

## **House of Lords Science and Technology Sub Committee on Genomic Medicine**

Further evidence, submitted after sub-committee meeting on 11<sup>th</sup> June 2008

### **Information on cost structures of 'next generation' DNA sequencing technologies**

#### **Introduction**

At the sub-committee meeting on 11<sup>th</sup> June, Lord Patel invited participants to submit further background information on the status and costs of 'next generation' sequencing technologies. Developments in these technologies are driving the acceleration in genomic research; for example 98% of all DNA bases ever measured have been sequenced since January 2008. These technologies also have potential for delivering improved health outcomes through clinical applications. The anticipated technology revolution in this area may be likened to a 'broadband revolution'.

This document presents an overview of the current capabilities and costs of existing technologies and presents a series of considerations that will be important when planning for the integration of future, emerging technologies into research and clinical strategies. Summary information is gathered from various public sources and references will be supplied on request.

#### **How to cost sequencing**

There has been much discussion of whole genome sequencing in the technical and scientific media. The cost of completing a whole human genome sequence is often used as a benchmark, resulting in the well publicised pursuit of the '\$1000 genome'. However, there is and will continue to be a need for affordable targeted sequencing of much smaller sections of DNA, and a need for the comparison of multiple samples.

A more relevant system for the cost of sequencing therefore divides the cost of sequencing into composite units, i.e. 'cost per DNA base' rather than the 'cost per completed human genome'. However, there is not a linear relationship between a single base cost and a full genome. There are fixed and variable costs and these will also depend on other factors including:

- The coverage requirement for the application e.g. bacterial genomes require 20-30x coverage just for resequencing, diploid genomes require 30-50x coverage
- Size of DNA segment to be sequenced
- Structural complexity of the segment e.g. repeats, polyploidy
- Raw vs finished data – varying error rates mean that re-reads are necessary to reach a satisfactory level of accuracy. The cost of 'finished' data rather than 'raw' data takes this into account.

When considering the real cost of sequencing, it is important to consider all costs: the capital cost of instrumentation, sample preparation, reagents, servicing, labour, energy, software, data storage and processing, and other overheads. There has been a recent trend of announcements that only discuss reagent costs, ignoring other costs and the infrastructure requirements needed to support these technologies. With all costs included, it is estimated that a completed human sequence will still cost several hundreds of thousands of pounds with current technologies.

An important part of this consideration is the cost and complexity of instrumentation. As current 'next generation' technologies cost hundreds of thousands of dollars and require multidisciplinary teams to operate, use of these systems is currently relatively centralised.

### Considerations when planning services and research

When planning for future research and clinical services based on sequencing, it will be important to consider the following:

- **Currently centralised:** The cost and workflow structure of existing sequencing systems suit a relatively centralised process that delivers large volumes of genomic data. This suits large-throughput projects, for example those centres currently participating in the 1000 genomes project.
- **Future technologies will be cheaper and simpler:** Technologies that are in development are likely to reduce the cost of instrumentation by a significant degree, and simplify workflows. Plans for future infrastructure and funding should imagine that future technologies will allow more laboratories direct access to affordable and powerful sequencing technology.
- The **rate of change** in these technologies is very fast. At the time of the 2003 White Paper, the funding structure for new technology assumed that it should be considered for replacement after five years. The existing technology pipeline indicates that a two year cycle would more appropriate for one technology to be replaced by the subsequent generation. Planning of the infrastructure and funding of genomic medicine would need to take this into account.

### Bioinformatics

A major challenge for all sequencing technologies is the downstream ability to process, store, organise and make sense of the data. Key parameters determine the utility of DNA sequence data. These are:

- **Read length.** If a system can read a longer length of DNA in one measurement, this makes reassembly of the data easier and can reduce the need for massive redundancy ('spare' data that has been sequenced but is not needed). This may be likened to having larger pieces of a jigsaw. Read length in particular determines the ability to assemble and align DNA data for genomics and genetics analysis.
- **Base call confidence.** The accuracy of each run of data analysis affects the number of times the sequencing needs to be repeated to produce high quality sequence data, and this in turn affects cost and time. Being able to work at "low coverage" also reduces the processor and storage overheads required to obtain and end result.
- **Cost.** See above and appendix

Any existing high throughput sequencer is likely to generate large amounts of primary (raw) and secondary (sequence) data, with relatively small amounts of tertiary data (mutation lists, genotypes etc). Systems that enable the end user to get to tertiary data quickly and cheaply are highly desirable and currently lacking

### Horizon scanning

There are a selection of new technologies in development that have the potential to deliver a significant improvement in the speed and cost of DNA sequencing. The speed of innovation means that a two-year cycle should be considered appropriate for replacement of sequencing technology.

Future generations of technology promise two specific improvements:

- **single molecule** analysis (removes the need for amplification of DNA in the sample, reducing complexity, time, cost)
- and/or **label-free** analysis (removes the need for fluorescent labels, which would simplify instrumentation and remove the cost and complexity of chemical reagents)
- **long read lengths** (minimise sample prep, speed up data assembly, greater accuracy, greater resolution of structural variation/polyploidy considerations)

A lower instrumentation cost and powerful throughput will in turn have a lower infrastructure need, allowing a broader range of laboratories direct access to powerful DNA sequencing. These factors should be taken into account when horizon scanning and planning.

One technology in development is at Oxford Nanopore Technologies, founded on the science of Professor Hagan Bayley of the University of Oxford. Nanopores are capable of direct single-molecule, label-free analysis of DNA bases and this technology is scalable. For more information please visit [www.nanoporetech.com](http://www.nanoporetech.com).

### **Applications**

Improvements in sequencing technologies have clear application for the research market. However, the NHS should consider how other health outcomes may be improved by an accessible, powerful, economic sequencing technology. For example, there may in the future be an economic argument for diagnosing infectious disease by sequencing the relevant infectious organism, or for near-patient sequencing to be a solution to a range of diagnosis needs.

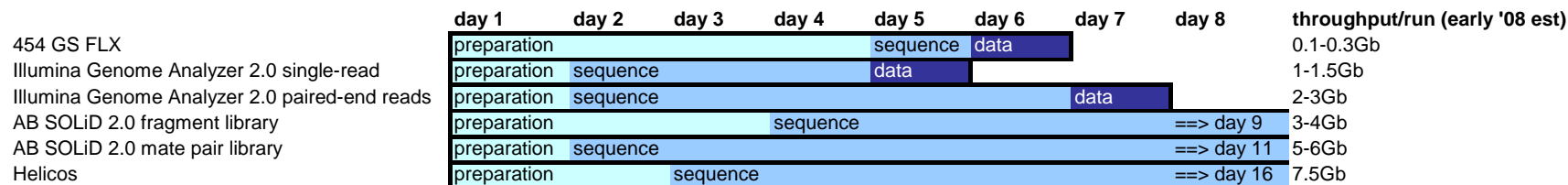
The Appendix on the next page shows a simplified representation of the workflow and cost structure associated with existing 'next generation' sequencing technologies. Specifications are taken from corporate literature and other public sources, however specifications may change due to ongoing development of these technologies.

### **References**

Tables were compiled from Corporate literature from Illumina, Applied Biosystems, Helicos, 454, as well as independent analyst reports: Pacific Growth Equities, 11 February 2008 and JP Morgan 9 October 2007. Sequencing specifications may vary according to the institution and application, and there may be further changes to these specifications in the future.

### APPENDIX: Workflow

Existing 'next-generation' sequencing systems include a relatively time-consuming series of steps for preparation of the sample, DNA analysis, and subsequent data reassembly and analysis. The graph below shows a simplified summary of these workflows. Because of the workflow, the nature of the support team needed to run the process and the volume of data output, these systems are well suited to large-scale projects. In some cases the systems suit/allow the parallel analysis of many samples rather than the analysis of a large volume of data from one genome.



Improvements in these technologies continue to increase the volume of data that can be generated in each run. Future technologies are also likely to deliver a shorter and simpler workflow with a lower skilled labour and management burden. These developments should be anticipated when planning sequencing technologies into future research and services.

### Overall cost comparison

The following table compares the existing 'next generation' technologies and outlines the cost requirements for running these technologies. With significant capital investment, these technologies are well suited to centres that can maximise their use, so are involved in projects with a high volume of genome analysis.

	sequencing instrument cost	annual service cost	Informatics costs per instrument (hardware, software, data storage, analysis, processing)	read length (base pairs)	sample prep (instrument+reagent. may need multiple instruments)	sequencing reagent costs per run	est. labour, management per run
454 GS FLX	\$500k + \$100k upgrade	\$50k	\$10k	200-300	\$25k+\$150/run	\$4k	\$150
Illumina Genome Analyzer	\$400k	\$50k	\$40k	36-50	\$25k+\$50/run	\$6k	\$1500
AB SOLiD	\$600k	\$50k	\$80k	25-35	\$25k+\$500/run	\$9k	\$3750
Heliscope	\$1.35m	\$50k	\$140k	30-50	\$25k+neg	\$18k	\$5250