UNIVERSITY OF CAMBRIDGE

Department of Health

wellcome trust

Applications of WGS for surveillance and DST

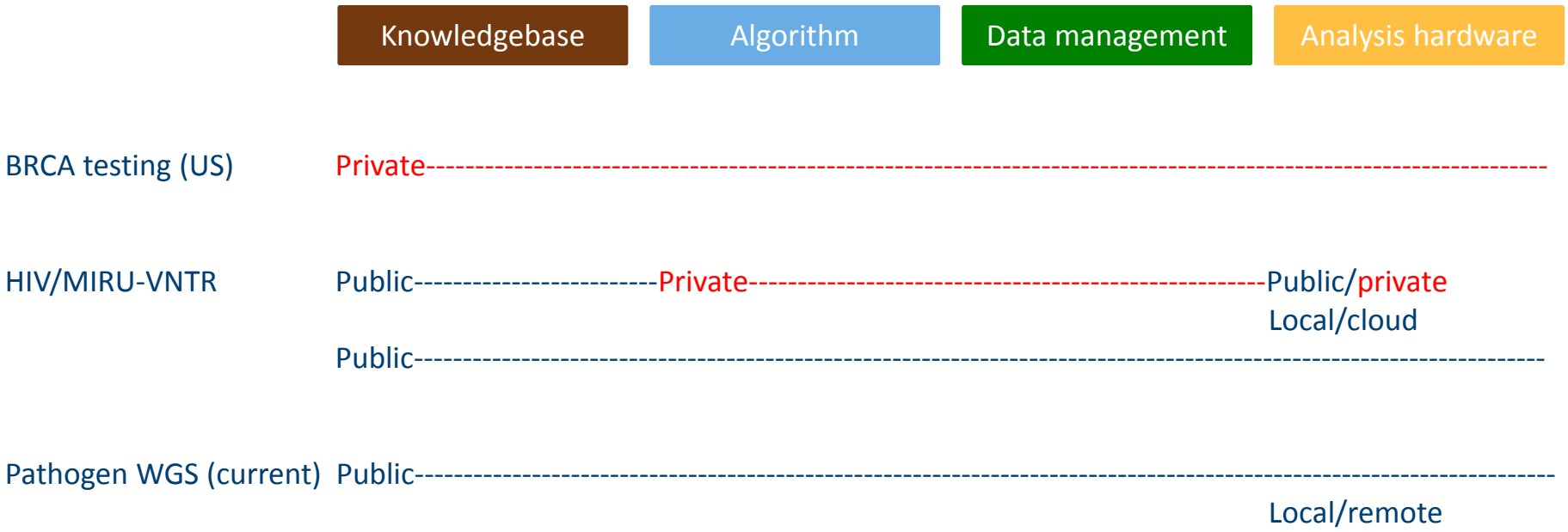
Dr Claudio U. Köser

Disclaimer

- This talk presents independent research supported by the Health Innovation Challenge Fund (HICF-T5-342 and WT098600), a parallel funding partnership between the UK Department of Health and Wellcome Trust. The views expressed in this talk are my own and not necessarily those of the Department of Health or Wellcome Trust.
- Bill & Melinda Gates Foundation and Janssen Pharmaceutica covered my travel expenses to attend a workshop and meeting, respectively.

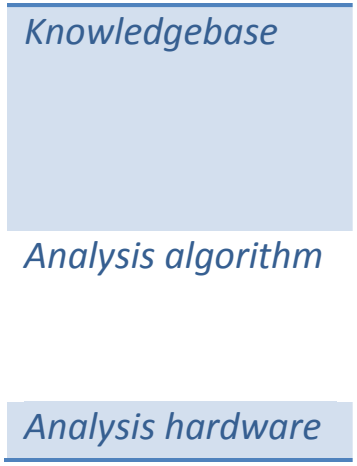
Acknowledgements

- Matthew Ellington & Jane Greatorex (Public Health England Cambridge)
- Michael Head (University College London)
- Danesh Moradigaravand (Wellcome Trust Sanger Institute)
- Sharon Peacock (University of Cambridge)


Knowledgebase
Algorithm
Data management
Analysis hardware
BRCA testing (US)
Private
HIV/MIRU-VNTR
Public
Private
Public/private
Local/cloud
Public
Pathogen WGS (current)
Public
Local/remote

MIRU-VNTR paradigm not appropriate for WGS

- Developed by academics who also maintained websites
- Commercial tools only emerged because they were more user-friendly (especially for data management and exchange), provided automated record-keeping and more intuitive ways to overlay metadata – customer service and legal liability



- The challenge is therefore how to improve and maintain surveillance infrastructure, particularly in light of what has happened to TB databases in the past

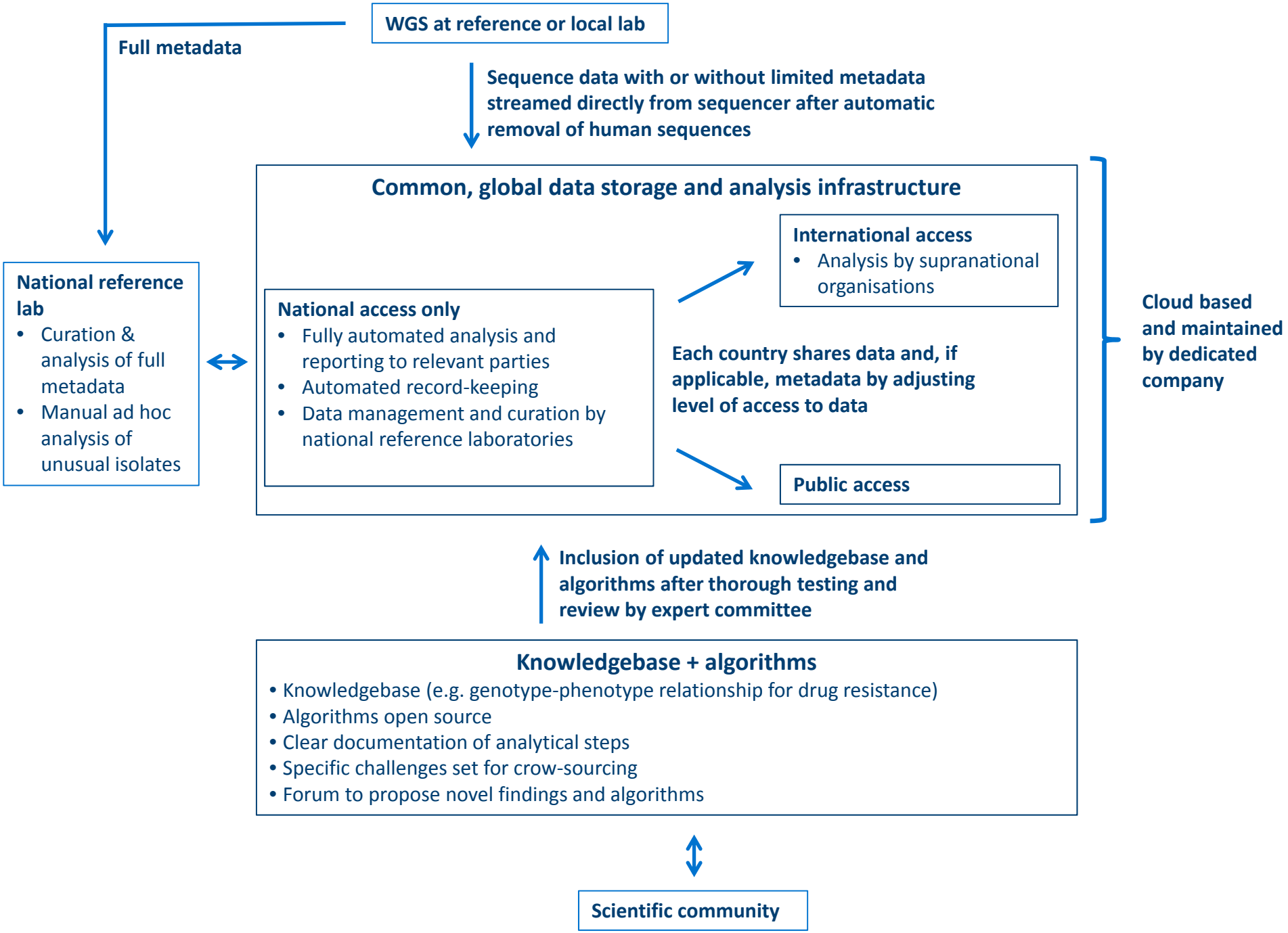
Surveillance driven by WGS diagnostics

- Initially, WGS for research/surveillance only
- WGS of every single new culture-positive case will become standard of care in well-resourced countries – i.e. WGS will be done for diagnostic purposes, rather than surveillance

Region	Country	Notified TB cases	Estimated culture-positive cases
EU & associated countries		69,893	45,876
North America	Canada	1,686	1,012
	USA	9,945	5,967
Wealthy Asian, Middle Eastern and Pacific countries	Australia	1,325	795
	Bahrain	225	135
	Hong Kong	4,969	2,981
	Israel	509	305
	Japan	21,283	12,770
	Kuwait	737	442
	Macao	406	244
	New Zealand	297	178
	Qatar	728	437
	Republic of Korea	49,532	29,719
	Saudi Arabia	3,833	2,300
	Singapore	2,364	1,418
	United Arab Emirates	85	51
Total		**167,817**	**104,630**

Surveillance driven by WGS diagnostics

- Analysis tools have to be robust and scalable to allow decentralised WGS
  - Sequence data streamed to the cloud for storage and analysis
  - Fully automatic analysis and reporting to relevant parties within minutes of completion of sequence run – manual analysis only for unusual isolates
- Pay for access to data (public could pay for or subsidise cloud infrastructure for storage and analysis at a central location such as EBI/NCBI)
- Individual countries need to retain control of their data
  - Data do not have to be shared with the public immediately
  - No metadata have to be shared


WGS at reference or local lab
Full metadata
Sequence data with or without limited metadata streamed directly from sequencer after automatic removal of human sequences
Common, global data storage and analysis infrastructure
International access
• Analysis by supranational organisations
National reference lab
• Curation & analysis of full metadata
• Manual ad hoc analysis of unusual isolates
National access only
• Fully automated analysis and reporting to relevant parties
• Automated record-keeping
• Data management and curation by national reference laboratories
Each country shares data and, if applicable, metadata by adjusting level of access to data
Public access
Cloud based and maintained by dedicated company
Inclusion of updated knowledgebase and algorithms after thorough testing and review by expert committee
Knowledgebase + algorithms
• Knowledgebase (e.g. genotype-phenotype relationship for drug resistance)
• Algorithms open source
• Clear documentation of analytical steps
• Specific challenges set for crow-sourcing
• Forum to propose novel findings and algorithms
Scientific community

Complexity of WGS is a blessing in disguise

- Clear economic/political incentives
  - Continued maintenance very likely
  - Everybody would get access to the same quality of analysis
  - Everybody would benefit from economies of scale
  - Differential pricing could be used to encourage adoption in high-incidence countries
- EBI and NCBI have already begun to build an open storage and compute infrastructure as part of Global Microbial Identifier Initiative
- Public and philanthropic funders have to take active role in shaping market
  - Private enterprise cannot develop analysis tools without access to the knowledgebase, analysis algorithms, sequence data/metadata and laboratories
  - Manufacturers of sequencers have strong incentive to grow these markets
  - Public/private company with international governance structure (akin to EUCAST) that reinvests profits in maintenance of infrastructure or research?

Thank you very much for your attention