue only ONE NARRATIVE → KNOWLEDGE MOBILISATION is KEY!

visual minutes by FEDE CIOTTI


https://warwick.ac.uk/hpruged/
Cookie preferences | Notify | Edit | Xavier Didelot 67
WARWICK
THE UNIVERSITY OF WARWICK
Search Warwick
About us | People | Research Themes | Knowledge Mobilisation | Involvement and Engagement | Capacity Development | Conference 2024 | Intranet

Health Protection Research Unit in Genomics and Enabling Data

NIHR | Health Protection Research Unit in Genomics and Enabling Data at University of Warwick

The Health Protection Research Unit (HPRU) in Genomics and Enabling Data is a partnership funded by the National Institute for Health Research (NIHR) between the UK Health Security Agency (UKHSA) and the University of Warwick, in collaboration with the Centre for Genomic Pathogen Surveillance and the University of Cambridge.

The HPRU in Genomics and Enabling Data is part of a network of 14 HPRUs funded by NIHR across England, part of a £58.7 million investment by the NIHR to protect the health of the nation. All HPRUs are partnerships between UKHSA and academic institutions with a specific disease or methodological remit. Each HPRU undertakes high quality research that is used by UKHSA to keep the public safe from current and emerging public health threats.

The specific mission of the HPRU in Genomics and Enabling Data is to provide the methodological backbone required to improve national public health using new rich data sources. In particular, recent and ongoing developments in whole-genome sequencing technologies have a widely acknowledged great potential to help us improve public health, but this potential is currently incompletely realised due to a lack of sound and scalable methodology to interpret the data in the correct epidemiological context.

Working in close collaboration with other HPRUs, we are developing new analytical methods that exploit large scale genomic, metagenomic and epidemiological data available on infectious diseases, in order to learn about the ways pathogens evolve, spread and cause diseases. These new methods are based on robust statistical methodology, thoroughly tested on both simulated and real datasets, and implemented into open source software tools that are easy to deploy and apply. We use probabilistic models and Bayesian inference to keep a clear understanding of any assumptions made and a full quantification of uncertainties inherent to any public health system. We give particular attention to the scalability of the methods to large amounts of data, and the practicality of their application in real-time situations, including the feasibility to respond to public health emergencies.

Our research work covers four complementary themes. We also perform additional activities for Knowledge Mobilisation, Patient and Public Involvement and Engagement, and Research Capacity Development.