ComMet

Microbiomes and Metagenomes

Warwick Business School, The Shard, London

3rd-4th March

2nd breakout group notes

To consider:

- Internal control
- Normalisation
- Standardisation
- Top3 things for comparability

Standardisation

- 16S which region to use? Illumina typically v3-v4 changes with technology. Use full length PacBio? Look at 16S rRNA/16S rDNA to see how active different fractions are.
- Difficult to standardise as so different with varying sources of material. Difficulties with low abundance OTU.
- Be good to have a 'kit' to allow standardisation of analysis
- Need to add the control prior to the extraction to account for extraction variation OR have a mock community alongside. There is a need for experimental standards to be added.
- Standards are pointless from protocol point of view as it all depends what you want to look at. Instead – be more candid with storage methods, extra flag, metadata.
- Need a mimimum level of information about metadata as standard.
- For a standard to be used it must be representative of the system.
- N.B. Bacteriodes disappears in freezer -20/-80C how store? Liquid N_2
- Depends on the claims- extraordinary claims need extraordinary evidence.

Normalisation

- Issues with copy number variation, can we standardise? Medical community use reference database based on genome database to estimate.
- Harder for environmental samples
- Even more of an issue for eukaryotes with log fold changes and no real database.
- Currently people use total read number for each run so use proportion rarefaction or the Chris Quince method or count based method. Rarefaction (See Susan Holmes paper 2014), phyloseq (manipulating OTU table in R).
- Considered best way to deal with singletons within a sample. Depends on replication across samples.
- What level of replication is OK? DSTL use triplicate

QC /Internal controls

- DSTL are comparing between platforms and methodologies with 2 types of control. 1. GC control 2. Process control spiked into sample. Ideally get same profile and stats. Currently evaluating to look at variability and error rate across systems.
- Need negative control - Considered contamination from PCR reagent or swab content - the kitome. A discussion to share what kits and swabs work best would be helpful. Try to change procedure. And/or remove bioinformatically.
- Separate processes out to reduce contamination
- Use 2 different methods to see if same results establish/verify data.
- Can we standardise the error by normalising the data across platforms? DNA extraction method is very important. Use mock community for standardisation.
- For meta-transcriptomics need to add internal control (Mary Ann Moran) to allow quantitative meta-transcriptomics

Metadata

- Need to record bioinformatics technology.
- It would be useful to be able to retro analyse data and compare different read lengths.
- EBI found that sequencing technology was not as important as method of preparation – needs recording accurately
- Explored if needed stats tests and comparisons. Depends on impact of getting it wrong, is it a pointer for directions to study or answer in itself.
- Instead of using p values can use effect size which should be more independent of methods.
- So much in database should we be using that first? Explore those with unknown function - use metatranscriptomics, proteomics, metabolomics data
- What is future? Oxford nanopore and many cells? Cf shotgun sequencing
- Often use big data and then focus on smaller questions with traditional methods – culture, metabolites, genes, expression
- For databases need cross-referencing and remembering that everything is in flux.
- Depends on what portal used what standards/metadata needed.
- What information should go into the supplementary materials, how much is needed and how much is too much!
- Journal of Biomolecular Detection and Quantification is a new journal that may be of interest.