AMR Prediction tools and databases: an overview

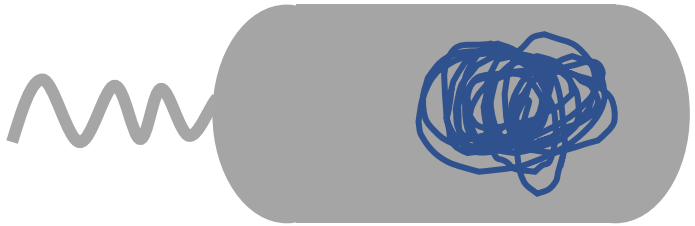
Dr. Narender Kumar

Senior Bioinformatician

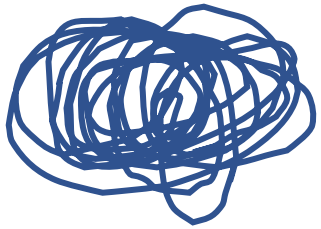
Global Pneumococcal Sequencing Project
Prof. Stephen Bentley's Group

wellcome sanger institute

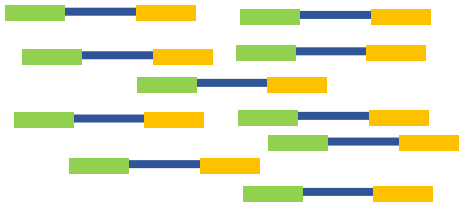
Whole genome sequencing



Bacteria



Genomic DNA isolation



Library Preparation

Reads

Read 1 Read 2

Illumina Sequencer

wellcome sanger institute

Applications

Whole Genome Sequencing

1. Pathogen diversity and surveillance
2. Transmission tracking
3. Outbreak investigation
4. AMR prediction and surveillance

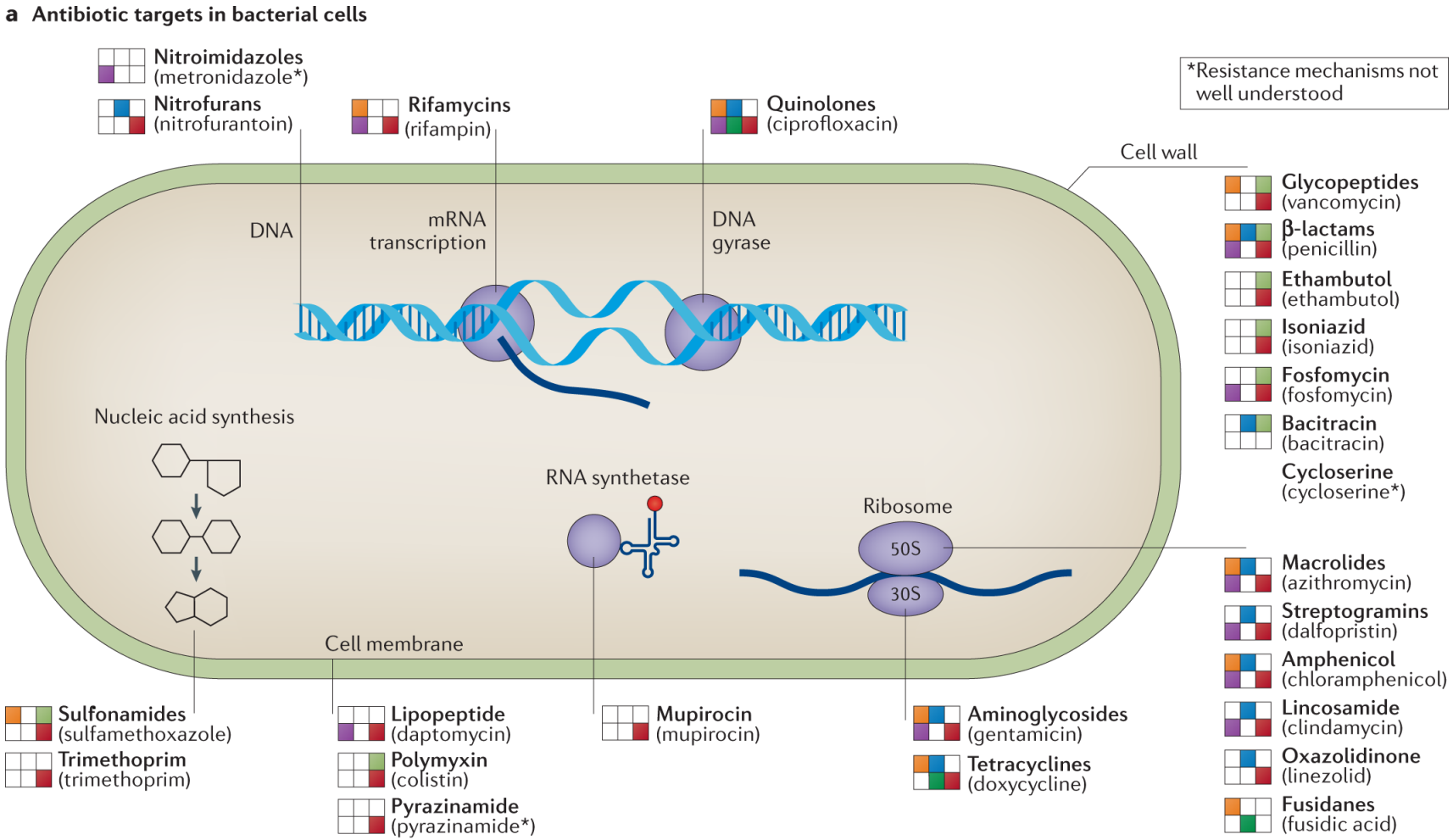
wellcome sanger institute

Predicting antimicrobial resistance

1. **Information on resistance mechanisms**
   - Acquired - Example: *mecA*
   - Chromosomal mutations - Example: *gyrA* mutations
   - Resistance databases: **CARD, ResFinder and AMRFinder**
2. **Whole Genome Sequence**
3. **Analysis and interpretation tools**

wellcome sanger institute

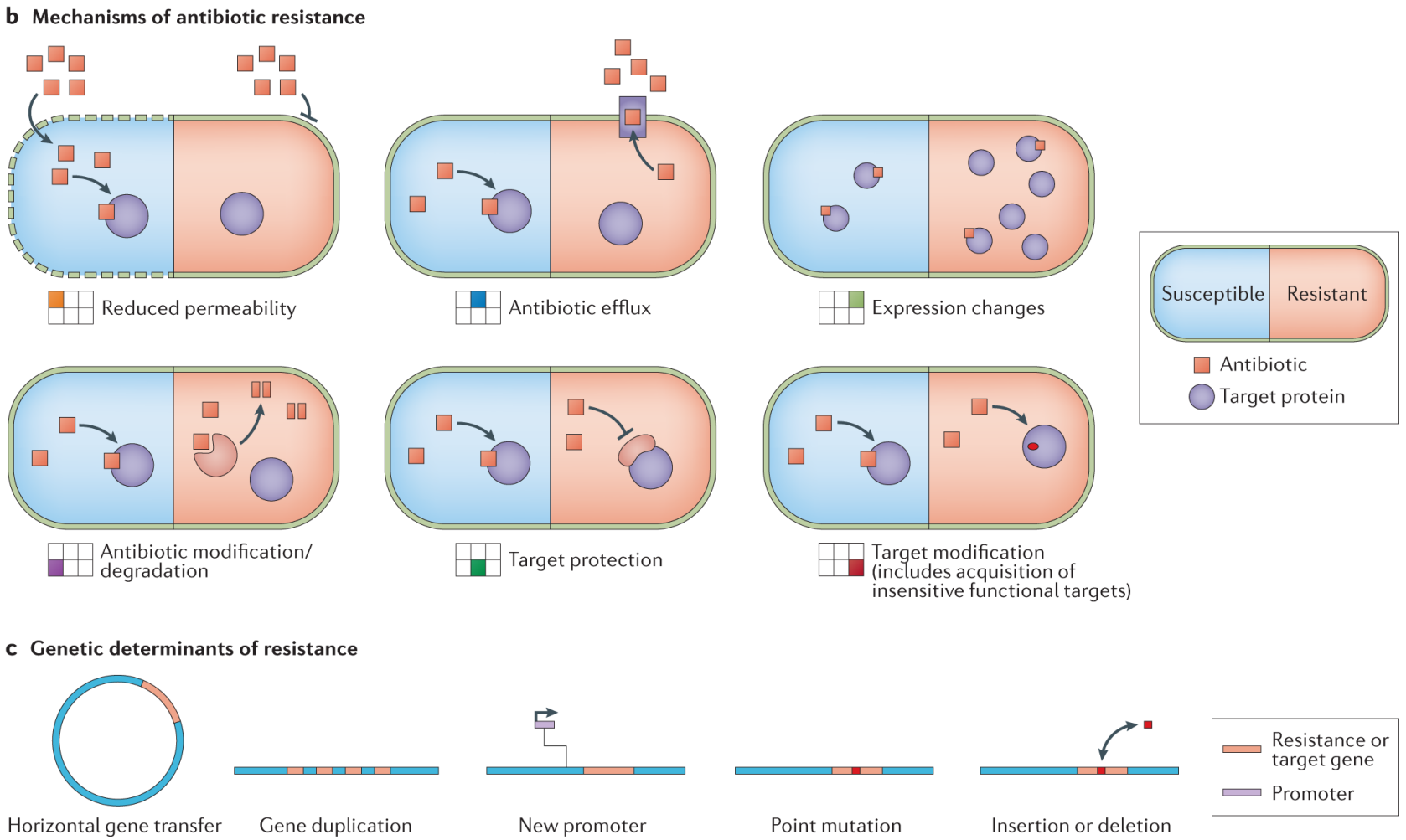
Antibiotics



wellcome sanger institute

Nature, 2019, doi: 10.1038/s41576-019-0108-4

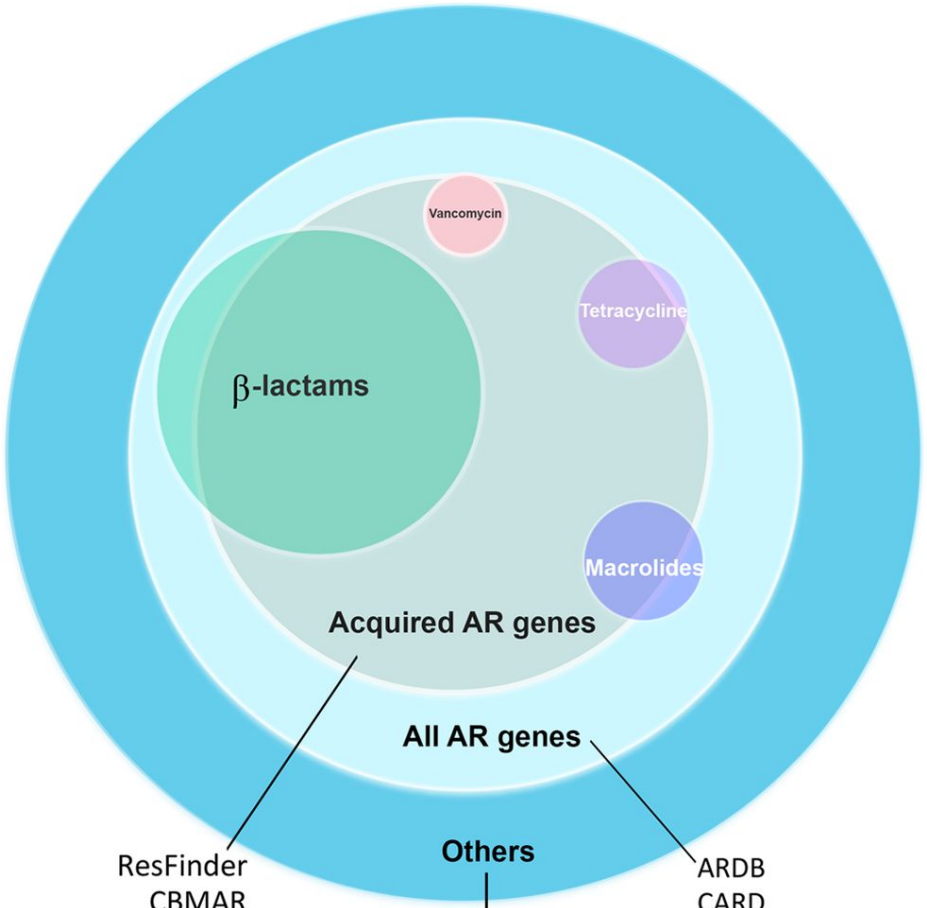
Resistance mechanisms



wellcome sanger institute

Nature, 2019, doi: 10.1038/s41576-019-0108-4

Antimicrobial resistance databases



wellcome sanger institute

Xavier et al. 2016, JCM

Predicting phenotypic resistance

AMRFinder (*Feldgarden M et al. 2019*):

Validated on 6,242 isolates across 3 different species

concordance >98%

Mycobacterium tuberculosis (*Hunt et al. 2019*):

validated against 10207 isolates

concordance 94%

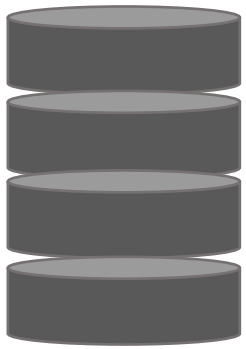
Staphylococcus aureus (*Kumar N et al. 2020*):

validated against 778 isolates

concordance 99%

wellcome sanger institute

CARD



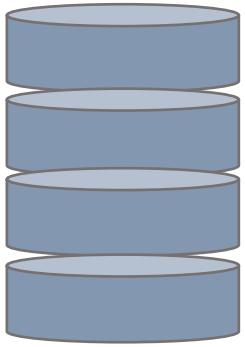
- Determinants from 263 pathogens (N > 100,000) genes, SNPs and plasmids
- Gene ontology structure (6453 GO Terms) for characterization
- Uses both BLAST, homology and SNP models
- latest versions with Kmer based algorithms
- Integrated within easy-to-use webtool

Usage considerations:

- Results need to be filtered based on the cutoffs used (manual curation)
- Species nonspecific predcitions
- Prediction limited to class of drugs in certain instances
- Limited upload size not suitable for large scale usage
- Accepts only assembled sequences

wellcome sanger institute

ResFinder



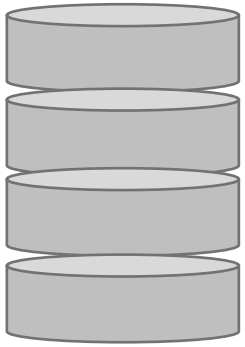
- Reference database contains 2919 genes, 266 SNPs and >57 antimicrobials
- Regularly curated and updated
- Uses Kmer based algorithms
- Integrated within easy-to-use webtool and command line version also available

Usage considerations:

- Species specific predictions
- Minimal curation of the results needed
- Prediction limited to class of drugs in certain instances
- Can accept multiple sequences at a time
- Accepts both assembled sequences and reads

wellcome sanger institute

AMRFinder



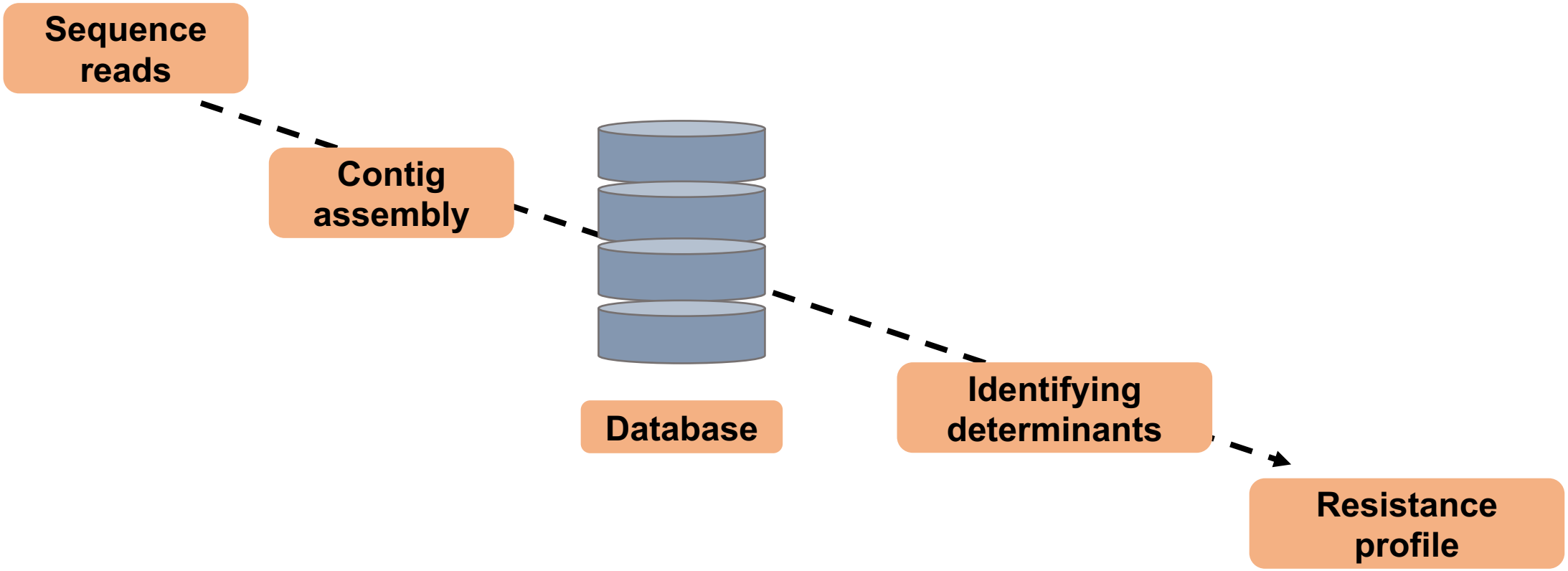
- Reference database contains 4579 proteins, 560 HMM
- Regularly curated and updated maintained by NCBI
- Allows identification of potentially novel mechanisms

Usage considerations:

- Command line only tool
- Results include drug class
- Can accept multiple sequences at a time
- Accepts assembled sequences

wellcome sanger institute

Detecting resistance



wellcome sanger institute

Prediction tools

Prediction tools	Databases used
Web tools	
Pathogenwatch	own
ResFinder	own
CARD-RGI	own
Command line tools	
NCBI AMRFinder	own
ResFinder	own
SRST2	Derived from ARG-ANNOT
ABRICATE	ResFinder, CARD, AMRFinder, PlasmidFinder, VFDB

Considerations

- Input files reads/assemblies
- Comprehensiveness
- Curation
- Regularly updated
- Easy to access
- Comparability
- Discrepancies between phenotype and genotype

wellcome sanger institute

Acknowledgement

GPS team

Prof. Stephen Bentley

SAGESA network

Thank you

wellcome sanger institute