MANCHESTER 1824
The University of Manchester

CfMR
Centre for Musculoskeletal Research

Integration of Research IT services within the Centre for Musculoskeletal Research

John Bowes
Research Fellow

Arthritis Research UK | centre for genetics & genomics

NHS
National Institute for Health Research

MANCHESTER 1824
The University of Manchester

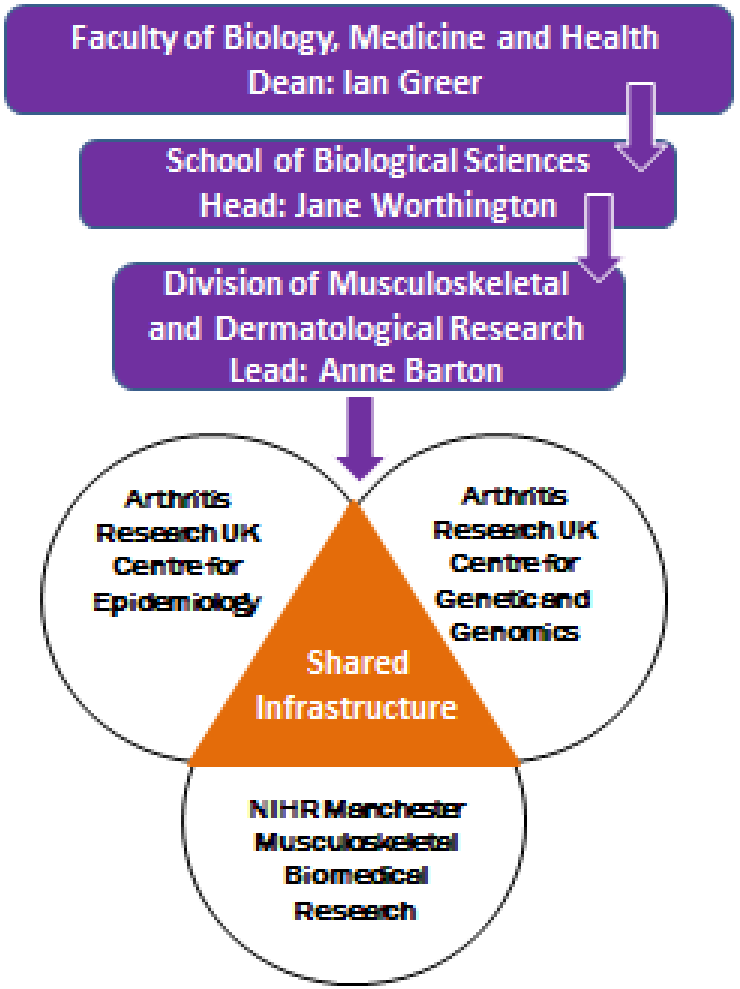
Aim...

1. .. to show how as a centre we have adopted core set of Research IT services as our routine analysis platform

2. Two examples where collaborating with Research IT has enabled to develop projects not previously possible

MANCHESTER 1824
The University of Manchester

Centre for Musculoskeletal Research



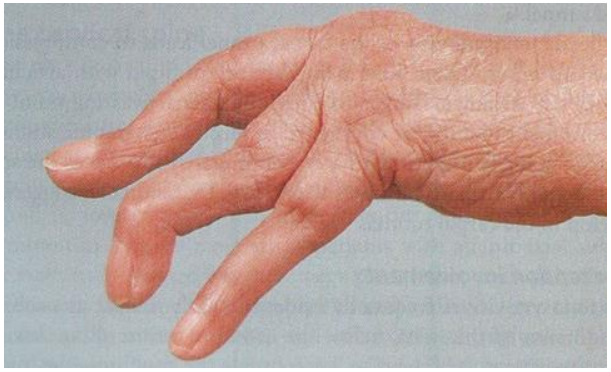
Centre for Genetics and Genomics

Study genetic basis of musculoskeletal disorders:

1. Rheumatoid Arthritis
2. Juvenile Idiopathic Arthritis
3. Psoriatic arthritis

Autoimmune diseases:

- significant disability
- increased morbidity
- increased mortality



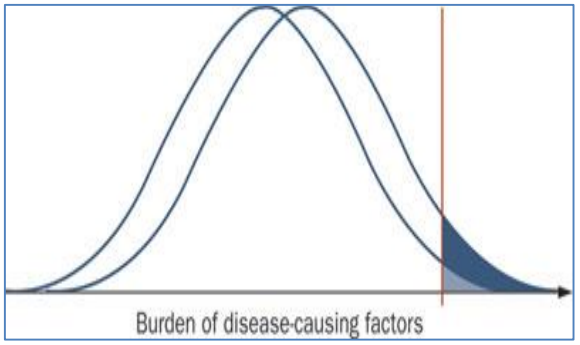
MANCHESTER 1824
The University of Manchester

Common complex diseases

- Complex disease

  1. Genetic risk factors
  2. Environment risk factors
  3. G+E = liability

- Goal: identify genetic variants that underlie disease

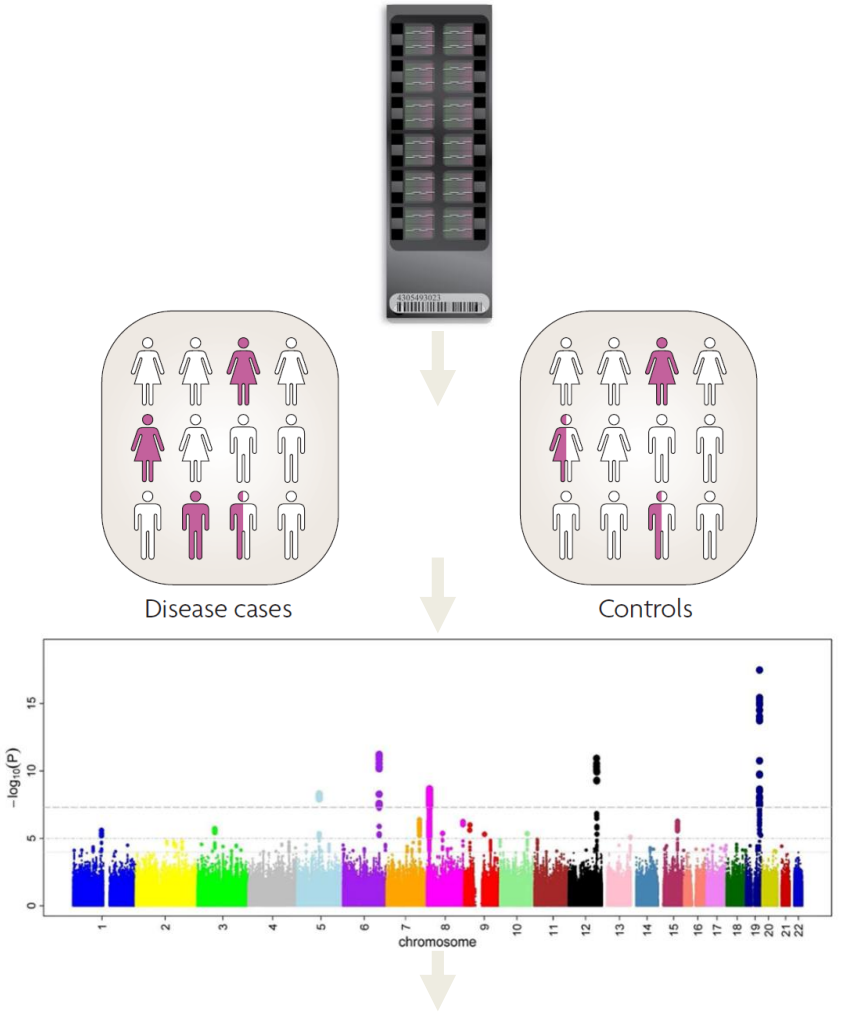


Genome-wide association studies (GWAS)

Genotyping
~1 million genetic variants randomly distributed in the genome typed in high-density arrays

Case-control study
Compare frequencies of genetic variants between disease cases and controls

Genome scan results
Significant differences in SNP allele frequencies indicate possible new disease genes and loci



Genotype-associated SNPs in independent case-control sample

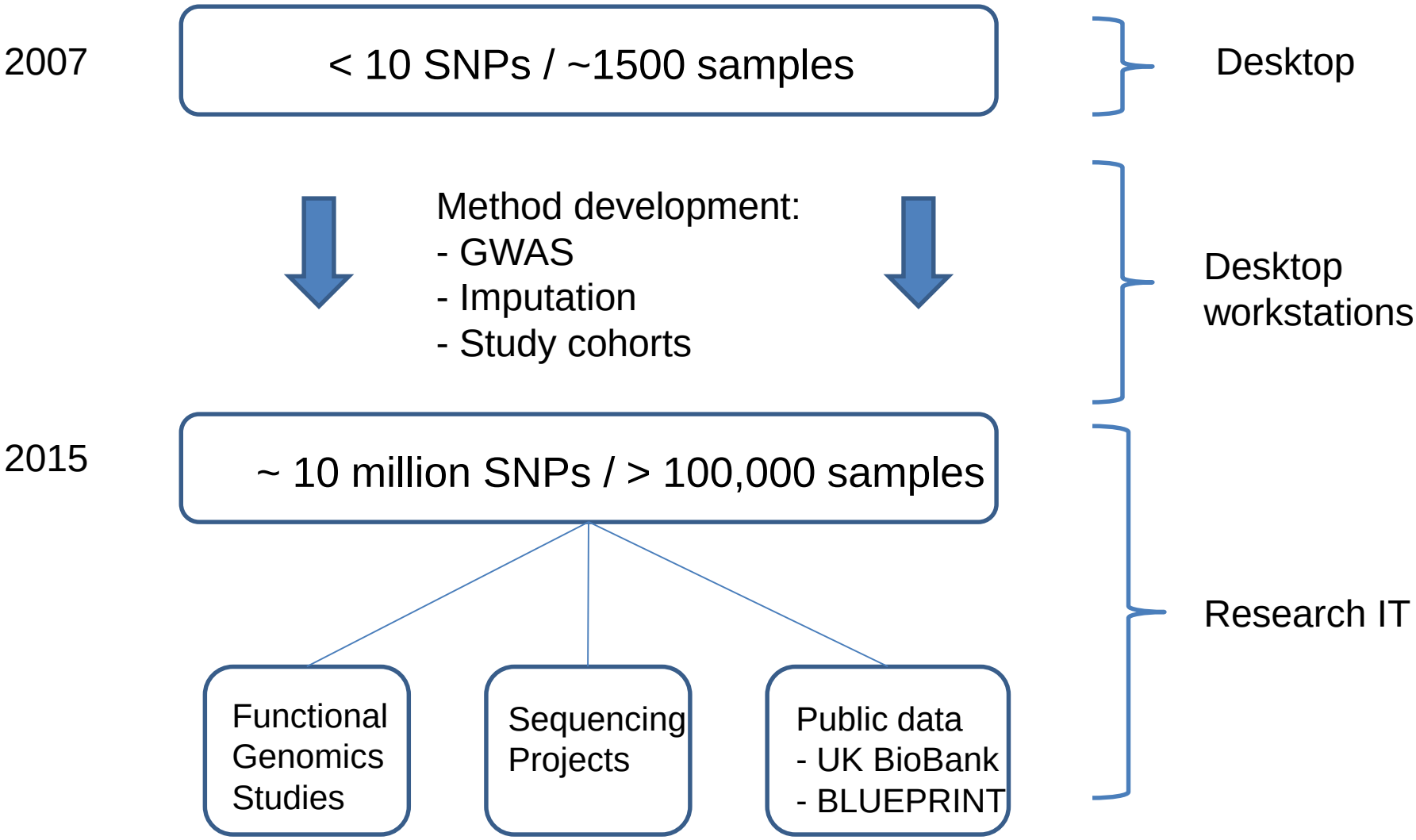
MANCHESTER 1824
The University of Manchester

Research interests:

1. **Understanding susceptibility**
   - risk prediction
2. **Disease outcome**
   - Longitudinal – disease severity
3. **Treatment**
   - drug response/repositioning
4. **Biology – functional genomics**
   - Chromatin confirmation, RNA-seq, CRISPR

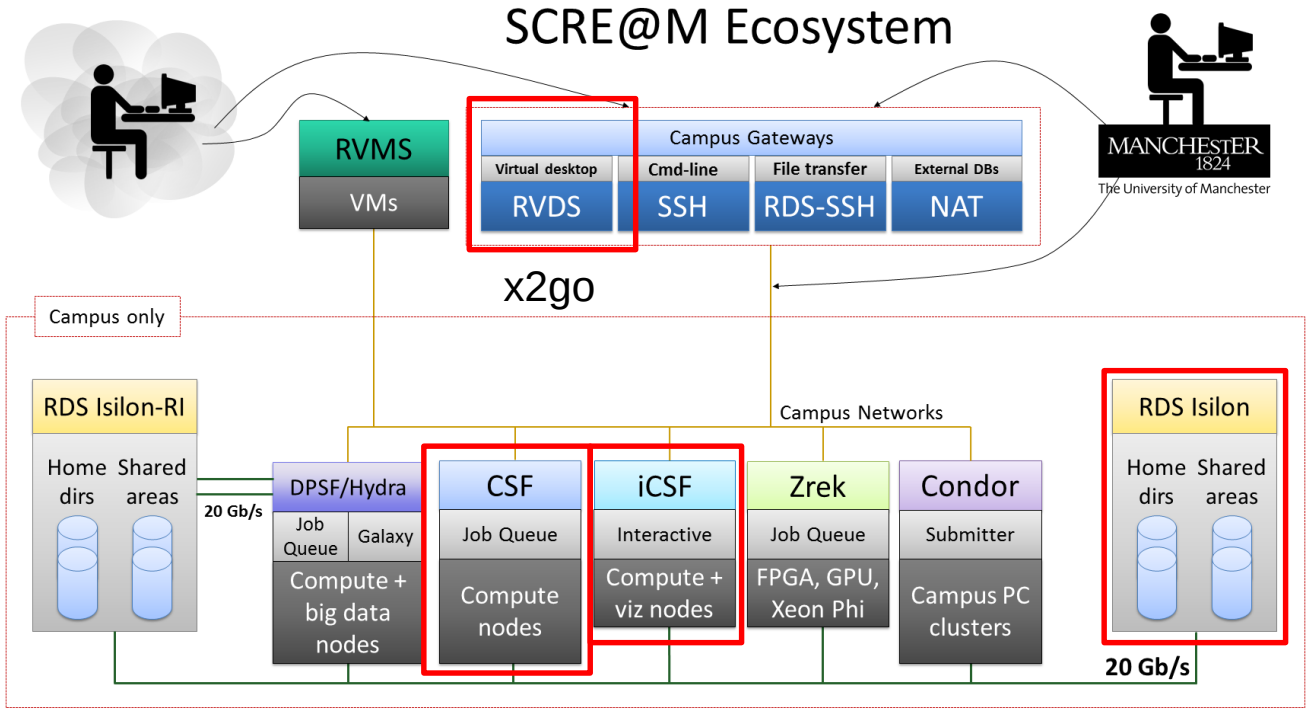
MANCHESTER 1824
The University of Manchester

Acceleration of data collection



MANCHESTER 1824
The University of Manchester

Computationally Intensive Resource (CIR) ecosystem

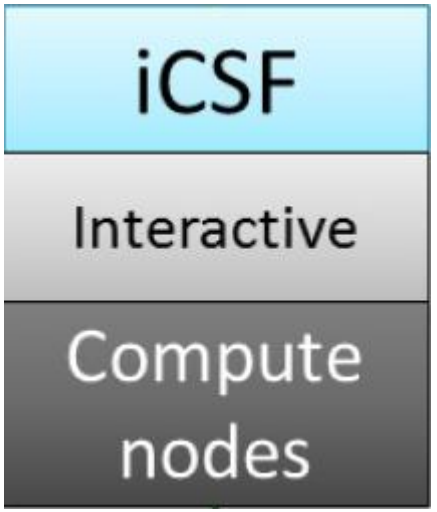


Rely on 4 core components:

1. RVDS
2. iCSF
3. CSF
4. Isilon

MANCHESTER 1824
The University of Manchester

iCSF overview



- Main analysis platform
- GUI-based Interactive computational work
  - Stata, rstudio, MATLAB
- No SGE
  - Log directly onto back-end node (12)
- Multi-user workstation (>60 users)

MANCHESTER 1824
The University of Manchester

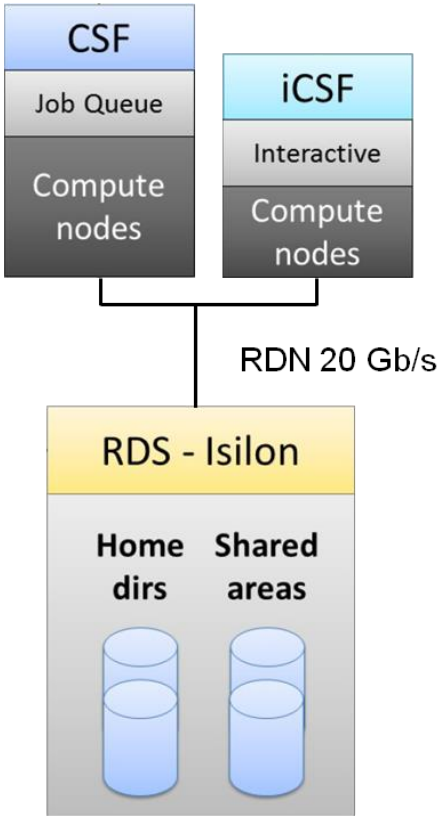
iCSF configuration

8 Standard Nodes	• 12 core nodes • 64 GB RAM
2 High-memory Nodes	• 16 core nodes • 256 GB RAM
1 Super-high-memory Node	• 2TB RAM • priority use for contributor
3 GPU nodes	• AMD FirePro v7800 GPUs

MANCHESTER 1824
The University of Manchester

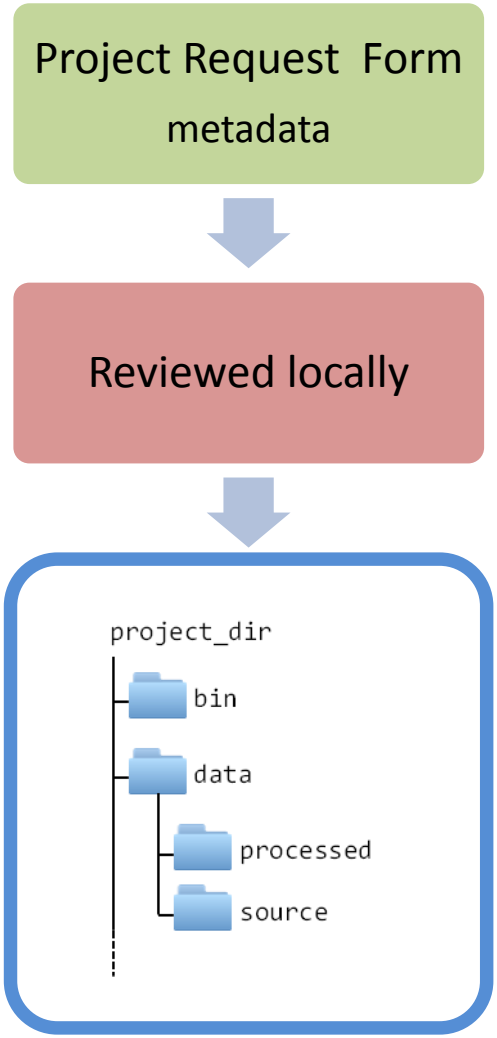
Research Data Storage

- Key feature: centralisation of compute and storage
- Tight integration of systems
- Storage linked to both clusters
  - High speed connection
  - Access home/project directories
- Highly resilient (snapshots and replication)
  - Not possible with our workstations



Project directory system

- Lots of users working on multiple projects (n=61)
- Important to organise this workspace
- Implemented a project directory system
- Currently have 89 projects (~20 TB)



MANCHESTER 1824
The University of Manchester

Transition

- Important transition period
  - Windows based systems to Linux
  - Most users had limited experience
- This was facilitated by a user workshop
  - Pen and George
- Core group of users
  - Now provide training/supervision to new users

MANCHESTER 1824
The University of Manchester

Three key issues

1. Dependent on existence
   - Changes would have an impact
   - Long term plans (cloud)

2. Maintaining access
   - Ad hoc grant contributions
   - Subscription (5 year access – multiple systems)

3. Archiving
   - Many completed projects
   - Interested in any centralised archiving processes

MANCHESTER 1824
The University of Manchester

Summary

Advantages of centralised systems:

1. No procurement, hosting, administration of local systems
2. Allows us to keep up the pace of data collection
3. Centralisation improves organisation
4. Not just about the hardware
   - access to expertise in Research IT is important

MANCHESTER 1824
The University of Manchester

Two projects:

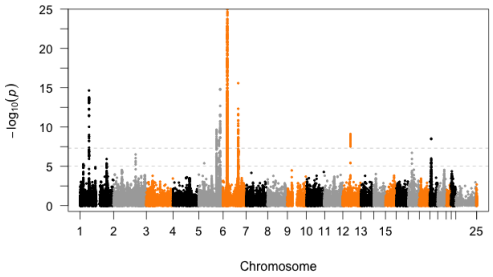
1. Data browser: RVMS and Research Software Engineers

2. Galaxy pipeline: DPSF and Centaurus server

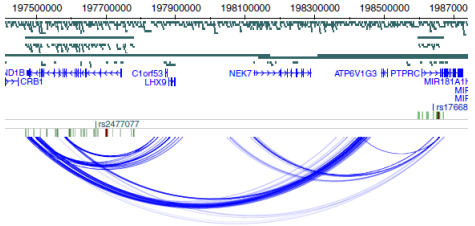
Data Browser

The Problem....

- We generate large volumes of data
- Not currently easy to browse in a unified way
- Sharing is ad hoc and with collaborators



Genetics



Genomics

Public data

MANCHESTER 1824
The University of Manchester

Data Browser

The Aim....

- Create a web-based data browser visualisation/exploration
- Publically available
- Mechanism for sharing data
- Encourage collaboration

WashU
EpiGenome
Browser

MANCHESTER 1824
The University of Manchester

Software engineers

- Developing with Rob Haines
- Team of professional software engineers
- Allows us to “hire” a RSE for a short period of time
- Not possible through conventional recruitment process

MANCHESTER 1824
The University of Manchester

RVMS

- Hosted on Research Virtual Machine Service (RVMS)
- Centrally hosted VMS
- Users maintain administrative control
- Tight integration with computing infrastructure

MANCHESTER 1824
The University of Manchester

CRAFT

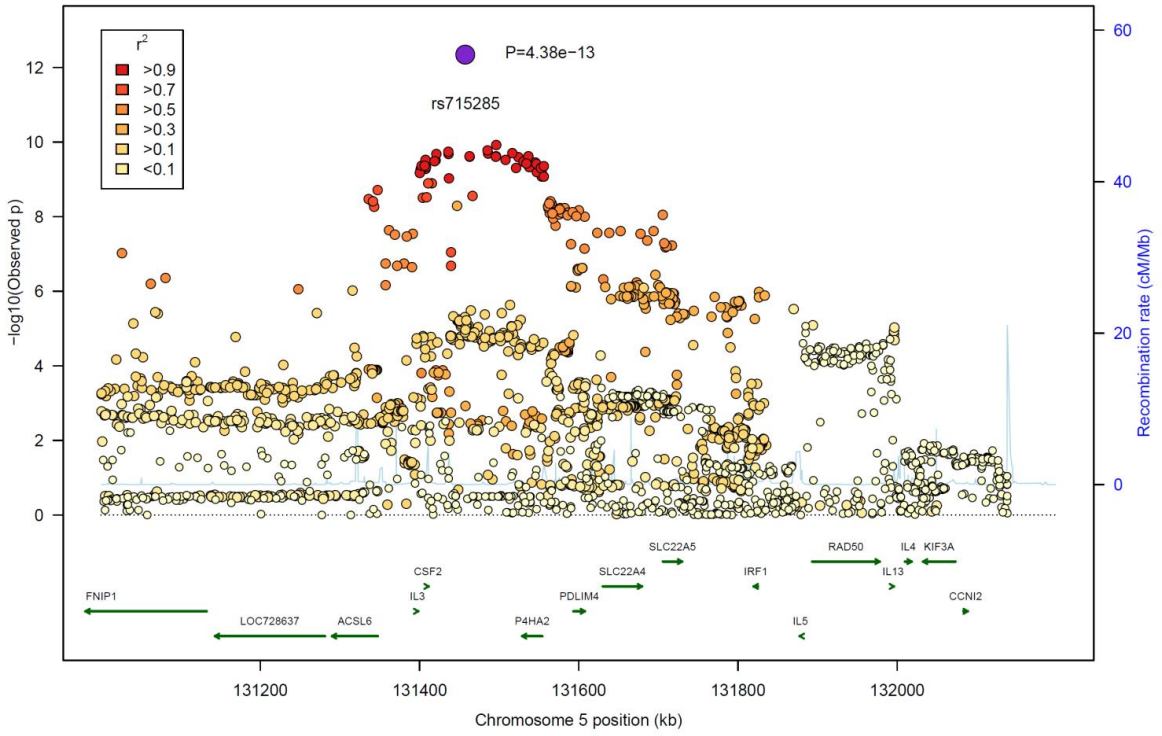
Galaxy

- **C**redible **R**efinement and **A**nnotation of **F**unctional **T**argets

- Pipeline for fine mapping genetic loci associated with disease

- Implemented in Galaxy
  - Bioinformatics platform for distributing workflows
  - Enables access for researchers without extensive computer experience

Fine mapping

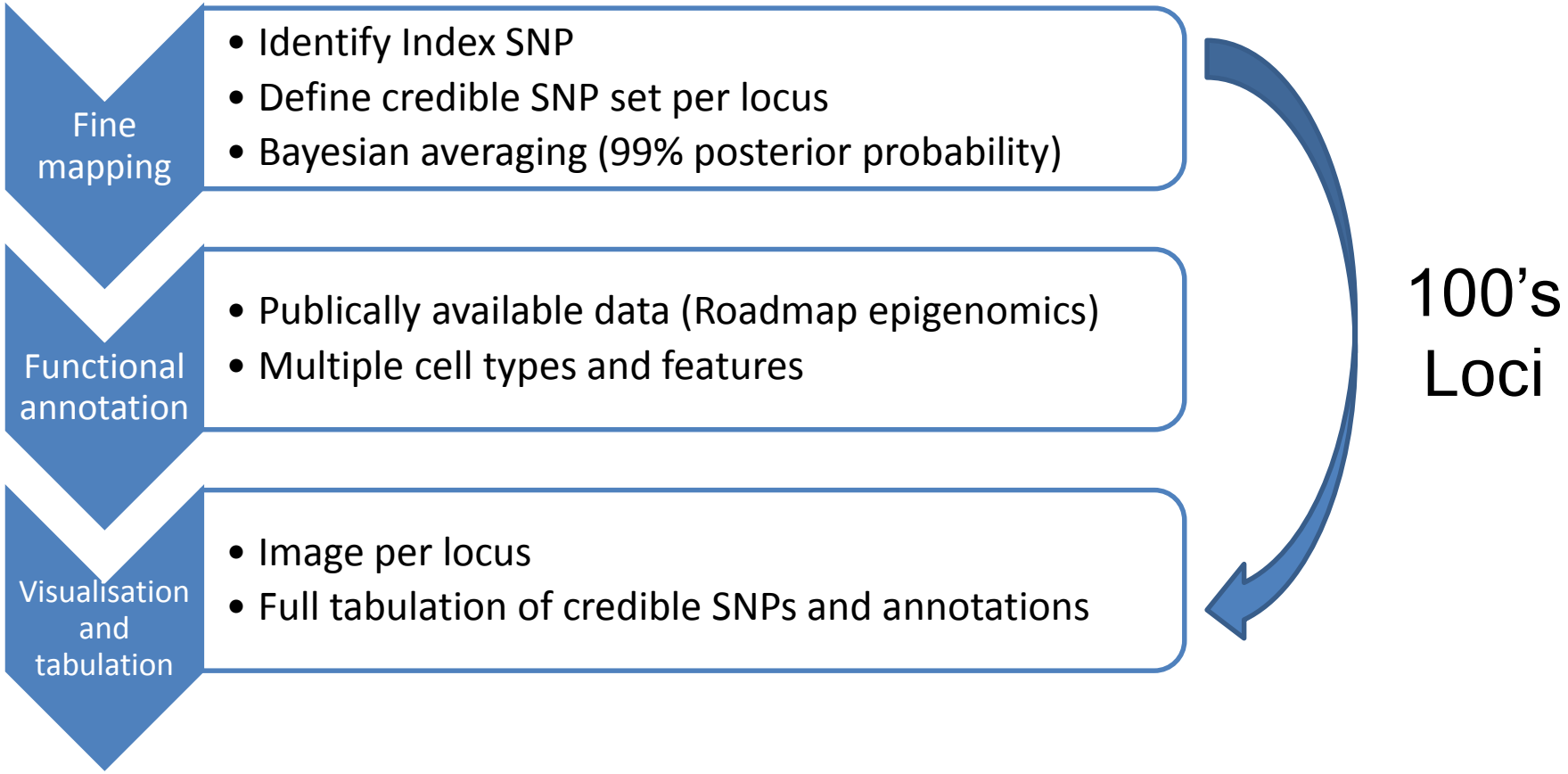


- Large region
- Lots of genetic variants
- High correlation between variants
- Multiple candidate genes

NEED: refine the association and prioritise functional variants

MANCHESTER 1824
The University of Manchester

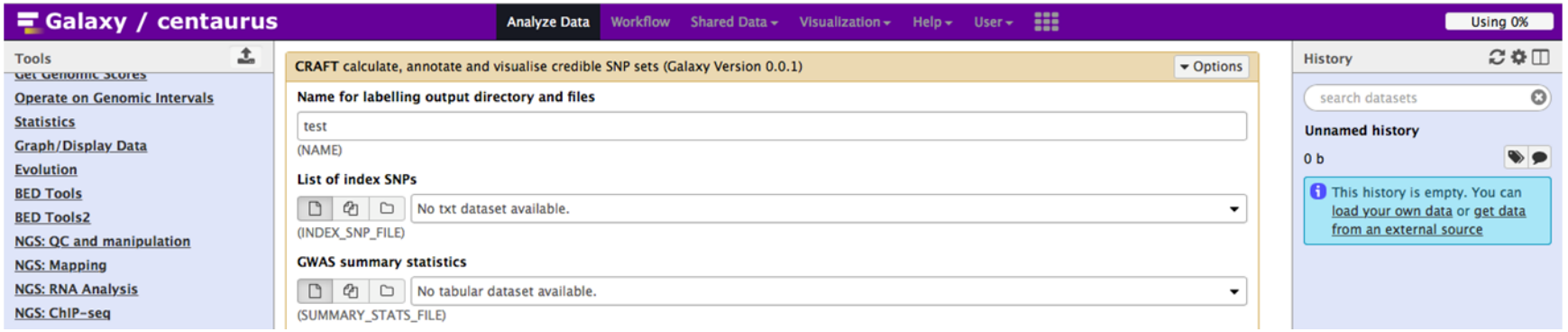
CRAFT stages



MANCHESTER 1824
The University of Manchester

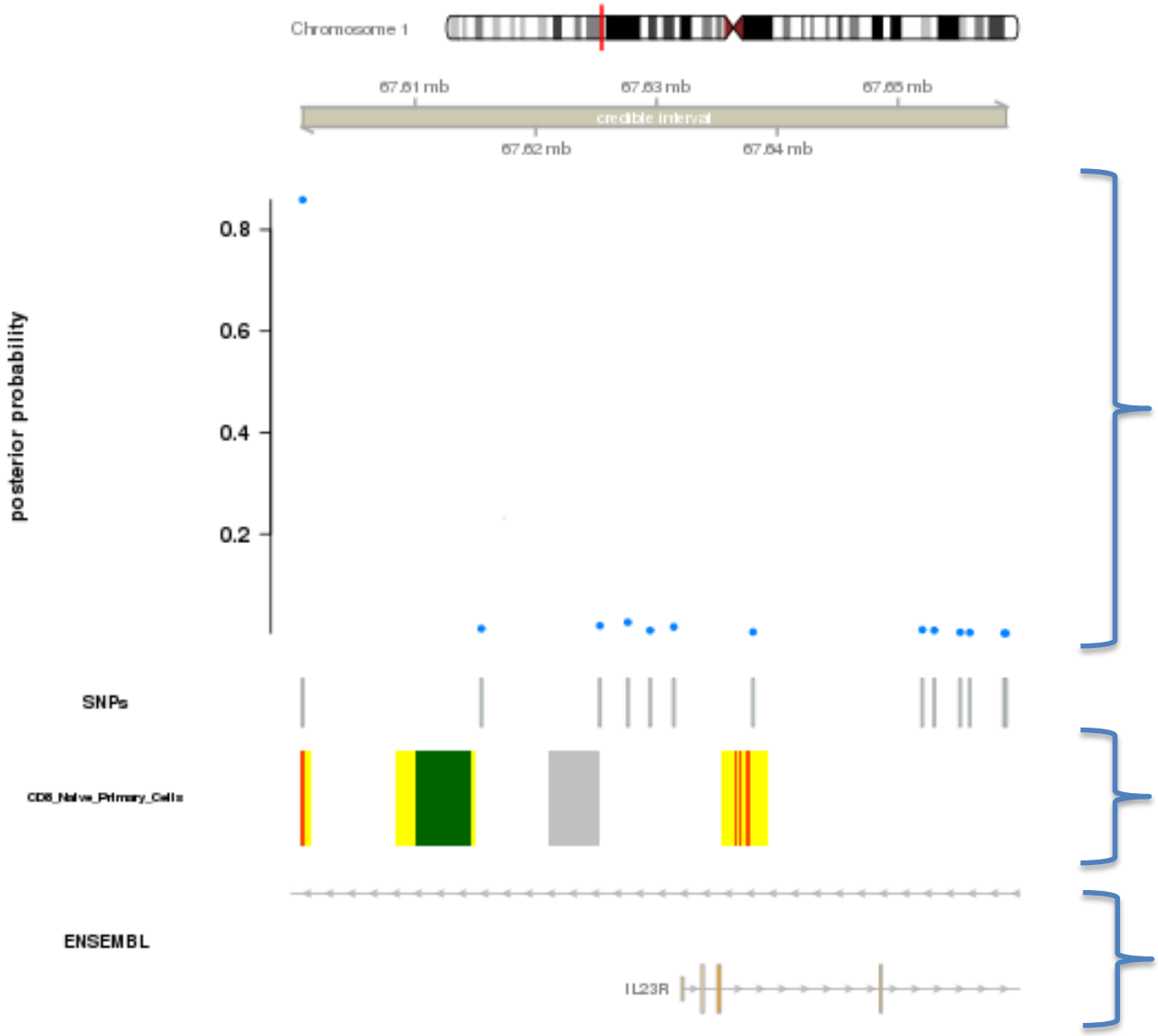
CRAFT: development

- Collaboration Peter Briggs (BCF)
- Hosted of Centaurus Galaxy Server
- Data Processing Shared Facility (DPSF)



MANCHESTER 1824
The University of Manchester

CRAFT: example



Credible SNP set (posterior probability)

ChrommHMM segmentations

Genes

MANCHESTER 1824
The University of Manchester

Conclusion

- Research IT systems now underpins the majority of our genetic research
- Access to both hardware and expertise has been important
- This has allowed us to keep pace with data collection

Acknowledgements

Research IT

Simon Hood

Pen Richardson

George Leaver

CfMR

Owen Stewart

Andrew Tracey

CRAFT

Peter Briggs

Damian Tarasek

Data Browser

Rob Haines

Funding

Jane Worthington

Anne Barton