Principles of Digital Biology Topics to Consider

Topic 1: Clinical Data Science

The range of data being used for clinical genomics is rapidly expanding, as is the list of stakeholders who want to access the results of the genomics medicine analysis. In addition to the clinical bioinformatics post, should the unit be investing in a specialist clinical data scientist position?

Questions:

What is the role of data science in genomics, is it different from the bioinformatics?

Would a clinical data scientist support the provision of cost-effective patient care?

Topic 2: Application of AI

The quantity of data available on human variants has massively expanded in the past year. It is expected that the amount data will increase even further (5 million genome project). The lab head has been to a number of talks which are highlighting the role that "deep learning" methods could play in interpreting genomics data in the context of disease. Should the lab begin to explore deep learning in its own work?

Questions:

What is deep learning?

Is it mature enough to deliver real benefit in a clinical setting?

Topic 3: Data Science and Data Quality:

The team is being asked to work with ever larger and more complex data sets. However, there are concerns about data quality.

Questions:

What tools are currently used within data science to provide a metric on data quality?

Are they useable for genomic data?

Is there any way to automatically generate data quality metrics.

Topic 4: Blockchain as a mechanism for sharing genomic data

There has been a lot of discussion as to whether blockchain – as a distributed ledger system – could be used as a safe way of sharing patient data between labs. Is this an area the lab should be exploring?

Notes: What exactly is the problem? What are the governance challenges of working with patient data

Questions:

What is blockchain?

Is the technology mature enough to be useful?

Is it safe?

Topic 5:Containerisation for reproducibility of work &/or easy distribution of genomic analyses pipelines

Containerisation allows researchers and professionals to easily share and reproduce computational workflows, which are often complex and require a lot of software dependencies. In addition, genomic data analysis is often complex, requiring specialist software tools to analyse the large amounts of genomics data. There have been recent advances have focussed on reproducibility and scalability of workflows with new platforms such as serverless computing. In the team they would like to know how recent advances and new technologies could help them run and share more complex analysis.

Questions:
What tools are available and how are they used?
What challenges are there when using it in a NHS environment?