University of Oxford:

Survey of DNA Sequencing Provision and Strategic Directions

Compiled by Rory Bowden, Wellcome Trust Centre for Human Genetics

Contents

1. Survey of DNA Sequencing Provision

Headline – the opinion of a key user

"…there is great deal to be gained by adopting the Broad[1] model which provides a high throughput, fast turnover, up to date sequencing facility at low cost to its scientists and this is reflected in the early high impact publication output – they really are on the front of the wave with lots of innovative uses of the technology coming in its wake. We should aim to have a similar philosophy here. Inevitably it means it has to be designed to run at lower than full capacity to give it flexibility and throughput but that is made up by the improved research output. At present we are maybe a couple of years behind them and will remain so."

Introduction

This report summarizes the provision of DNA sequencing capability within the University of Oxford. It has been compiled at the instigation of Oxford's Wellcome Trust Institutional Strategic Support Fund Committee in order to help in evolving "a more co-ordinated strategy for major equipment purchase, so that when major competitions arise, the University can make the best possible application in the light of up-to-date knowledge of existing capacity and a well co-ordinated view of upcoming need".

This summary of provision is not designed to be a comprehensive assessment of even all the currently available technologies; not every technology included or excluded can be explained in a single brief document. However, an effort has been made to mention each of the technologies that is currently relevant or is likely to have an impact over the next three years.

Sanger Sequencing vs. Next-Generation sequencing

The basic mechanism underlying Sanger dideoxynucleotide chain-termination sequencing of DNA is likely to be familiar to most readers. Invented in the early 1970s, automated in the 1980s with the introduction of fluorescently labelled dideoxynucleotides or oligonucleotides, and supercharged with capillary electrophoresis in the 1990s, the human genome was sequenced using Sanger technology, in which each of the thousands or millions of individual template had to be prepared as thousands of identical copies in a separate tube or well. In contrast, in the first 'next-generation sequencing' (NGS) platforms, both library making and sequence detection were streamlined such that whole genomes could begin to be sequenced in a single process.

How NGS works

The principles of next-generation sequencing have been described elsewhere. In brief, the term NGS refers to all post-Sanger, highly-parallelised, short-read technologies in which many (millions of) DNA molecules are simultaneously prepared (fragmented and joined to adapter DNA molecules) to form a library suitable for sequencing on one of several platforms. Limitations in the sensitivity of detection technologies have hitherto meant that

[1] The Broad Institute of Harvard and MIT: https://www.broadinstitute.org/

the library needs to be loaded onto the platform and then amplified into a sub-population or cluster of identical molecules to facilitate detection, e.g. by optical means. NGS technologies have included 454, SOLiD, Illumina's SBS[2] and Ion Torrent. The impact of upcoming (so-called 'Third-Generation') technologies, defined by their ability to directly detect sequencing signal from single molecules, is described in a later section.

Sequencing Provision Survey

Several departments or units for whom sequencing is a substantial activity were approached for responses on (a) their in-house provision of sequencing platforms and (b) their experiences and strategies regarding sequencing provision. The survey was augmented with an analysis of University-wide financial purchasing data for the 2014-2015 Financial Year.

Sequencing Platforms in Oxford

The provision of sequencing platforms (where and when sequencers have been purchased) in Oxford historically has been driven by a combination of general research area, Departmental strategy and the interests of individual research groups. Of course, success in a particular area of research (where that may include sequencing and sequencing data analysis) tends to lead to a strengthening of activity in that area, so the current disposition of facilities reflects, more than anything else, this history rather than a top-down initiative to provide research facilities or the results of a strategic assessment of Oxford's needs.

Sequencing Install-base

There are approximately 23 sequencing instruments in regular use within the University (including NIHR-BRC[3]). This sum includes a total of three ABI 3730 / 3730xl capillary sequencers (two in Zoology and one at the WIMM) still in active use, but none of the Roche/454 instruments nor one known Ion Torrent 'Ion Proton' instrument that have been installed in Oxford, which are not thought to be actively used. Early-access Oxford Nanopore Technologies instruments are also not included.

The dominant NGS technology is Illumina's SBS. Illumina divides its instruments into four classes with increasing output, a classification that is helpful in describing Oxford's combined sequencing capacity.

The **MiSeq** is a **bench-top instrument** capable of producing up to[4] 15Gb of data as paired 300b reads in a 2.5-day run, that is marketed for individual experimental labs, for focused clinical sequencing and for microbial genomics, to be used by non-specialists with operation facilitated by the use of cartridge-based consumables. Oxford currently has 8 MiSeq instruments in operation, consisting of two each at WTCHG-HTG, OTMD and in the MMM consortium, and one each at WIMM and Zoology. Originally marketed as a near-equivalent to the MiSeq, five Ion Torrent PGMs (Personal Genome Machines) are also in use in Oxford:

---

[2] Sequencing by synthesis.

[3] For Departmental abbreviations, see Appendix.

[4] Illumina's quoted outputs are maximum realistically attainable data volumes in favourable conditions, not guaranteed minimum amounts.

two at OTMD, two at WTCHG (not within WTCHG-HTG) and one at WIMM, for sequencing of targeted panels of amplicons.

The **NextSeq 500** is a **mid-to-high-output bench-top instrument** which emphasizes fast-turnarounds and ease-of-use comparable to the MiSeq, producing up to 120Gb of data as up-to-150b reads. It is targeted at smaller, standalone research and clinical facilities. There is a NextSeq 500 at both of NDORMS/TDI and WIMM.

The **HiSeq** range of instruments, in which the HiSeq 3000 (single-flow-cell) and HiSeq 4000 (two flow cells) have recently succeeded respectively the HiSeq 1500 and HiSeq 2500, is designed for **high-throughput operation** for a range of applications. The HiSeq 4000 can produce 900Gb of data per flow cell in 3.5 days in the form of paired 150b reads, and operational flexibility is maintained by dividing the flow-cell into 8 equally-sized lanes that can each run a different experiment. For the moment, the HiSeq 2500 maintains relevance as the only high-throughput machine that enter 'Rapid' mode to produce paired 250b reads for up to 150Gb of data per flow cell in a 2.5 day run. There are now three HiSeq 4000 instruments (two at WTCHG and one at OTMD) and four remaining HiSeq 2500 instruments (two at WTCHG-HTG and one at OTMD) in Oxford. The purchase of HiSeq 4000s at WTCHG-HTG during 2015 will trigger the retirement from service at the end of the year of three older (3-4 years old) HiSeq 2000 / 2500 platforms not included in the numbers above .

Illumina also offers the **Hi-Seq X ultra-high-throughput** instrument configuration, in which sets of 5 or 10 upgraded HiSeq instruments are bundled together. With proportionately higher purchase cost and per-run output but ~50% lower per-unit consumables costs than individual HiSeq 3000 / 4000 instruments, a HiSeq X Ten system can in principle sequence ~18,000 human genomes, to industry-standard 30-fold read coverage, per year. At a purchase cost of $US10M ($US6M for HiSeq X Five) and running costs of ~$US1000 per genome sequenced, the 4 system installations in the UK to date (WTSI, Glasgow, Edinburgh, Illumina's Genomics England Facility) have necessarily been tied to large-scale, funded projects such as Genomics England's 100,000 Genomes Project. At release, the HiSeq X system was initially commercially restricted for use solely on high-coverage human whole-genome sequencing although this restriction being progressively relaxed. In spite of substantial local interest in the sequencing and analysis of whole human genomes for clinical purposes, Oxford has not (yet) purchased an ultra-high-throughput sequencing platform. (The lack of funding for genome sequencing outside of Genomics England, the potential for national and global competition to supply what has become a generic commodity and the large investment required all contribute to uncertainty about the viability of a new ultra-high-throughput facility.)

Sequencing-capable Departments in Oxford: How Oxford's sequencers are used

In several Departments / Units in Oxford, sequencing capacity primarily reflects internal demand. For instance:

- Zoology has maintained automated Sanger sequencing capacity particularly in connection with microbial population genomics (especially multi-locus sequence typing) for almost two decades, also offering population-scale amplicon sequencing

to external users, and has recently invested in a MiSeq platform for mid-scale microbial and other sequencing based on Departmental demand.

- WIMM continues to run an internal quick-turnaround capillary sequencing service (for characterizing clones and constructs) and since 2013 has had a well-used MiSeq for internal use (self-operation by trained users). The WIMM use-case exemplifies the idea of bench-top sequencers as a fast way of getting a readout from any of a wide variety of experiments, enabling the next experiment to proceed without waiting a long time for off-site service. WIMM has just purchased a NextSeq 500 instrument for mid-scale projects, which will also be operated by researchers themselves.
- A consortium of NDORMS / TDI researchers has recently (late 2014) purchased a NextSeq 500 that is made available for use by University researchers.
- Oxford is home to an active translational clinical genomics programme under the NIHR-BRC banner, based in a dedicated clinical lab at the OTMD and at WTCHG and focused on clinical applications of whole-genome sequencing and focused cancer sequencing.

At WTCHG-HTG, Oxford possesses a *de facto* University-wide High-Throughput Genomics (arrays and NGS) core. WTCHG-HTG has since 2009 operated a Solexa/Illumina sequencing service, combining high-throughput instruments (HiSeq or equivalent) for whole-genome, exome and transcriptome / epigenome sequencing with, since 2011, MiSeq bench-top sequencing for test runs and microbial and focused genomic sequencing. Although not given any special institutional strategic support, The WTCHG has been designated an MRC Sequencing Hub (since 2010). In recent years, the WTCHG-HTG has operated under a self-assigned remit to provide the widest range of high-throughput genomics capability to Oxford (primarily) and the external scientific and commercial community. Funding is on a cost-recovery, fee-per-service model in which equipment, a substantial and recurring cost as new instruments are introduced, has been obtained from a combination of grants to collaborators and user charges. Recently, equipment replacement costs have been stripped out of charges to Oxford users. The funding basis for WTCHG-HTG is in contrast to comparable centres (the MRC Hubs, BBSRC's TGAC in Norwich) whose existence and management is driven by higher-level Institutional and/or Research Council strategy. WTCHG-HTG does not have specific clinical accreditation.

Sequencing Activity at the University of Oxford

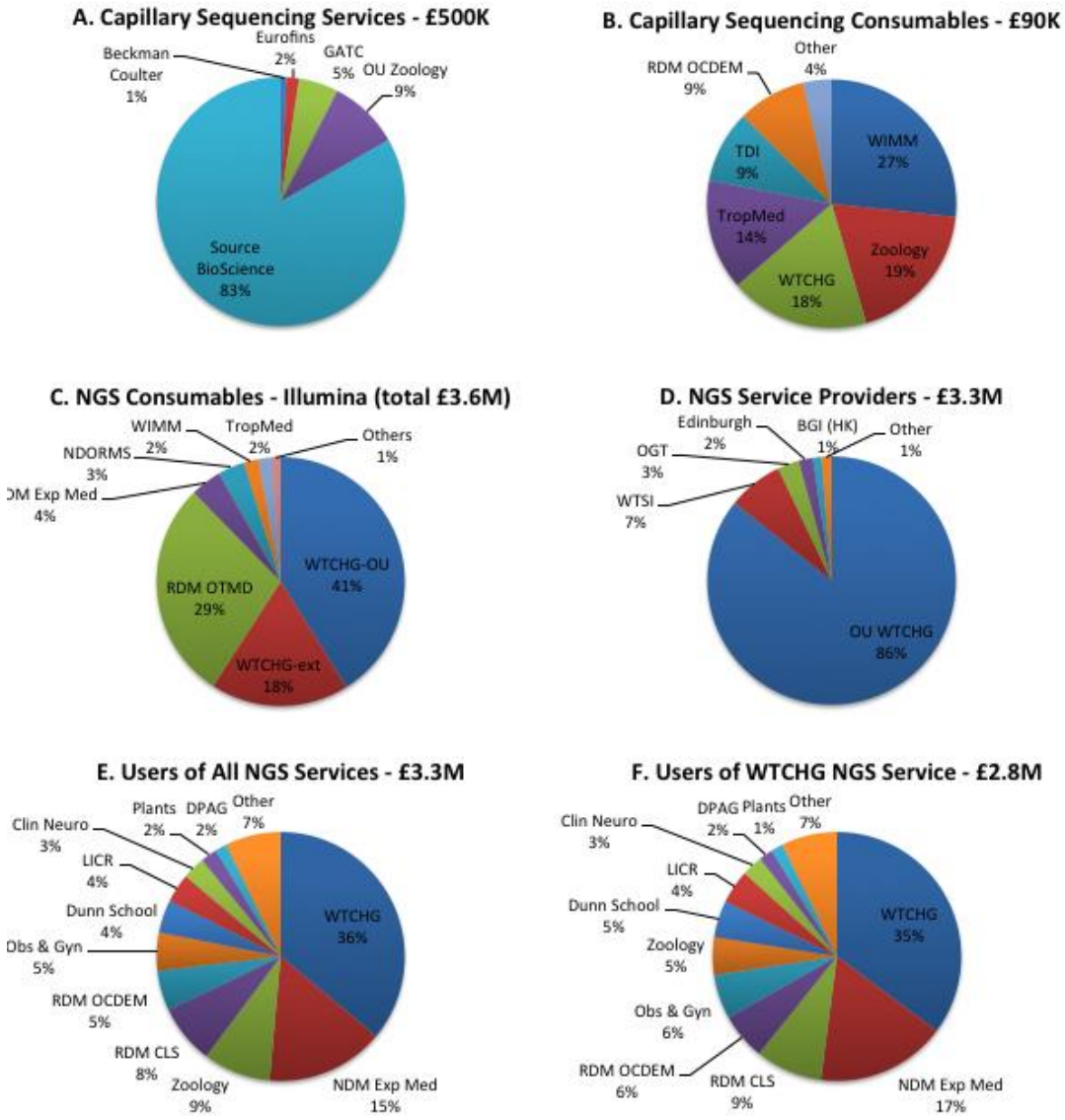
A snapshot of sequencing activity – the 2014-2015 Financial Year

Capillary sequencing activity

Although perhaps not a strategic priority, at ~£500K pa (excluding WIMM's internally facing service) the demand for Sanger capillary sequencing services (largely for single-locus sequencing and screening/checking of constructs, via overnight, pre-paid service) is surprisingly strong. The dominant supplier is Source Bioscience (83% of identified purchases), with four other suppliers including Zoology each accounting for <10% of the total activity (Figure). In addition to the above service activity, Oxford's own capillary sequencers required the purchase of ~£90K in capillary sequencing reagents, equivalent to an unknown extra volume of service sequencing. Given the expectation in this area for

highly flexible volumes and ultra-fast turnarounds, the potential benefits of centralised, University in-house provision seem marginal. On the other hand, Oxford might benefit significantly from a negotiated framework or University-wide tender that could foreseeably drive prices down.

Figure: Snapshot of Annual Oxford Sequencing Activity



Breakdown of spending on sequencing activities, financial year 2014-2015. A & B: Capillary (ABI 3730) sequencing services provided to the University (not including Zoology's and WIMM's internal provision) (A) and Oxford's spend on consumables (B). C: The great majority of Oxford's spend on Illumina consumables is for service activity by WTCHG-HTG (Oxford-to-Oxford: "WTCHG-OU" and to external bodies: "WTCHG-ext") or associated with NIHR-BRC / OTMD translational genetics research. D – F: WTCHG-HTG provides the majority of sequencing services bought by the University (D) and WTCHG is also the largest purchaser of services, both from all providers (E) and from itself (F).[5]

[5] Further Department / Unit abbreviations are in the Appendix

NGS activity

The total value of NGS activity undertaken by Oxford researchers is difficult to calculate because it is made up of a combination of NGS experiments undertaken or libraries made by researchers themselves (reflected in consumables spending), NGS services provided by internal (almost exclusively WTCHG-HTG) and external suppliers, and activities that cannot be accounted for because they do not involve spending within Oxford (large-scale collaborations involving direct funding of a major sequencing centre or grants held elsewhere). In any case, the vast majority of sequencing is done on Illumina platforms: as a guide, Oxford spend on Ion Torrent *consumables* is ~1/25, and on PacBio *services* ~1/100, that on Illumina equivalents. A reasonable estimate of total (Illumina) activity for the 2014-2015 financial year, excluding the categories of spending that cannot be tracked, would be ~£6.6M if all work were priced as a service, or ~£4.0M in estimated consumables costs for libraries and sequencing.

Scale of WTCHG-HTG activity

WTCHG-HTG ranks between about 5th and 10th in capacity / output among sequencing centres in the UK, and behind the ultra-high-throughput facilities at Illumina, The Sanger Institute and Edinburgh/Glasgow is one of about 5 broadly comparable core sequencing facilities (others are at Edinburgh, Liverpool and TGAC in Norwich). WTCHG-HTG provides the majority of Illumina sequencing *services* used by Oxford researchers while undertaking more than half of the NGS *by value* in Oxford (Figure – pie charts). The facility is run autonomously by WTCHG "on behalf of" the University, without external strategic or management influence, and has historically split its sequencing activity approximately equally between internal (ie WTCHG-affiliated), other Oxford and external projects. This mix of projects has enabled WTCHG-HTG to maintain scale to control costs and help with smoothing out periods of variable demand that could affect service delivery. The lack of a direct link to institutional strategy or financing appears to be unique to Oxford among UK centres.

Stakeholder Views

Summary of 7 responses (Units, Departments or Research Groupings), including representatives of almost all extant NGS capacity and the vast majority of demand, to part 2 of a survey of platforms and opinions about sequencing.

Blockages and Potential Issues in Sequencing in Oxford

Survey Question: *What do you see as the main issues that may (potentially) threaten the effective provision of sequencing in Oxford?*

WIMM: "Biggest bugbear from groups is length of wait for sequencing from third parties. Cost continues to be a challenge. Scale has been addressed by the different Illumina platforms - the choice is usually clear depending on the number of reads required. Short half-life of the technology is a disincentive to invest in local machines. Local expertise is rarely a problem as many in the WIMM are early adopters of the technology and are constantly looking for novel uses of the technology. However the bottleneck is always the analysis and most groups need a dedicated bioinformatician embedded in the group. What

is currently missing in Oxford is adequate bioinformatics training for Graduate students to equip them to analyse their data."

MMM: "Speed and quality of the data are currently of key importance but cost pressures are increasing…"

DPAG: "Most genomics researchers work in partnership with the WTCHG. They provide the samples handling and sequencing, initial QC and processing. Groups then analyse the DNA variants to identify disease associations. Price is always an issue, as is taking advantage of technological advances. We would like to use more next-gen sequencing of patients in the future and also carry out the sequencing and analysis in Oxford. For this. It will be vital to have in-house expertise available. In addition, the pricing should be competitive, otherwise we would be tempted to out-source the sequencing but would need assistance with the analysis."

RDM CLS: "Some issues that may threaten the effective provision of sequencing in Oxford might include adequate funding for next-generation sequencing provision (costs for setting up and running sequencing facilities, as well as funding for individual research groups to cover sequencing costs), speed (quick turnaround time), cost (competitive cost per sample) and analysis (availability of bioinformatics support)."

LICR: PI 1 – "Currently my group is happy with sequencing service provided by WTC. Speed, reliability, expertise and price are all good. We use it for gene expression analysis, genome-wide mapping of protein-DNA interactions and DNA modifications. We plan to expand a number of applications including but not limited to mapping of chromatin accessible regions and transcription initiation sites. Although speed of delivery is acceptable now, it could still be improved. Lower price is of interest as well. The implementation of web-based LIMS type system to track project samples and progress have been discussed more than a year ago and not have been implemented." PI 2 – "The main threats are in my view cost and analysis of data. Particularly the latter point is an issue as the cost per sample will continue to decrease. Due to the wealth of data generated, subsequent analysis requires specially trained staff in order to get maximum information out of the data. With an increasing demand on high-throughout sequencing, provisions for handling and analysing the increasing data load should be a priority." PI 3 – "Now the Ludwig has a dedicated technician funded, the turnaround time for sequencing at [WTCHG-HTG] is acceptable. Cost of RNA-seq remains high, especially if multiple runs are needed to obtain statistical significance. For some groups there remains a bottleneck in getting sequencing data analysed, though we are attempting to get around this by providing low level training to students and post-docs which will increase capacity. Nevertheless, bioinformatics support can be an issue."

NDORMS-TDI "Previously turn-around time and support for NGS activities through other facilities at Oxford have been unsatisfactory. Purchase of Nextseq 500 was a great choice to adapt in a highly flexible manner the need to get samples sequenced immediately, and it is now widely used through many external (i.e. non Botnar) users at University. Bottleneck is generally the lack of training possibilities (exception is CGAT) in computing at Oxford, either for students or for interested postdocs or senior researchers. "

There is a strong link between a consensus that a central sequencing facility for Oxford is desirable and recognition of the need for data processing and analysis capability; the expectation is that one should accompany the other, and be linked to training of analysts / provision of computational capability.

Potential Oxford-wide Strategies for Sequencing Provision

Survey Question: *Although not formally part of the information-gathering on existing provision, to not ask for opinions on whether an Oxford-wide strategy is desirable or achievable, and what it should include, would be a missed opportunity.*

WIMM: "The ease of use of the Miseq and Nexseq500 platforms has meant that these machines are very popular for in house sequencing provision. They allow rapid generation of data and for the groups to be nimble and adapt quickly to new developments and techniques such as ATAC-seq. I don't think central provision on this scale is necessary at least for the WIMM users. I anticipate that the WIMM machines will continue to be heavily used. For larger projects where time is less critical then the WTCHG is the ideal provider. A large central provider does have the ability to cope with the short half life of the platforms and provide cheaper sequencing at scale …"

MMM: "Having a central agreement with sequence equipment providers is essential. Having an Oxford wide strategy to sequencing would be beneficial but only if it maintained the creativity and innovations of the research groups. Care would be needed not to restrict access, application etc too tightly …"

DPAG: "I am happy using the services provided by the WTCHG. I would like it to be cheaper, and I would like new technological advances to be implemented sooner. Both of these are best served through a centralised service."

RDM CLS: "An Oxford-wide strategy for sequencing provision is clearly desirable (and is essential). It should provide quick and cost-effective next-generation sequencing together with the necessary bioinformatics support. The work of [NIHR-BRC Genomic Medicine and OTMD] is of great importance in this regard. It is important that Oxford is recognised as one of the worldwide leading universities in the field of next-generation sequencing in medicine. This technology will play an essential role in personalised medicine (diagnostic/prognostic markers and treatment selection) in the near future."

LICR: PI 1 – "We are currently happy with sequencing provisions. The WTC sequencing facility keeps-up with the technology and is efficient to serve our needs." PI 2 – "An Oxford-wide facility would be desirable and along with it should be a core to support analysis at different levels, including necessary training of staff. PI 3 – "Ideally Oxford should coordinate use of currently scattered bioinformatics resources including processing capacity. Not sure how this might be achieved given that different departments have invested independently in hardware and expertise. What is clearly needed is training capacity in bioinformatics (that we are partly addressing by using training courses at the WIMM) and perhaps an Oxford wide bioinformatics service."

NDORMS-TDI: “It is probably fair to say that provision, access and downstream training (computing, bioinformatics) in NGS related activities *for all users* at Oxford really falls short, and in comparison to other (especially US) Universities has to improve significantly if we want to stay competitive.”

Summary of Stakeholder Responses

Turnaround time

Almost all responders cited this as a potential or actual issue, including for outside suppliers. Approaches to successfully address the issue include in-house bench-top machines and specific negotiations or arrangements with WTCHG-HTG to define standards of service for a routine set of applications, including provision of dedicated staff acting between users and service. (Note that the cost of this solution is likely to be cheaper than instrument purchase.) Even when projects don’t run late, the projected time to data may be unacceptable, or the problem may be a lack of flexibility to run a same-day experiment, whose immediate results can allow the next experiment to proceed. Such flexibility is obviously much easier to attain in the absence of multiple competing demands.

Centralised Facility (WTCHG-HTG)

Responses supported the role of a centralised facility for Oxford, either for all or for the higher-volume parts of their sequencing needs. Most (Medical and Life Science) Departments continue to have relatively limited sequencing requirements (after the top 10 consumers of sequencing, the remaining 7% of activity in our snapshot was consumed by 11 further Departments) and so will probably continue to lack the expertise, let alone the demand, to sustain even bench-top instruments. Separately, anecdotal experience suggests that individual groups welcome the service they get from a default, centralised facility. In contrast to capillary sequencing, relatively few consumers of NGS within the University actively choose to get their sequencing elsewhere. Major users tended to support a strategic role for a University-supported facility, combined with the hope that strategic support might solve resource (including bioinformatics) and cost issues. One exception to the remarkable ‘loyalty’ shown to the WTCHG-HTG is in the area of large-scale human whole-genome sequencing for research, an area in which WTCHG-HTG’s ability to compete is limited.

Bioinformatics support and training

The lack of adequate bioinformatics support remains a consensus issue, expressed in several ways. From the point of view of a biology / medical researcher, it is reasonable to wish for more help and speed in providing interpretation of sequencing data. There is a lack of clarity about what level of support should come (free) with sequencing and a tendency to lump together as ‘bioinformatics’ several stages in the evaluation, preliminary processing, detailed analysis and modelling (not to mention experimental design issues) of experimental sequence data. Thus although the desire for straightforward support in data analysis is understandable, it is not surprising that there may be disagreement over how to provide for it. All agree on the need for more training of full-time or part-time bioinformaticians to support sequencing.

General Outlook

Platforms that have lost out to Illumina

A critical examination of the history and market for sequencing technologies tells us that intrinsic, patentable features of technology (rather than cost, marketing or other aspects) are essential to success. Unlike perhaps for FACS or even micro-arrays, when choosing a sequencer it is important to back a winner. Illumina has consistently occupied a position of market dominance for bench-top and larger-scale NGS instruments, based on intrinsic advantages of its SBS technology that enabled it to out-compete the pioneering 454 and almost all subsequent NGS market entrants.

Technologies that have eventually failed, for commercial and/or technical reasons, to make a lasting impact include the 454 pyrosequencing platform (the first successful next-generation platform, quickly overtaken on throughput and reliability, eventually equalled on read length), SOLiD (never installed in Oxford) and Ion Torrent's Proton system. Ion Torrent's PGM has to some extent democratized sequencing without attaining market dominance; its use, in Oxford at least, is limited to targeted sequencing of established multiplexes in a clinical research context. Ion Torrent's manufacturer Thermo Fisher has recently announced a new instrument, the Ion S5, marketed for targeted sequencing with a simplified workflow. Who would bet on its success?

Bench-top vs. High-Throughput Platforms

The market and application areas for next-generation sequencers have broadened as the top end of the market has achieved ever-higher capacity: today's bench-top instruments produce more data than the high-throughput machines of 5 years ago. Thus there are now separate niches for laboratory (bench-top), departmental (mid-range), university (high-throughput) and genome centre (ultra-high-throughput) platforms, and a case can be made for each in a University such as Oxford. The relative importance of unit sequencing cost and convenience is the single biggest question in determining whether to equip individual labs or units with their own platforms: instruments get more use in a central facility but the immediate availability of lab-based sequencing allows overnight return of data.

Ultra-High-Throughput Platforms

The Illumina HiSeq X platform has recently (mid-2015) been joined by a new platform, Revolocity, from a re-entry to the sequencing market, Complete Genomics, now owned by BGI. Complete Genomics uses the Combinatorial Probe Anchor Ligation technique for sequencing and the Revolocity is an end-to-end (raw sample-to sequence) integrated platform with comparable capacity and cost to the HighSeq X. Several platfoms have been placed around the world, including a first in the UK at the Epilepsy Foundation in Chalfont, Bucks and it appears to be a viable alternative and challenge to Illumina's dominant position.[6]

[6] The November 2015 announcement of a change in commercial strategy at BGI places a cloud over even these pre-installation, early adopters.

Third-Generation (Single-Molecule) Methods

At present, the vast bulk of all sequencing worldwide for research and clinical use is undertaken using 'short-read' technologies, or which Illumina is the dominant exponent. However, single-molecule, or third-generation technologies have practical and theoretical advantages, specifically the possibility to read the native molecule directly, producing long reads that make possible the *de novo* reconstruction of difficult, repetitive ('low-complexity') sequences or in principle the direct reading of non-canonical or modified DNA bases. The only current commercially established single-molecule platform is that of Pacific Biosciences, but nanopore-based sequencers are becoming available. '

Pacific Biosciences

PacBio's SMRT (Single Molecule Real Time) technology relies on the imaging, enabled by zero-mode waveguides, of the addition of individual fluorescent nucleotides to single DNA molecules by a slowed-down DNA polymerase. Released commercially in 2011, the RS (and then the RSII in 2013) have been large, expensive, relatively low throughput, relatively inaccurate instruments that have nonetheless found specialist uses, notably in streamlining the *de novo* sequencing of pathogen genomes, in resolving complex re-arrangements and in assisting with the assembly of novel larger (animal and plant) genomes.

With the availability of SMRT sequencing as a service from several specialist centres, no individual Oxford Department has been able to justify the purchase of the PacBio platform. Depending on the continuing development of nanopore sequencing, PacBio's release of the faster, neater Sequel SMRT platform in late 2015 might prompt a general re-evaluation of the technology for specialist applications requiring ~10kb sequence reads, but its commercial success is likely to depend strongly on progress of the intrinsically more straightforward, compact, nanopore sequencing technologies.

Oxford Nanopore Technologies

Launched to end users with an early-access programme in 2014, Oxford Nanopore Technologies' MinION is both the first nanopore sequencer and first truly portable, pocket-sized device, capable of sequencing a bacterial genome with median 8-10kb reads (and 100kb+ maximum read lengths) in a single run. With abbreviated library methods and by dispensing with amplification of individual library fragments, the MinION can be quick to get started and has been shown to be effective in identifying organisms (such as pathogens) quickly from metagenomic samples, in making near-complete bacterial genome assemblies and in resolving phase or structure in longer genomes. A scaled-up version of the same technology, the PromethION, with 48 independently addressed flowcells and eventually capable of 6Tb+ of raw data output per day, is due for early-access release in early 2016. With individual-read accuracies (optimized by reading both strands of each molecule) still little higher than 90%, there are questions about how much data will be required for e.g. human whole-genome re-sequencing, but the small size and ease of use of the instrument and the proposed business model, where users would not need to buy equipment and might pay a flat rate for a particular data volume, open the way for a major upheaval in the sequencing market once technical questions are answered.

The Core Service Perspective on Stakeholder Views

Cost of service

- Consumers of generic sequencing 'products', from microbial to human genomes, are aware of competitor costs in the market.
- Some researchers are sceptical of pricing under the current model, particularly when an instrument in their Unit would not be subject to full cost recovery for space and servicing.
- If bioinformatics support is offered alongside sequencing, its full costs are difficult to recover.
- Full cost recovery operations models for sequencing are not realistic and are sub-optimal for science.
- Elsewhere, facilities typically get support (for capital and other costs) from funders, philanthropists and institutional strategic funds.
- Charge-out costs for internal users at other institutions may cover just consumables costs (the biggest single cost for most sequencing projects) or may include some labour or facilities costs.
- Funders vary on allowable costs, but some will not permit capital cost recovery / depreciation and it is difficult to run multiple costing formulas so the lowest cost will normally win out.
- For these reasons, academic facilities struggle to break even. As a part of WTCHG within a single Oxford Department, Oxford's main sequencing facility is constrained by funders allowable charges and under pressure to subsidise non-WTCHG users with Centre resources, without a mechanism for cross-departmental or institutional support, a key advantage.
- Illumina has successfully stratified the sequencing market on size (preferred, larger customers) and application / instrument capacity (per-Gbase costs of consumables fall steeply as run data output increases.
- Oxford cannot currently compete on cost for whole-genome sequencing which is available for large sequencing centres on the ultra-high-throughput HiSeq X system at <50% of the unit cost of output from the otherwise state-of-the-art HiSeq 4000.
- We have known for some time that Oxford is not competitive on standard human WGS and Illumina has further released restrictions on species of origin for HiSeq X so there is a danger of being undercut even for non-human WGS.
- Apart from instances where a collaboration or funding arrangement has required the sequencing to be undertaken at another institution, large-scale sequencing projects have certainly been lost to Oxford because of cost issues. If a case can be made that doing more sequencing within Oxford would be strategically beneficial, then there would be an incentive to address issues of equipment provision and ancillary costs.

Central facility

- That researchers in Oxford should use, support and benefit from a central sequencing facility was a strongly supported consensus, with the obvious caveats

that standards of service and cost must not be unfavourable. This general support did not depend on whether a Unit had bench-top sequencing capability.

- Some researchers use other facilities (for justifiable reasons – either the realities of the market make Oxford sequencing unaffordable or the research is tied through funding or samples to another sequencing facility) but this is not a first choice.
- Genomics specialists still have a role, not least because they are expert at optimising data yield and quality, especially for non-standard sequencing applications, and can advise researchers who only use sequencing part-time. The flip-side of a strategy to ensure bench-top sequencing is available where it is needed ought not be a degradation of centralised facilities at for those who prefer them.
- Bench-top or on-demand sequencing can also be provided by central facilities: the Broad has a call-off, walk-up HiSeq 2500 Rapid facility for which the researcher orders a lane of sequencing (with only a few hours notice) and then simply provides a pre-normalised library for overnight processing.
- Turnaround time to useful data: Some Oxford users need assistance with all stages from sample to analysis while others are early adopters / experts on producing sequencing-ready libraries who still require bioinformatics expertise. So data may be available but useless because of a lack of analytical capacity, a frustrating situation for both service provider and user.
- The solution surely is a bioinformatics team that gets ahead of demand enough to develop automated tools and environments for analysis that can in turn maximise efficiency. There is a lack of trained personnel for analysis, a lack of (or unwillingness to assign) resources (especially across Departments) and a lack of recognition that sequence data analytics is a valid and valuable technical vocation in the same way that specialised molecular biology skills are indispensable for sequencing.
- High capital cost and short half-life of technology are dominant factors in the affordability of current next-generation sequencing platforms and strongly influence the purchasing and disposition of sequencers around the University. There is an impression that purchasing decisions are not made lightly.
- A further related and important factor is the risk of purchasing an ultimately unsuccessful technology in the face of limited budgets and pressure / incentives from manufacturers to place novel platforms. If the University is to commit strategic resources to sequencing platforms, then it would be logical to enforce an evaluation and oversight of technology choices, perhaps by internal and external expert advisors.
- University Purchasing also has a role in ensuring the appropriate service and performance guarantees are contractually mandated.

The benefits of on-site sequencing capability

- In principle, having sequencing capability (NGS but applies similarly to Sanger sequencing) on-site within the University can provide benefits in the areas of convenience, cost, turnaround, flexibility and expertise. Some of these benefits (convenience, turnaround) apply most when the facility is operated within the same Unit/Department or site, but others (latest technology, cost, expertise), since they assume economies of scale, are applicable at the level of the institution.

- It has taken some time for the arguments in favour of Departmental NGS provision to be won, but the availability of bench-top platforms, frustrations about turnaround times and gradually increasing demand for, and awareness of sequencing have pulled NGS into individual labs and Departments/Units.
- The main hypothetical objection to Department-level facilities is that they might not be fully or expertly utilized, so that sequencing may end up more expensive than in an institutional (or external) service. It seems that individual Departments see this as a risk worth taking, especially since, in comparison with a centralized fee-for-service facility, turnaround times remain in control of the researchers and labour may be provided by retained staff / students rather than by the service lab, resulting in lower apparent costs to the grant account.
- The arrival of cheaper instruments such as Oxford Nanopore platforms is likely to transform the landscape of Departmental sequencing, so long as sequencing assays can be 'ported' to the new formats.

The benefits of an Institutional shared sequencing facility

- There is little doubt that the ready availability of NGS within Oxford provides a substantial benefit to Oxford research. However, these benefits are not so much in the obvious areas of convenience or turnaround (where organizational and resource issues and their tension with cost can interfere with service delivery) but rather in the molecular biology and sequencing assay expertise that is directly available to researchers from other Departments (and other Universities) and the quality of data.
- The most common feedback from users of the WTCHG-HTG facility is in three areas, often combined: that turnaround can be very slow and there may be a lack of ready support for data evaluation and analysis, but data quality and volume are both very high.
- While the WTCHG-HTG lab is in many ways run on commercial lines, the lack of a profit motive and a strong desire for satisfied users means that the service can focus on quality and achieving optimal data volumes while supporting a wide variety of applications. The difference in these areas between WTCHG-HTG and other service providers is often cited as a positive aspect of user experience.
- Many of the arguments for an institutional facility like that at WTCHG can also be made for a potential ultra-high-throughput facility. Oxford's science and reputation would benefit, and a non-trivial by-product would be the cheapest available sequencing for even small numbers of genomic samples. Even if it remains difficult to start a WGS facility here, there could be strategic benefit in institutional support of cheap sequencing for all.

2. Strategic Directions

A University-Wide Strategy?

Why should we have a University-wide strategy on sequencing?

- The lack of a strategy and co-ordination for equipment provision could mean that Oxford researchers may be disadvantaged in the provision of sequencing equipment, either because the right bids for funding are not made, or because such bids are less likely to be successful in the absence of a clear plan.
- Thus having a strategy for sequencing equipment provision would make it easier to fund new equipment, both for Core and Departmental platforms.
- A strategy for sequencing could be helpful in gaining University support for sequencing-related initiatives.
- Co-ordinated initiatives for provision of analytical support, including through training programmes, would be more likely to be successfully conceived, funded and implemented.
- Having a strategy would allow the University to better come to grips with issues of inter-departmental funding and overall cost of sequencing.

What should be included in a University strategy for sequencing?

- General plans for provision and replacement of sequencing platforms, including aims for provision within Departments, taking into account demand and location
- A strategy for closely related technologies
- Awareness of new sequencing technologies
- A policy / strategy regarding ultra-high-throughput sequencing for Oxford
- Co-ordination of purchasing for external suppliers and service organizations
- A framework for more formal co-ordination between Departments, including the ability to set up cross-Department core facilities
- Acknowledgement of the importance of data processing and analysis in the successful exploitation of sequencing
- The ability to guard against costly purchasing mistakes by sharing expertise and risk
- Minimum standards for sharing of centrally or Institutionally funded equipment
- Strategies for support of Oxford Tropical Network researchers
- An exploration of ways to share financial risk of centralised provision across stakeholder Departments / the University

Substantive Strategy Suggestions

Bench-top to high-throughput provision

- Develop co-ordination between areas of strength in sequencing and data analysis in WTCHG, WIMM, CGAT (and possibly others) to develop a programme for training and support.

- Form a contact group with representation from a senior academic level, Purchasing, stakeholders and existing strategy groups at WTCHG and WIMM that can advise the University on priorities.[7]
- Recognize the WTCHG-HTG group as the University's default core sequencing service, eligible for high-level institutional strategic support to maintain its capabilities and improve Oxford's access to value and expertise.
- Adopt a strategy for making central funding applications for sequencing capacity.
- Funding applications and purchasing decisions that involve University or Departmental Institutional support should be evaluated for technical feasibility and value for money and approved by a small group of Oxford / external experts
- A statement should be produced to clarify the sequencing and related facilities available to researchers in different parts of the University.
- The availability and need for shared bench-top NGS capability in under-served (geographic) parts of the University should be evaluated, taking into account the potential impact of upcoming technologies (Nanopore).
- Support the development of 'walk-up' sequencing facilities[8] accessible to the widest group of Oxford researchers, in an efficient and cost-effective manner. [The WTCHG-HTG facility may be in a position to offer more walk-up services.]
- Maintain an up-to-date view on possible provision of long-read technologies for the whole University

Ultra-high-throughput provision

The future development for Oxford of ultra-high-throughput (population-level) sequencing capacity is strategically desirable in order to strengthen Oxford's existing translational genomics research and also for specific-disease-area research. Even though sequencing can be done at a distance from patients and analytical expertise, it is likely that research strength will continue to accumulate around sequencing centres. Therefore the University should:

- (take appropriate steps to …) Prioritise the formation of an ultra-high-throughput facility for Oxford.
- Investigate diverse and innovative options to bring ultra-high-throughput sequencing to Oxford, including disease-specific, charitable, philanthropic and industry collaborations and alternative (non-Illumina) platform(s).

The existing WTCHG-HTG group and the OTMD group are both in a good position to contribute to the formation and running of an Ultra-high-throughput facility. In any planning, the progress of other (especially nanopore) approaches will need to be carefully monitored, although it seems likely that feasible alternatives for whole-genome sequencing are still at least two years away.

---

[7] A Genomics Oversight Committee, with representation from across the University, has recently been convened and may be in a position to provide input in this and other related areas.

[8] Generally understood to involve reserving capacity for last-minute bookings where the user either provides ready-to-sequence libraries to the sequencing team, or runs the machine themselves.

3. Appendix

Acknowledgements

Thanks are due to designated representatives of stakeholder Departments who responded to the initial survey request and commented on a draft of this report, and to staff in the University Purchasing team who helpfully provided bulk data on relevant categories of purchasing.

Abbreviations of Departments/Units

NIHR-BRC	NIHR Oxford Biomedical Research Centre
WTCHG	Wellcome Trust Centre for Human Genetics
WTCHG-HTG	WTCHG-High-Throughput Genomics
OTMD	Oxford Translational Medical Genomics Centre
NDM	Nuffield Department of Medicine
MMM	Modernising Medical Microbiology (NDM Experimental Medicine)
WIMM	Weatherall Institute for Molecular Medicine
Zoology	Department of Zoology
NDORMS	Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
TDI	Target Discovery Institute
RDM	Radcliffe Department of Medicine
RDM OCDEM	RDM Oxford Centre for Diabetes, Endocrinology and Metabolism
TropMed	NDM Centre for Tropical Medicine and Global Health
DPAG	Department of Physiology, Anatomy and Genetics
Plants	Department of Plant Sciences
Clin Neuro	Nuffield Department of Clinical Neurosciences
LICR	Ludwig Institute for Cancer Research (Oxford)
Dunn School	Sir William Dunn School of Pathology
Obs & Gyn	Nuffield Department of Obstetrics & Gynaecology
RDM CLS	RDM Nuffield Division of Clinical Laboratory Sciences
NDM Exp Med	NDM Division of Experimental Medicine
CGAT	Computational Genomics Analysis and Training