

# Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance

Prof. Willem van Schaik

Professor in Microbiology and Infection



UNIVERSITY OF  
BIRMINGHAM

Institute of Microbiology and Infection  
University of Birmingham, United Kingdom  
[w.vanschaik@bham.ac.uk](mailto:w.vanschaik@bham.ac.uk)



WvSchaik

# Sequence-based pathogen surveillance



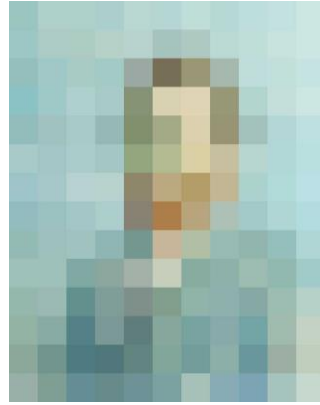
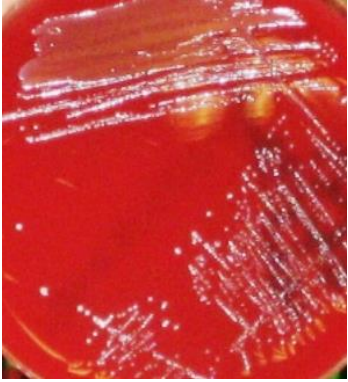
use sequence data to infer relatedness of strains



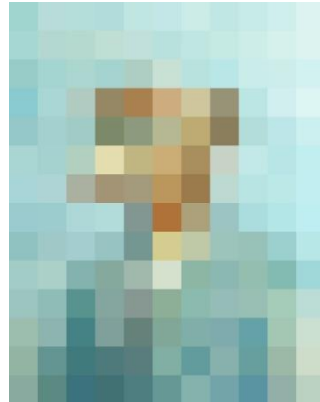
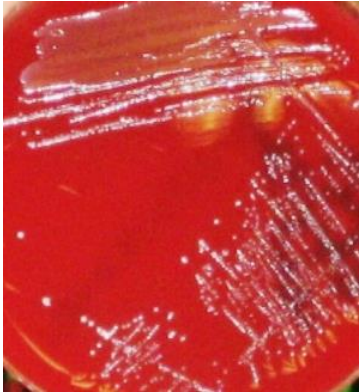
## MLST

(allelic profile of 5 – 7 housekeeping genes)

Strain A



Strain B



Low resolution

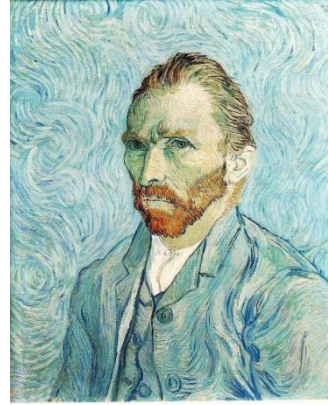
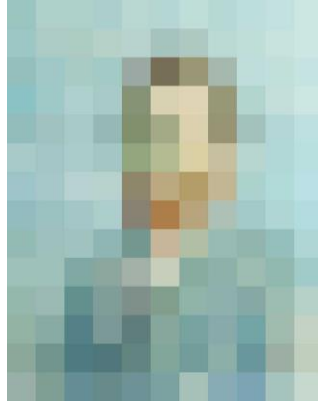
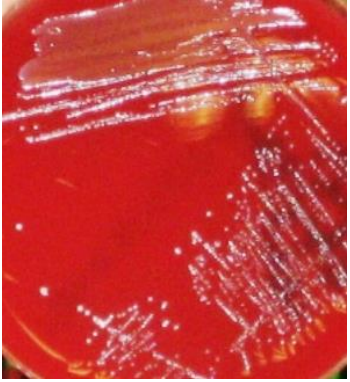
Differences may be apparent, but not clear

Provides a nomenclature to describe clones (e.g. *E. coli* ST131)

MLST

WGS

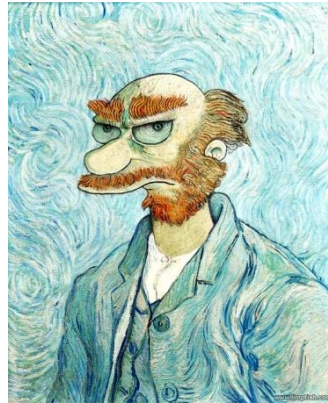
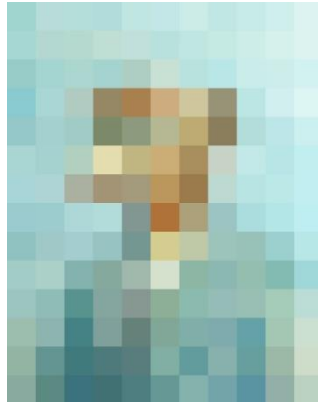
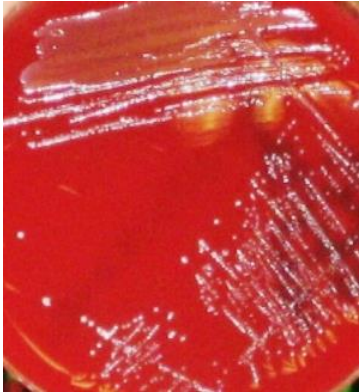
Strain A



High resolution

Core genome phylogeny  
(SNP-based, core  
genome MLST)

Strain B



Whole genome  
phylogeny  
(whole genome MLST)

Many tools

Which ones to choose?

Or use the tools  
developed by  
Torsten Seemann  
(Univ of  
Melbourne)



AMERICAN  
SOCIETY FOR  
MICROBIOLOGY

Clinical Microbiology  
Reviews®

## **Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis**

Scott Quainoo,<sup>a</sup> Jordy P. M. Coolen,<sup>b</sup> Sacha A. F. T. van Hijum,<sup>c,d</sup>  
Martijn A. Huynen,<sup>c</sup> Willem J. G. Melchers,<sup>b</sup> Willem van Schaik,<sup>e</sup>  
Heiman F. L. Wertheim<sup>b</sup>

Clin Microbiol Rev 30:1015–1063

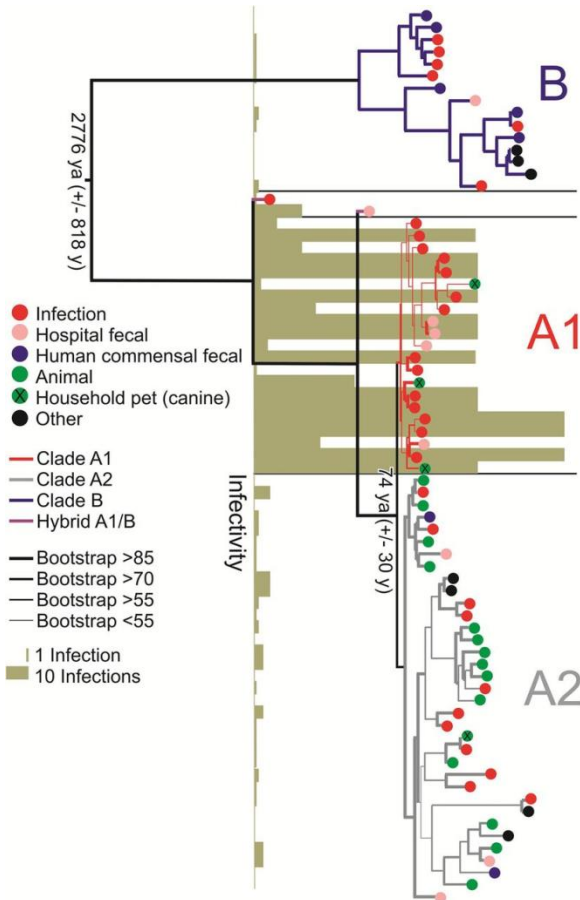
But there is more in a whole genome sequence...

Antibiotic resistance genes

Presence of mobile genetic elements



# Vancomycin-resistant *Enterococcus faecium*



“Human commensal”

“Clinical isolates”

“Animal isolates”

Maximum likelihood tree: concatenated alignment of 1344 core genes in 73 *E. faecium* genomes

Lebreton, van Schaik, McGuire *et al.*, 2013. *Mbio* 4:e00534-13

## WHO priority list of pathogens

### Priority 1: CRITICAL<sup>#</sup>

*Acinetobacter baumannii*, carbapenem-resistant  
*Pseudomonas aeruginosa*, carbapenem-resistant  
*Enterobacteriaceae*\*, carbapenem-resistant, 3<sup>rd</sup> generation cephalosporin-resistant

### Priority 2: HIGH

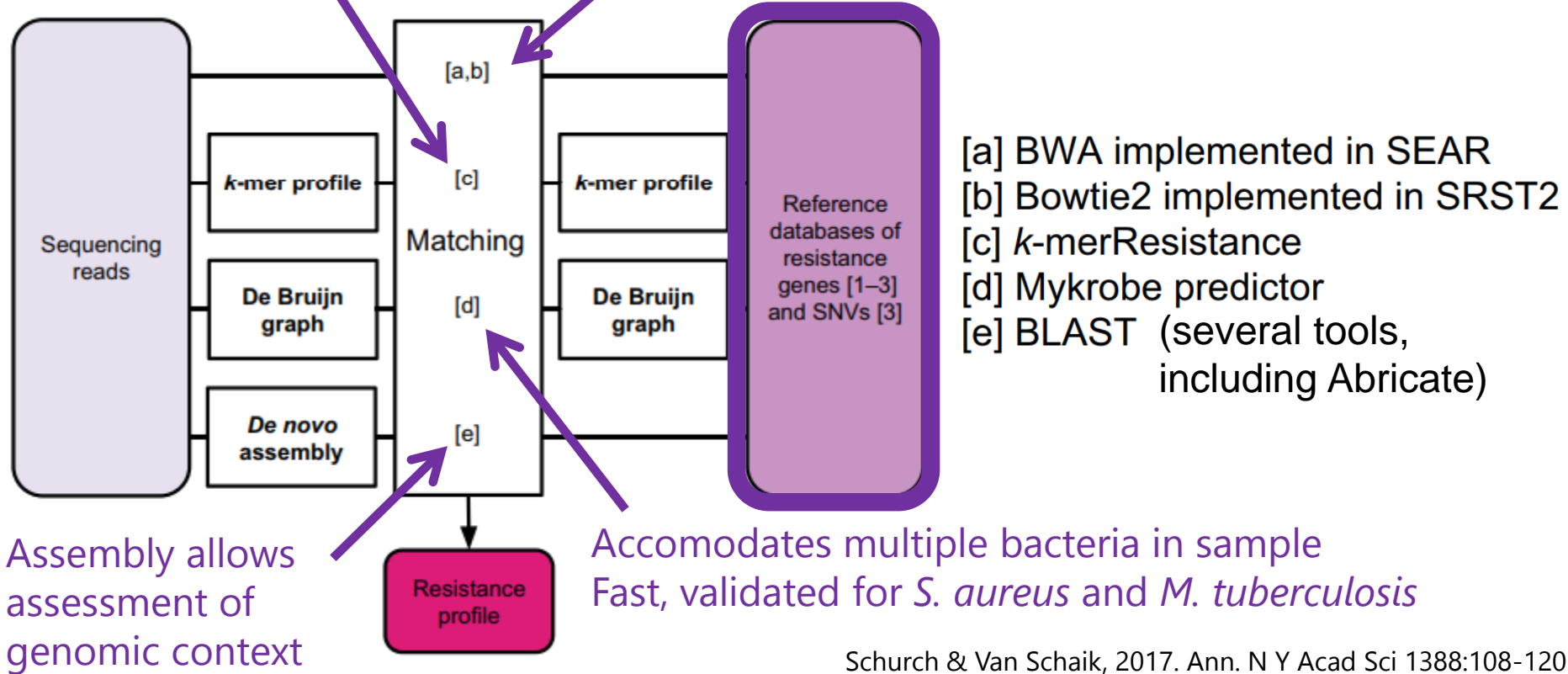
*Enterococcus faecium*, vancomycin-resistant  
*Staphylococcus aureus*, methicillin-resistant, vancomycin intermediate and resistant  
*Helicobacter pylori*, clarithromycin-resistant  
*Campylobacter*, fluoroquinolone-resistant  
*Salmonella* spp., fluoroquinolone-resistant  
*Neisseria gonorrhoeae*, 3<sup>rd</sup> generation cephalosporin-resistant, fluoroquinolone-resistant

### Priority 3: MEDIUM

*Streptococcus pneumoniae*, penicillin-non-susceptible  
*Haemophilus influenzae*, ampicillin-resistant  
*Shigella* spp., fluoroquinolone-resistant

Fastest option

Direct mapping of reads: no loss of data





# What is an antibiotic resistance gene?

“a gene that confers resistance to an antibiotic in an otherwise susceptible microbial host”

 Horizontally acquired resistance genes: e.g. *bla*KPC, *ermB*, *vanA*  
associated with mobile genetic elements

Mutations in housekeeping genes that confer resistance  
e.g. target modification, porin mutation 

 Conserved genes that make a species intrinsically resistant  
e.g. efflux pumps

# Antibiotic Resistance Gene Databases

ARDB: no updates, many intrinsic resistance genes: do not use

CARD: frequently updated, based on ARDB. Contains some intrinsic resistance genes; database for 'resistance SNPs' but can give false positives

ResFinder	}	Frequently updated. Acquired resistance genes only, small number of intrinsic resistance genes.
ARG-ANNOT		

ResFams: prediction of resistance genes from metagenomes, many false positives or unvalidated genes

# Antibiotic Resistance Gene Databases

*Subject Section*

**ARGs-OAP v2.0 with an Expanded SARG Database and Hidden Markov Models for En-**

**hanc  
of A  
tal M**

**SOFTWARE**

**Open Access**

Xiaole  
R. Col

**DeepARG: a deep learning approach for  
predicting antibiotic resistance genes from  
metag**



**MEGARes: an antimicrobial resistance database for  
high throughput sequencing**

Gustavo Arangi

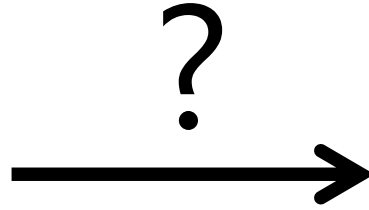
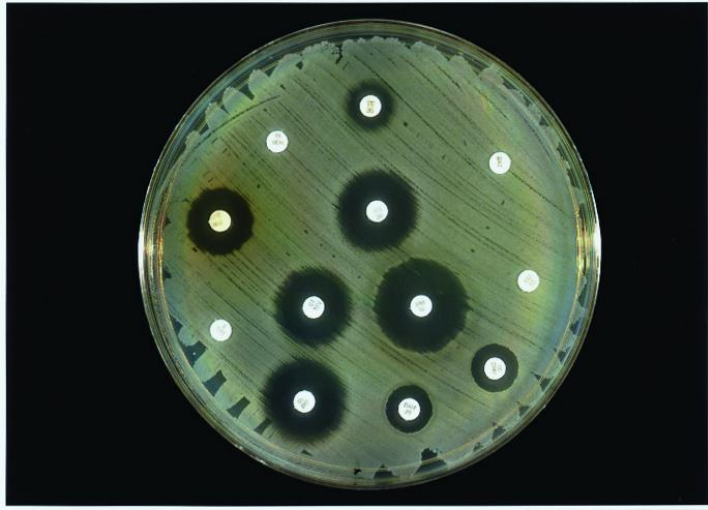
Steven M. Lakin<sup>1,†</sup>, Chris Dean<sup>2,†</sup>, Noelle R. Noyes<sup>2,†</sup>, Adam Dettenwanger<sup>3</sup>, Anne  
Spencer Ross<sup>3</sup>, Enrique Doster<sup>1</sup>, Pablo Rovira<sup>4</sup>, Zaid Abdo<sup>2</sup>, Kenneth L. Jones<sup>5</sup>,  
Jaime Ruiz<sup>6</sup>, Keith E. Belk<sup>4</sup>, Paul S. Morley<sup>1</sup> and Christina Boucher<sup>6,\*</sup>

# Antibiotic Resistance Genes

When reporting antibiotic resistance genes....



# Sequence-based diagnostics of antibiotic resistance genes?



# Sequence-based diagnostics of antibiotic resistance genes?

*Mycobacterium tuberculosis*



Very slow growth

No horizontal gene transfer

Large collection of strains +  
resistance phenotypes

WGS accurately predicts  
susceptibility 93.6% – 99.0%  
to four key antibiotics

CRyPTIC Consortium and the 100,000 Genomes  
Project. New Engl J Med. 2018 379:1403



# Sequence-based diagnostics of antibiotic resistance genes?

Common causes of bloodstream infections

*Escherichia coli*  
*Staphylococcus aureus*  
*Enterococcus*

Fast growth

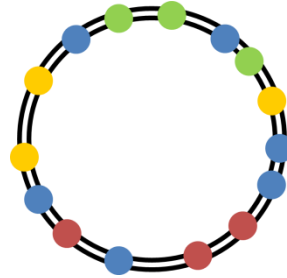
Horizontal gene transfer

Emergence of new resistance determinants (e.g. *mcr*)

Fast WGS (e.g. of bloodstream cultures) for screening while susceptibility testing is ongoing



# Prediction of Plasmids from Illumina Data



Sequence repeats in  
plasmid sequences  
complicates assembly

42 completely assembled genomes  
12 genera

148 plasmids (1.5 – 339 kbp)

?

Illumina + PacBio sequences  
available

# Prediction of Plasmids from Illumina Data

MISSION: IMPOSSIBLE

or

MISSION: POSSIBLE

## Comparison of completely automated approaches

PlasmidFinder: web-based tool to detect replicon sequences in contigs *Carattoli et al., Antimicrob Agents Chemother. 2014 58:3895*

cBar: detects differences in k-mer composition between plasmids and chromosomes *Zhou & Xu Bioinformatics 2010 26:2051*

Recycler: extracts circular sequences from De Bruijn assembly graphs *Rozov et al. Bioinformatics 2017 33:475*

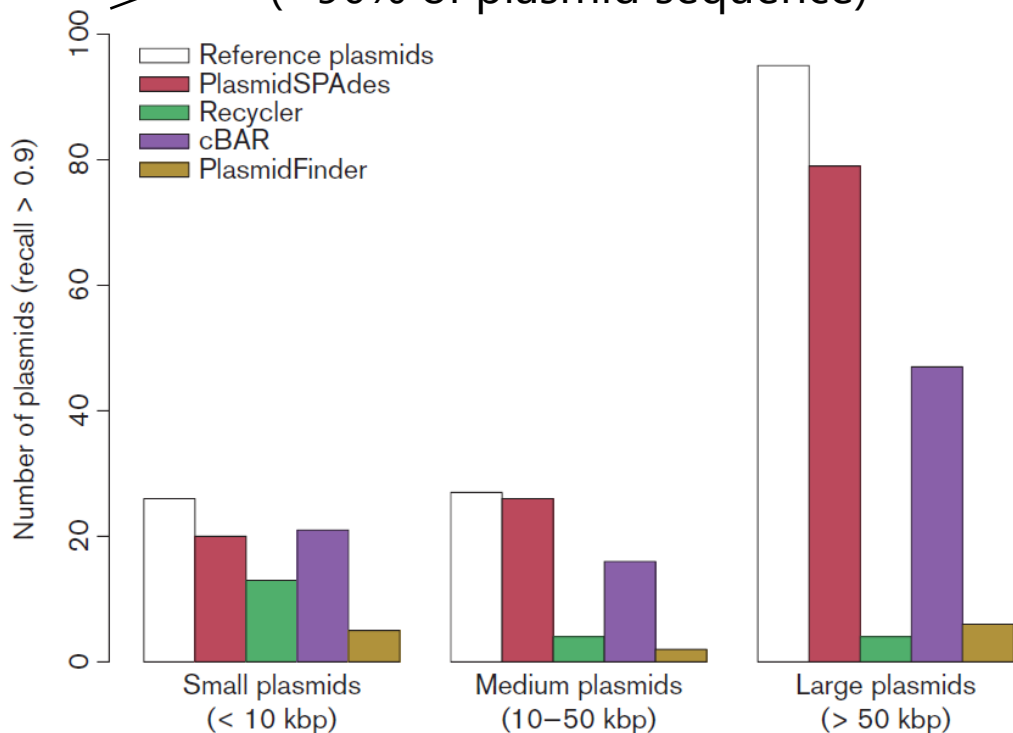
PlasmidSPAdes: plasmids have different coverage (higher copy number) than chromosome *Antipov et al. Bioinformatics 2016 32:3380*

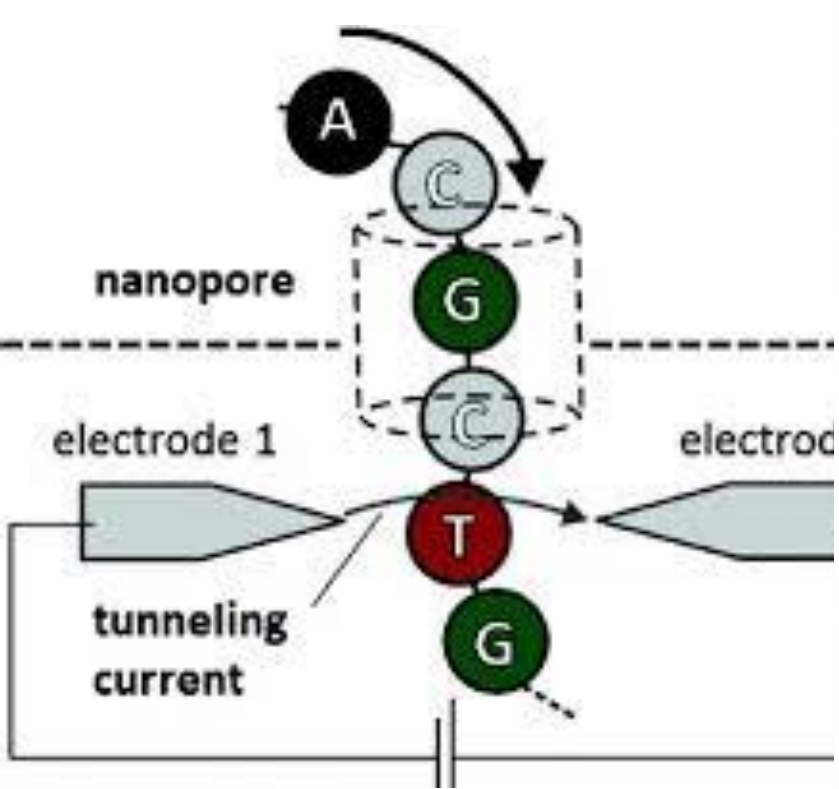
# Prediction of Plasmids from Illumina Data

Number of plasmids correctly predicted  
(>90% of plasmid sequence)

Results are variable

Good performance  
PlasmidSPAdes, but all  
plasmids in single bin

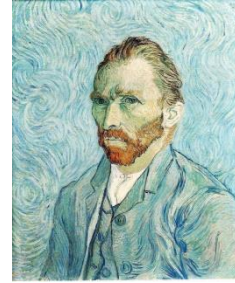




Long reads can cover repeats

# Summary

Whole genome sequencing is required for high-resolution typing







UNIVERSITY OF  
BIRMINGHAM

Ann Snaith  
Maria Papangeli  
Steven Dunn  
Alan McNally



UMC Utrecht

Sergio Arredondo-Alonso  
Rob Willems  
Anita Schürch



University Hospitals  
Birmingham  
NHS Foundation Trust

Abid Hussain  
Katie Hardy

## Funding



UNIVERSITY OF  
BIRMINGHAM

