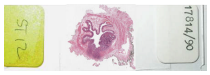
Visualising image, genomics and health records data

Professor Roy Ruddle

School of Computing, and Leeds Institute for Data Analytics
University of Leeds, UK
www.comp.leeds.ac.uk/royr/
r.a.ruddle@leeds.ac.uk

UNIVERSITY OF LEEDS

A world of unlimited data



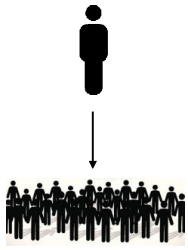
Pathology slides
10 gigapixels each
1−100+ slides/patient

systmone
emis

Health records
100s of events/patient per year

$1000 genome sequencing?

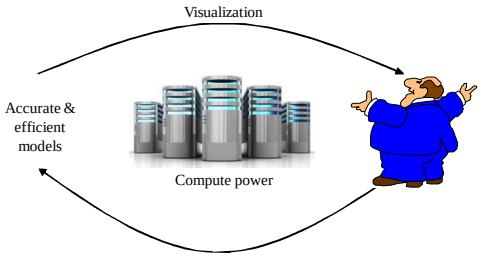
fitbit
Personal lifestyle



Cohorts & population

UNIVERSITY OF LEEDS

What do we need?



UNIVERSITY OF LEEDS

Visualization

- **Leeds Virtual Microscope (2006 to date)**
  - **Trillion-pixel image collections**
- **Orchestral (2012-13)**
  - **Comparative genomics (copy number data)**
- **Paramorama (2009-12)**
  - **Data processing pipelines**
- **QuantiCode (2016 to date)**
  - **Health records**

UNIVERSITY OF LEEDS

Leeds Virtual Microscope

- **Designed for pathologists**
  - **Fast cancer diagnosis from digital slides[1-3]**
  - **https://youtu.be/oZGkhKkDG5o**
- **Solution for Powerwalls & workstations[4,5]**
  - **Used at LTHT (since 2010)**
  - **… and 10 other Yorkshire & Humberside hospitals (2016)**
- **Novel concept & user interface (patented)**
  - **Being commercialised by Roche**

The Leeds Teaching Hospitals NHS
NHS Trust

[1]Treanor et al. (2009). *Histopathology.*
[2]Randell et al. (2013). *Histopathology.*
[3]Randell et al. (2014). *Human Pathology.*
[4]Randell et al. (2012). *CHI work in progress.*
[5]Ruddle et al. (2016). *ACM ToCHI.*

UNIVERSITY OF LEEDS

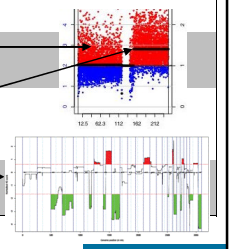
A killer application for Powerwalls

- **Investment**
  - **Powerwalls and/or 4k visualization workstations**
    - **Anatomy, Clinical Sciences, Computing, Earth & Environment, LICAP, LIDA, LTHT (St James' Hospital)**
  - **Free for you to use with any Windows software**

Orchestral

- **Proof of concept for analysing genomics data**

Steps in the data analysis pipeline	
Each sample	
1	Sequencing and align DNA
2	Calculate copy number variation (CNV). One data point per region of sample
3	Smooth/segment CNV data to remove noise
All samples	
4	Statistical analysis of significantly aberrant regions

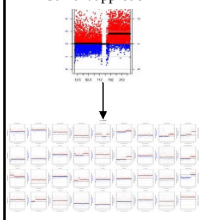


UNIVERSITY OF LEEDS

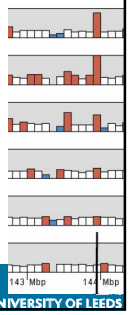
Fine-grained data in context of smoothed

- **Insight**[1]
  - **"data looks abnormally similar" → discovered processing error**
  - **Rare pattern is in fact common → limitation of CNAnorm**

Current approach



Orchestral

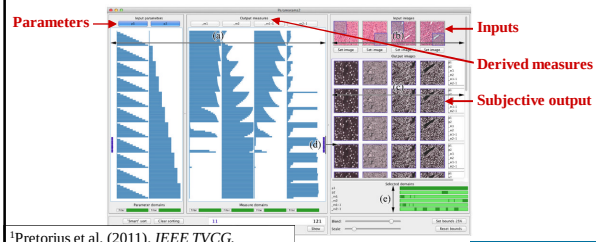


[1]Ruddle et al. (2013). *Proc. Biovis.*

UNIVERSITY OF LEEDS

Paramorama

- **Novel tool for designing & optimising data processing pipelines**
  - **Small but significant quality improvement (cell segmentation)**[1]
  - **Identify a subtle logic error (image segmentation)**[2]
  - **An assumption is invalid (colour deconvolution)**[3]



[1]Pretorius et al. (2011). *IEEE TVCG.*
[2]Pretorius et al. (2012). *Proc. SIGRAD.*
[3]Pretorius et al. (2015). *BMC Bioinformatics.*

UNIVERSITY OF LEEDS

QuantiCode

- **Event sequences**
  - **E.g., health records; social care; supermarket sales**
  - **Research into novel data mining & visualization techniques**
- **Visualization techniques to**
  - **Investigate data quality (missingness, etc.)**
    - **Tool available early 2017**
  - **Level of detail for analysis (abstraction models)**
  - **Stratification**

UNIVERSITY OF LEEDS

Collaborators & funders

Project	Collaborators
LVM	Rhys Thomas, Rebecca Randell, Phil Quirke, Darren Treanor (LTHT)
Orchestral	Peter Sondergeld, Waleed Fateen, Phil Quirke Darren Treanor (LTHT)
Paramorama	Hannes Pretorius, Yu Zhou, Duane Carey, Derek Magee, Andy Bulpitt, and John Fisher (Welmec PI) Darren Treanor & Stephen Smye (LTHT) Mark-Anthony Bray, Anne Carpenter (Broad Institute)
QuantiCode	Georgios Aivaliotis, Mark Birkin, Justin Keen, Kevin Macnish, Alex Markham, Chris Megone, Muhammad Adnan, Anna Palczewska, Jan Palczewski. aql, Bradford Institute for Health Research, Consumerdata, Leeds City Council, Leeds Informatics Board, NHS Digital, Sainsbury's,

EPSRC
NHS NIHR
NHS Health Education Yorkshire and the Humber
Pathological Society
yorkshire cancer research
Jisc
hefce

UNIVERSITY OF LEEDS