A centre for high resolution imaging

Background

This document for discussion follows on from the recent imaging community consultation for the ISC, and concentrates on optical imaging, which was addressed by the Life Sciences Panel. The panel discussion encompassed three main frontiers of optical imaging could potentially be addressed by the ISC: high-throughput imaging, *in vivo* animal imaging, and super-resolved/singe molecule imaging. Additionally, the development of software and facilities for image processing and long-term storage was universally considered to be an important need and a potential role for the ISC. Ideally these "frontiers" would be considered in the context of scalable experimental designs aiming to explore the relationships between molecular dynamics and functional dynamics. This would require molecular, cellular, and intravital imaging strategies, utilising with prior knowledge with the software resources to access, retrieve, analyse and synthesize data on all levels. It is clear, however, that the resources available to the ISC will be insufficient to adequately address all these areas. *In vivo* imaging is being discussed in parallel by a group within the MRC Mouse Genetics Centre at Harwell and a proposal for super-resolution/single molecule imaging is being developed by David Clarke and colleagues at Harwell. There is a consensus that high throughput imaging would be the most expensive and challenging to address and that it would not be appropriate to locate such a facility at Harwell. This document is intended to explore the potential for a national centre for high throughput imaging that could be located elsewhere.

High throughput imaging is perhaps most associated with assays for drug discovery and is related to high throughput genomics approaches to biology, including the screening of siRNA libraries with the associated challenges of curating and analysing large data sets. Fluorescence is widely but not exclusively used to readout such assays and the biological samples are typically arrayed in multiwell plates. The form of such assays ranges in complexity from single channel measurements made on homogeneous solutions, through low resolution imaging to "count" multiple cells in each well (analogous to cytometry) to imaging fixed or live samples with subcellular resolution and ultimately to imaging whole organisms such as zebrafish. The use of imaging to read out assays is often referred to as High Content Analysis (HCA) and the extension to time lapse imaging of live cells to establish phenotypes represents the current state of the art, as exemplified by the "*Mitocheck*" assay developed at EMBL by Jan Ellenberg and Rainer Pepperkok.

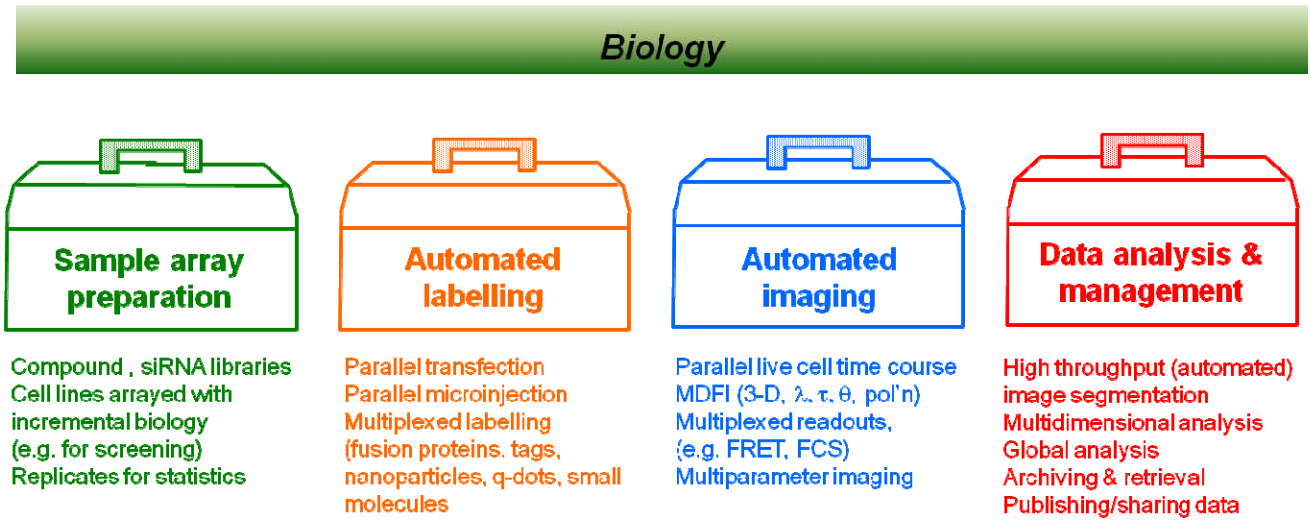

Figure 1. key components of high throughout imaging with possible implementation methodologies

Figure 1 aims to highlight the key competencies required to deliver state-of-the-art high throughput imaging assays. The power of the approach is to accelerate biological research by parallelising it and this requires appropriate sample preparation and labelling technologies for the biology under investigation. For time lapse imaging of live cells labelled with genetically expressed fluorophores, this is particularly challenging – even before considering the preparation of the sample array from e.g. an siRNA library. One approach to labelling arrays of live cells with GFP and other fusion proteins is the technique of Reverse Transfection developed by Sabatini and pioneered in the UK by Mike White's group. This involves "printing" an array of transfection reagents using a spotter and culturing cells on this array for subsequent time lapse imaging in a microscope or multiwell plate reader incorporating an incubator. To date such live cell arrays have been readout using fluorescence intensity in wide-field or confocal microscopes but there are now many sophisticated "multidimensional fluorescence imaging" techniques that have been developed for microscopy, including fluorescence lifetime imaging (FLIM) for Förster resonant energy transfer as a read-out of protein interactions or conformational changes. Resolving fluorescence signals with respect to multiple spectral and spatial dimensions leads to Gb data files per image, particularly when multiple dimensions are combined, and so time-lapse imaging of multiwell plate arrays rapidly exceeds Tb data volumes. This presents a major challenge in terms of analysing, storing and sharing data that requires supercomputing resources. To progress beyond intensity imaging and to combine image segmentation with spectral discrimination appears highly desirable but has not yet been realised anywhere. For high throughput screening it is therefore necessary to match supercomputing resources with new software tools for analysing image data on an unprecedented scale.

Demand for a national high throughput imaging facility

The ability to screen e.g. protein interactions or conformational changes as part of a cell signalling network against an siRNA library is becoming increasingly desirable and is likely to become of increasing strategic importance. We estimate that life scientists at Imperial College London are currently considering 10 such screens and ~50 such screens would be a reasonable estimate of demand once the technology was proven. We extrapolate this to a likely demand of ~ 200 screens from life scientists across Greater London and at least 400 across the UK. This demand would far exceed the capacity of EMBL or any other current facility to deliver it and suggests that a national facility would find more than sufficient users. It also highlights the importance of developing new, scalable high throughput imaging technology. While individual institutions can make important contributions towards this goal, no single institution offers competence in all the necessary component areas and this also speaks to the need for a co-ordinated national effort.

Strategic imperative

The *Mitocheck* assay represents a multimillion pound investment at one of the world's leading molecular biology laboratories. It is unlikely that any UK institution could replicate it at this time and it provides an exemplar of what might be achievable in a national centre for high throughput imaging. The capabilities to undertake such an assay have been established elsewhere in Europe, e.g. in Germany and Switzerland, and there is a strategic risk that such capability may be out of reach of UK scientist, except through international collaborations as a junior partner. It is therefore essential that UK scientists have access to this type of facility if the UK is to maintain its world-leading position in the life sciences. It is also vital to have access to the data from such experiments – the *Mitocheck* data is not yet freely available. As well as the impact on discovery in biology and medicine, it is also important that the UK build this expertise in the light of its importance to the pharmaceutical industry.

The need for scalable high content analysis (HCA) is ubiquitous. For drug discovery, it is important to understand that high content (HC) screens to date have only found a fraction of potential new leads. Even *Mitocheck* did not perhaps meet its full potential although it was able to conduct a sophisticated HC siRNA screen on an unprecedented scale. Conventional HC screens are far more time consuming and the issues of scale are highly constraining. For example, Dundee is just finishing a primary screen for chromosome proteins that load at specific points in mitosis. The proteomic list amounts to a few hundred proteins. Having conducted a secondary screen on worm mutants and culled this list to 60, there are still too many proteins to work on with currently available technology and so the next screen is siRNA knockdown in HeLa, followed by high resolution time lapse imaging of GFP-histone H2b. Having accessed their internal human genome siRNA library, they are, night by night, doing time lapse imaging of all sixty proteins, which is estimated to take ~4 weeks, doing 5 at a time, plus controls. This could, in principle, be addressed with a single run on a Mitocheck-style HCA system using reverse transfection.

A further key role of a national facility would be to act as a *national resource for siRNA libraries*. This could offer very significant cost savings compared to multiple UK universities buying the same such libraries. A national centre would therefore widen access to such resources and broaden the range of scientific questions that a single academic institution could address.

Today's proteomics methods, e.g. IP, Chip-Seq, and others, regularly develop lists of 50-200 gene products for functional follow-up and this suggests that there is significant demand for defined functional screens as a service to the UK academic research community. This functional validation (secondary/tertiary screening) of what are essentially primary screens would complement full genome -wide approaches. Currently there is nowhere in the UK to meet such demands and even EMBL's Mitocheck only offers a partial solution. It appears to be highly desirable that the international community, including the UK, develop new and better high throughput imaging methodologies, including siRNA design and synthesis, transfection, imaging, and data analysis. The latter software issues resonate with the emergence of open source tools such as OMERO for data archiving, sharing and analysis that are also set to make a significant impact in the wider imaging community, as discussed below.

Software development for Imaging

It is almost universally acknowledged that the exploitation of advanced imaging methods is paced by the availability of suitable software tools. Indeed, data analysis and management tools are going to be ever more important to address the big questions concerning structure, function, and ontology, hopefully allowing the broad gaps between the genomics and physiology to be filled and enabling fundamental biology to be studied, described and understood. Software development for image analysis, data management and system biology is, of course, extensive, but this has generally taken place on an *ad hoc* basis resulting in a vast number of non-compatible developments and a complete lack of universal formats for image acquisition and storage. It is reasonable to compare the current state of imaging software with the early stages of software development for protein crystallography, where groups developed their own software independently, to suit the needs of their specific science programmes. This was revolutionised with the establishment of the CCP4 project, which developed tools for crystallography data analysis, and established universal data formats.

With the establishment of the ISC, the opportunity exists to put in place for imaging a similar structure for the development of universal software solutions for data capture, analysis, and storage. As a provider of central facilities, STFC has a vast pool of expertise that can be exploited in this area, so the ISC is the logical choice in which to locate the core of any imaging software developments. In the same way as CCP4, a core staff with access to high-power computing facilities would be located on the Harwell campus, and nodes of the software development project could be located at universities to ensure close collaboration with the user community. In particular, the opportunity exists for the ISC to collaborate with groups developing the Open Microscopy Environment (http://openmicroscopy.org). This project aims to

develop standard file formats for microscopy and has developed an advanced image management platform (OMERO). The UK's OME development team is based at Dundee and headed by Prof. Jason Swedlow. Prof. Swedlow has indicated that his group would be willing to form a partnership with the ISC to promote this area. A national high throughput imaging centre could work in partnership with the suggested ISC imaging software core at Harwell, transferring capabilities and working in partnership to develop agendas and priorities.

Data archiving and storage is key area which is becoming increasingly important and the establishment of a centre in which imaging data can be stored for the longer term could be provided by STFC and the ISC, e.g. working with the ATLAS centre at Harwell, which was established to handle and store very large volumes of data resulting from high energy physics experiments, and which has spare capacity (and expertise) that could be used for the establishment of a long-term image storage facility.

Outline capital costing

This proposal for a high throughput imaging node of the ISC comprises 4 equipment elements, with which appropriate personnel would have to be associated.

- Robotic sample array preparation and handling
- Equipment for labelling sample arrays
- Automated imaging the sample arrays
- Hardware and software for image acquisition, analysis and curation

Robotic sample array sample preparation and handling: Liquid handling system, compound libraries, automatic pipetting, multiwell plate management *(other equipment?)*
Estimated £1.5M. (£1.3M hardware and £200k staff costs to interface/commission the equipment)

Equipment for labelling sample arrays: (Equipment details, e.g. reverse transfection equipment, FACS etc *(other equipment?)*
Estimated £0.7M (£1Mhardware and £200k staff and consumables costs to interface/commission the equipment).

Equipment for imaging sample arrays: (Equipment details, e.g. multiwell plate readers & microscopes etc.
Estimated £2.5M (£2M hardware for 4 workstations and £400k staff costs to interface/commission the equipment)

Hardware and software for image acquisition, analysis and curation: Acquisition software that integrates microscopes with robotics; data acquisition and data analysis in the different instruments, archiving and sharing. (anything else?)
Estimated £1.2M, including £1M for a super computer cluster and data storage systems and £200k for commissioning the IT infrastructure).

In addition, a significant investment (~£500k) is required for biological laboratories at the facility. It should be stressed that the approximate costings below do not include laboratory space or infrastructure.

Operational Issues

- ***Capital cost***

Subsystems	Capital cost (£)
Liquid handling and pipetting system	500k
Robotic plate handing system	300k
Compound library	500k
Multiwell plate readers	2M
Array labelling technology	1M
Data management	1M
Biology laboratories	500k
Commissioning staff effort	1M
Total	**6.8M**

- *R**ecurrent annual costs (non staff)***

Subsystems	Recurrent cost (£)
Liquid handling system	20k
Multiwell plate readers	60k
Robotic plate handing system	20k
Compound library	100k
Data management	50k
Total	**250k**

- ***Recurrent annual staff support required***

Staff	Persons
Assay design and implementation (biological)	2
Assay design and implementation (technical)	2
Dedicated operators for automated/robotics equipment	2
Compound/sample curation and preparation	1
Data analysis and curation	2

Location and space requirement for equipment (and any special requirements)

A national high throughput imaging facility capable of running four independent screens and including the compound libraries etc might require ~500 m^2, depending on the space allocated for biology labs and excluding the space required for the IT resources. Such a facility would require appropriate incubators, low temperature refrigeration, fume cupboards and at least Class 2 biological safety cabinets.

It is considered that a national high throughput imaging centre would be best located in easy reach of the life scientists wishing to use the facility. Given the concentration of university and research institute laboratories in London, an obvious suggestion would be to consider locating it in the new St Pancras research institute. A second obvious suggestion is to locate the centre near existing clusters of expertise such as Hinxton, Cambridge, in the proximity of the EMI. Clearly London is desirable in term so of easy access from almost anywhere in the UK or abroad but any associated building costs could be high.

The space required for visiting users (both short term & long term)

Visiting users will require access to laboratories for sample preparation and characterisation (see above). Office space will also be necessary, particularly for long term visitors.