# Tailoring bespoke medicine using big genetics and genomics data in the NHS

Joanna D Holbrook
Karen Temple,
Christopher Woelk,
James Batchelor

# NIHR Biomedical Research Centre: Building a Learning Health System

**Aims:**

1. Integrate high-dimensional e-health data such as imaging, electrograms, activity trackers and molecular profiles, so as to implement the advances of stratified medicine.

2. Use patient-relevant data from across the hospital and university to improve clinical care by data-driven decision making

3. Refine phenomarker profiles by iterative interrogation of patient data and enable replication and validation studies

# To do this, we must

1. Get all the data together in a secure environment (BRC Datawarehouse) with structure, standardisation and normalisation

2. Interrogate it in the context of known and putative phenomarkers for stratified medicine.

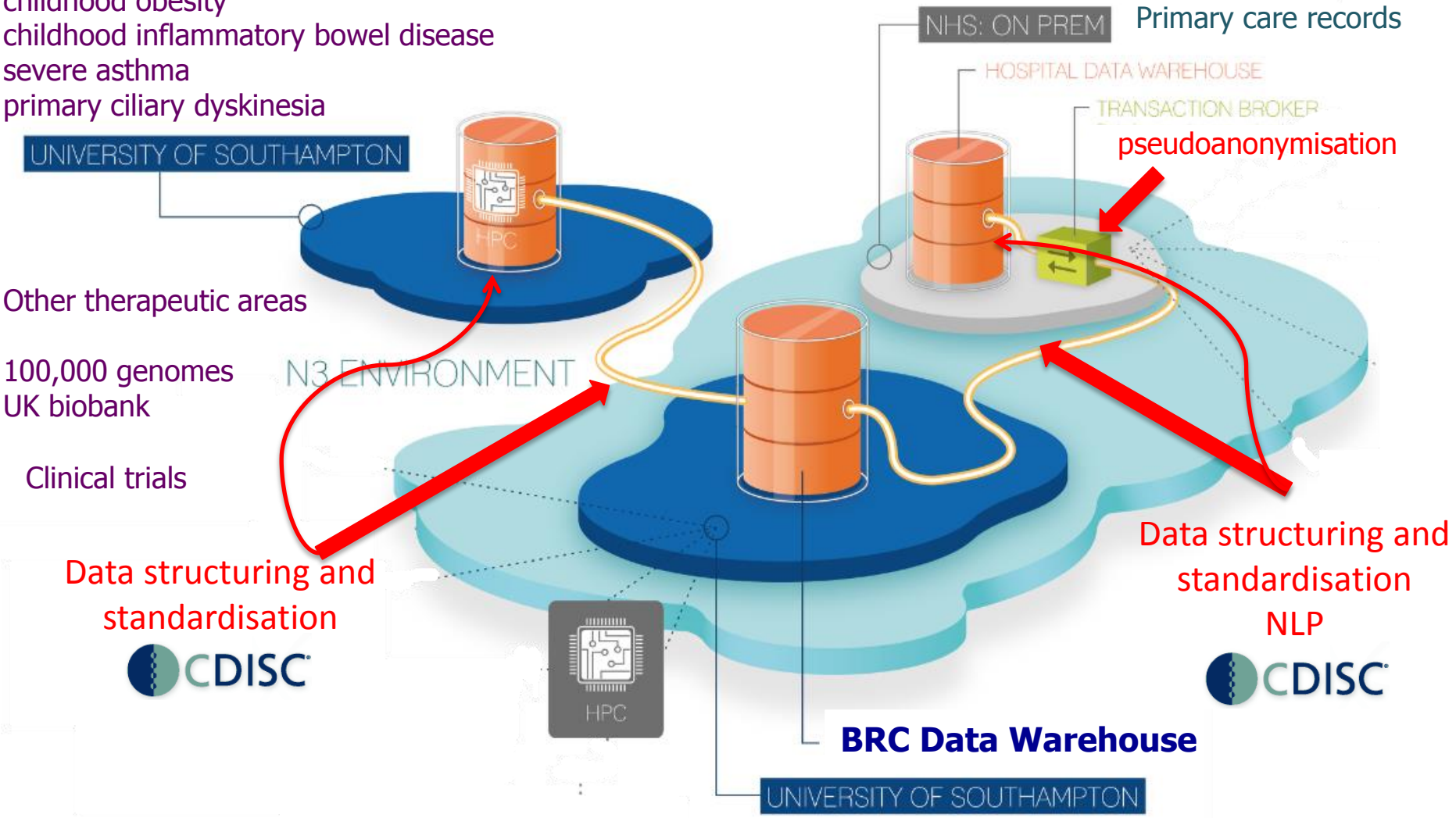
3. Feed the results of these analyses back into the data and clinical care.

# Where the data comes from

**Research cohort databases**
childhood obesity
childhood inflammatory bowel disease
severe asthma
primary ciliary dyskinesia

**SGH:** patient identifiable data,
electronic health records (HER)

Primary care records




Other therapeutic areas

100,000 genomes
UK biobank

Clinical trials

Data structuring and standardisation

CDISC

Data structuring and standardisation
NLP

CDISC

**BRC Data Warehouse**

# Hurdles to integrated data

- **Data safety and security**
  - N3 compliant but need HPC facility too.
  - Data warehouse will be pseudoanonymised but there will be the ability to map back to patient identifiers.

- **Some is unstructured and all must be standardised**
  - Apply Naturel language processing and standardise to CDISC
  - Feedback structure and standards to individual databases
  - Data cleaning may be necessary, have to evaluate and apply *ad hoc*

- **Patient consent for use of clinical data in research and new contacts with subject for further investigations**
  - Comprehensive ethical framework with pro-active consenting of SGH patients.

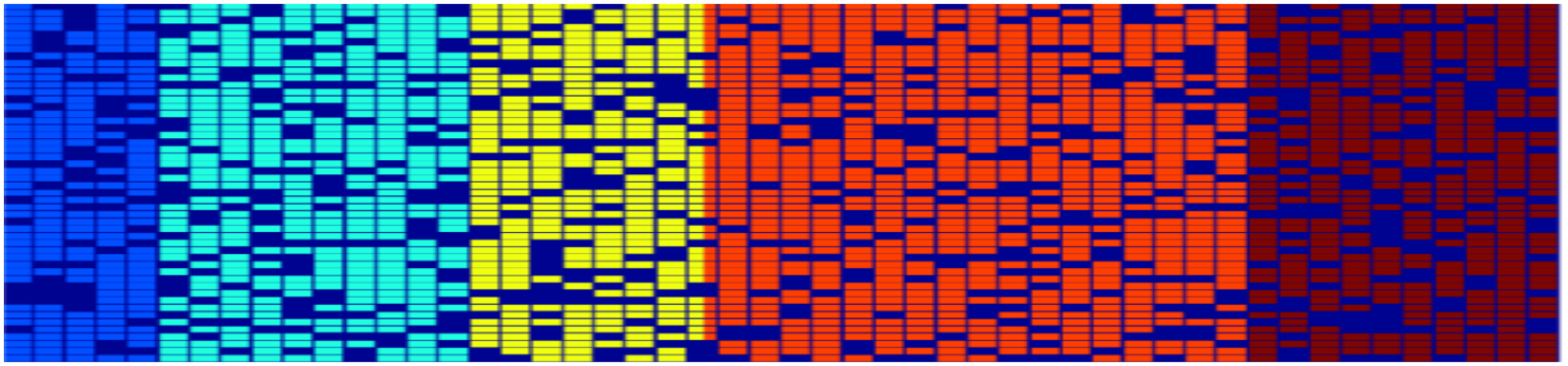
# What the data will look like

Virtual data matrix: rows are patients, columns are variables

**Pseudoanonymised IDs demographics**

**Biochemical, immunological and electogram measures**

**(Processed) Imaging data**
DTI/ MRI/DEXA
(100s datapoints/ timepoint)



**Diagnoses, clinical decisions and outcomes**

**(Processed) Genetics and genomics data**
Germline variants (> 10 million)
DNA methylation in various tissues ( ~ 4 million)
Transcriptomes in various tissue (> 20K)
Microbiota from various sample sites (1000s rising to millions)

Nextgen: Activity trackers and other mobile connected apps

- Short and wide -> multiple testing problems, over-fitting problems
- Some data types will have high degree of missingness and/or associated confidence metrics
- Complex structure of longitudinal and/or hierachical relations e.g. DNA->RNA

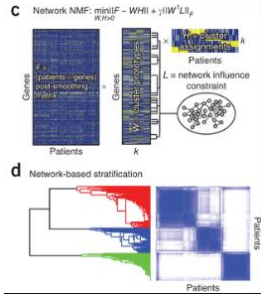
# How do we want to interrogate it?

**Replicate**

Many phenomarker signatures have already been reported with varying degrees of confidence in translation to clinical decision making. We can make these readily available to clinicians and use more data to improve on them.

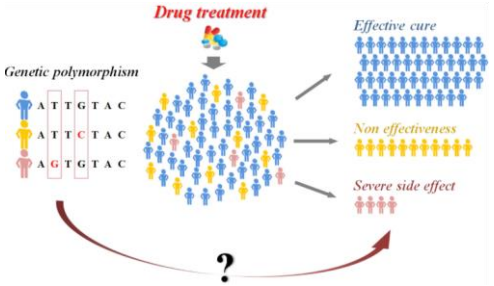
**Discover new patterns**

Define diseases, disease subclasses and prognosis by supplementing symptom data with high-dimensional data.



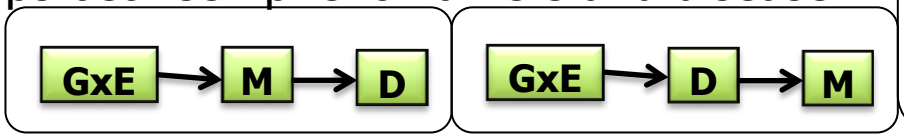

**Make new predictions**

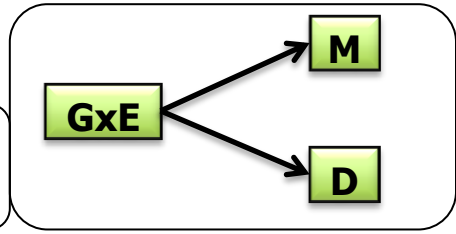
Predict response to treatments given full dataset for patient (includes pharmacogenomics)




**Learn causal relationships**

Longitudinal data analysis and causal inference (MR), to unpick casual relationships between phenomarkers and disease



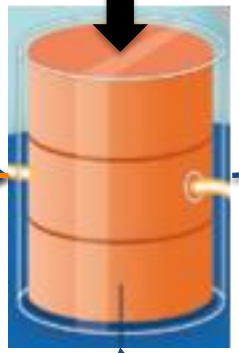
# Where we are going: What a true learning health system looks like

samples

HD data

symptoms

Known and putative phenomarker associations



Predictions of diagnosis and treatment response

treatments

outcomes

✓ Deliver personalised medicine and refine it within system

Refinement of phenomarker signatures
New phenomarker signatures
Drug repositioning

✓ Data driven discovery applicable more widely

Replication in other health systems (Other Trusts and BRCs)

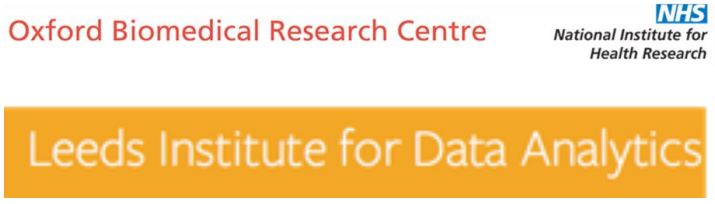
Randomised control trials
Select patients with particular profiles from data warehouse

# We are not alone












# Acknowledgements


- Karen Temple
- James Batchelor
- Christopher Woelk

- Sarah Ennis, Jacek Brodzki, Borislav Dimitrov, Ben MacArthur, Simon Cox, Jermey Wyatt, Peter Smith, Colin Newell, David Cable, Nick Sheron, Anneke Lucassen, Sandy MacKinnon, Tony Postle, Chris Mattocks, Anthony Williams, Spiros Garbis, Paul Skipp, Michael Bennett, Andrew Collins

- NIHR Biomedical Research Centre: esp. Rob Reed, Keith Godfrey, Ratko Djukanovic