There was a comment from one of the experts that you cannot have too much data. There must be some kind of cost limit on storing data (amount and for how long)? What is this limit? How much data is too much?

Answer from Dr Kristen Nowak:

There is not an exact answer to be given to these questions unfortunately – we don't know a limit on how much data is too much and how much money is too much to pay to store the data. This has not yet been discussed or decided upon at the national level.

• Currently, different public laboratories around Australia have different abilities to store genomic data.

- For example, here in Western Australia (at PathWest, our statewide public pathology provider), we have only started generating whole genome sequences this year. We don't yet have an appropriate way to store more than a small number of these genomes.
- So at the moment in WA, storing more than a few hundred genomes' worth of data would be technically impossible, let alone thinking of the cost and getting this funded.

• Storing large amounts of data will cost money, but:

- presumably this cost would go down over time.
- presumably there may be economies of scale with the cost organisations that provide cloud storage would charge (e.g. the more you store, the less it might cost per unit of data).
- the more ways you can use the data the more economical the cost of generating and storing it becomes.

• There are different ways you can store genomic data, based on if and how often you may wish to access it again:

- It is cheaper to store data that you think you may not need to look at again, or seldomly look at again (it then becomes more expensive to access it when you do need to though).
- It is more expensive to store data that you are likely to wish to access again (e.g. to reanalyse).

• There will come a point where it would be cheaper to generate new genomic data for specific people rather than store previously generated data for all who it is generated for.

- This would depend on how many people's genomic data is to be accessed again, and over what time frame. E.g. if only say 5% of newborns who have their genomes sequenced ever need to have their data looked at again, is it cost effective to store the data of all newborns indefinitely?
- The cost of generating genomic data is steadily coming down in price.
- The quality of genomic data is increasingly improving as technology improves.
- The time it takes to generate genomic data is also becoming faster.
- Therefore, what would the value of genomic data generated at the current time have in say 10 or 20 years' time, when the technology then is far superior, and a new test would likely be performed very rapidly?

• Ultimately, the value of the genomic data for a given timeframe would need to be approximated, and then the cost of doing this weighed up against generating new genomic data in the future for whoever may need it.

- Also, the value would need to weighed up against other things the money could be used for in health care.