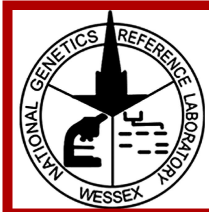


National Genetics Reference Laboratory - Wessex

Specification

The SCOBEC High Throughput Screening Facility

Revision 02
February 2004

Contributors

Person	Role	Institution
C. Mattocks	author	NGRL (Wessex)

Review and Approval

Person	Role	Institution

Distribution

To	Date

Revision History

Issue	Author	Date	Description
HTSF_01	C. Mattocks	12/04	1st draft
HTSF_02	C. Mattocks	02/05	Sample handling detail added Options clarified Polymorphism screen added (section 8.2)

Contents

List of Abbreviations **4**
1. Introduction **5**
  1.1. Aims *5*
  1.2. Operational Outline *6*
2. Sample acquisition **7**
3. Sample login **7**
4. Pre-PCR DNA quantification and quality checks **8**
5. Sample storage **9**
6. Batching **9**
  6.1. Sample collection *9*
  6.2. Primary screen batching *10*
  6.3. Secondary screen batching *13*
7. PCR setup **13**
  7.1. PCR reagents *13*
  7.2. Primary screen PCR setup *14*
  7.3. Secondary screen PCR setup *15*
8. Post PCR processing **16**
  8.1. Primary screen *16*
  8.2. Polymorphism screen *167*
  8.3. Secondary screen *18*
9. Capillary electrophoresis **19**
10. Sample and reagent tracking **20**
11. Analysis **20**
  11.1. Primary screen *20*
  11.2. Failures *21*
  11.3. Secondary screen *22*
12. Reporting **22**
13. System control **23**
Appendix A: Standardised primer design and optimisation (SPODS) **24**
Appendix B: Outline of proposed usage of the HTSF **25**
Appendix C: Simplified process diagram for the HTSF **26**

List of Abbreviations

1°	Primary screen
2°	Secondary screen
BRCA 1	Breast cancer 1 gene
BRCA 2	Breast cancer 2 gene
CSCE	Conformation sensitive capillary electrophoresis
cv	Coefficient of variance
EXO 1	Exonuclease 1
FAP	Familial adenomatous polyposis
GS	Gene specific
HNPCC	Hereditary nonpolyposis colorectal cancer
HTSF	High throughput screening facility
IT	Information technology
LIMS	Laboratory information management system
MALDI-TOF	Matrix assisted LASER desorption ionisation – time of flight
NF1	Neurofibromatosis type 1
NGRL (W)	National Genetics Reference Laboratory (Wessex)
NPfIT	National programme for IT
PCR	Polymerase chain reaction
QA	Quality assurance
rfu	Relative fluorescence units
SAP	Shrimp alkaline phosphatase
SCOBEC	Salisbury, Cambridge, Oxford, Bristol, Exeter, Cardiff
SPODS	Standardised primer optimisation and design specification
WRGL	Wessex regional genetics laboratory

1. Introduction

The SCOBEC high throughput screening facility is being developed to address the eight week reporting times for large genes outlined in the Genetics White Paper (June 2003). The facility will service the six constituent labs of the SCOBEC consortium and will initially focus on implementing a full screen for the hereditary breast cancer genes BRCA1 and BRCA2, covering all the coding exons. All aspects of the system have been designed to allow the transfer of the technology to other large gene disorders or multi-factorial disorders that require comprehensive screening.

The system will be installed in a new laboratory, currently under construction and due to be completed in April 2005. NGRL (Wessex) are working closely with SCOBEC to facilitate design and implementation of the systems required. Once the system is operational control will be transferred to WRGL to provide the ongoing service. It is anticipated that the facility will be staffed with one senior scientist and two technicians. The experience gained as a result of this project will be important resource for other diagnostic laboratories or consortia wishing to implement a similar system. This document represents a preliminary dissemination of this experience.

1.1. Aims

Aside from meeting the White paper guidelines, the structure and screening methodology for the HTSF has been designed with a number of aims in mind. These aims are outlined below:

- To maximise efficient use of instrumentation.
- To minimise sample turn around time.
- To maximise efficient use of scientific expertise.
- To redistribute work load to appropriate level.
- To maintain skill mix (technical and theoretical) at local level.
- To improve equality of service across disorders.
- To free time to allow development of local research and development links (e.g. with universities).
- To improve QA by standardisation of process.
- To provide the service without compromising future IT requirements (i.e. NPfIT)

1.2. Operational Outline

Initially the HTSF will undertake sample screening only; predictives / pre-symptomatic testing will not be undertaken. Ultimately however, this type of work will be managable within the secondary screening scheme. For tests that are carried out by the HTSF primers and protocols will be available for other labs to use on request. The process design has been built around a standardised primer design and optimization specification (SPODS). This is essentially a four primer system that uses a standard thermocycling profile for all templates and incorporates universal tails (M13) onto each end of the PCR product (Refer to appendix A for further details of the SPODS). This use of universal tails allows any fragment to be labeled with a wide range of fluorescent dyes or other useful tags. The system also affords great flexibility allowing a wide range of detection methodologies to be used with the same primer designs.

The HTSF will be following a pre-screening strategy using conformation sensitive capillary electrophoresis (CSCE) as the primary screen. All fragments found to contain a variation in the primary screen will be confirmed and characterised using bi-directional sequencing (secondary screen). It is envisaged that the analysis of the primary and secondary screens will be performed on different instruments (ABI 3730/3130 respectively). Each instrument will be dedicated to its purpose and overseen by a single operator.

In the event that CSCE is found to be unsatisfactory in any way, a full sequencing strategy will be adopted. This sequencing strategy is anticipated to cost in the order of five times more than the CSCE approach. We have also investigated a MALDI-TOF mass spectrometry approach (MassCleave™) but, although it appears to be a very promising approach for the future, it was not considered to be fully operational in the short term.

To limit the amount of data transfer between the referring labs and the HTSF, samples will be processed with reference to a sample number only; no patient information or clinical details will be exchanged. Samples will be processed and all sequence variations characterised at the DNA level and the protein level where appropriate. Reports will be returned to the referring lab by electronic transfer with results associated to the sample reference number. The reports will be in a text format suitable for inclusion in the final clinical report for the referring clinician and will deal with the patient in isolation with no reference to any other family members. Each centre will retain clinical control and management of the testing on local referrals. Since predictive / pre-symptomatic testing will not be carried out by the HTSF other arrangements will need be made where this is required. Refer to appendix B for an outline proposal.

2. Sample acquisition

Primary labs will use pre-bar-coded tubes for sample transportation. These will be 'Matrix' (TrakMate) style tubes with a pre-assigned 2D barcode printed on the base of the tube. It is anticipated that 0.75ml tube will be suitable for the purpose since they allow sufficient volume for the purposes of HT mutation scanning but are shallow enough to allow simple sample removal without the need for specialised pipette tips. Tubes will be sealed using SepraSeal caps; we have been assured that these provide a suitable seal for postage but trials will be carried out to determine the most suitable posting format.

TrakMate tubes will be purchased by the HTSF and supplied pre-racked in standard 96 well plate format to the regional labs. Each tube rack will identified using a linear barcode (font to be specified: either code 39 or code 128). This coding will be used for sample storage in the HTSF only. Initially the HTSF will purchase a stock of racks that can be filled with samples as they arrive. However, the regional labs should forward empty racks to the HTSF as they become available and this will provide the ongoing supply of storage racks.

During sample preparation in the originating lab the 2D tube bar-code will be read and this number stored against the relevant sample in the local database. The referring lab will log a new test in the HTSF database by defining tube [sample] number, primary lab reference and the test required to be carried out on the sample. Ideally this will be achieved by direct access to the HTSF database but other less direct methods will also need to be considered (e.g. email). Samples will be transported to the HTSF in the allocated tube; no supporting paperwork will be required.

3. Sample log-in

On receipt at the HTSF the sample will be logged as received by scanning the bar-code on the received tube. At this point the LIMS will allocate the next available storage position on the current rack. If a new rack is required the LIMS will request that a new rack ID is scanned and this will become the current rack. Sample racks will also be manually labelled using a simple system (e.g. year_plate no.) to facilitate easy identification and selection from storage. This number will also be stored against the relevant samples.

Sample log-in will activate the sample in the outstanding tests list and the disease specific batch collection list will be updated (see sample batching). Where a test has been defined on the database by an originating lab but no sample is received within a given period of time (e.g. one week) this will be flagged and a communication sent to the referring lab by email.

We will be investigating the use of 'liquid wax' (e.g. Chill-Out from MJ Research) as a tube sealing option for all stages of sample processing within the HTSF. This is

liquid at temperatures above 10°C and solid below 10°C. It is anticipated that a small amount of liquid wax will be laid on top of each sample as part of the log-in procedure. For the purpose of this document it has been assumed that trials of this method will be successful. Thus all references to liquid wax in this document are subject to trials once the robotics is operational. The main alternative to this sealing option is manual capping/de-capping of tubes and sealing of plates using adhesive foils, both of which present considerable logistical problems.

A variety of quantification and quality tests will be performed as part of the sample log-in procedure (see section 4 below).

4. Pre-PCR DNA quantification and quality checks

At least some level of standardisation of DNA concentration should be carried out in the primary lab. This will allow the required quantity of DNA to be sent to the HTSF at a dilution that will be at a suitable concentration for storage purposes and for a working dilution to be made on receipt (e.g. 20µl at 100-500ng/µl). It is likely that, in most cases, the output of the DNA extraction protocol will be sufficiently standard for this purpose.

Optical density spectra from 220nm to 350nm will be taken on a Nanodrop spectrometer for each sample as part of the logging in process in the HTSF. This will allow suitable quantification and quality values to be calculated (e.g. DNA purity value = $[OD_{260}-OD_{320}]/[OD_{280}-OD_{320}]$). A suitable homogenisation procedure will be carried out on all samples received by the HTSF before measurement (e.g. agitation on a roller at room temp for at least 2hrs). Measurements will be taken in triplicate and the variation between readings should fall below a defined acceptable limit (e.g. 20% cv). Samples failing to meet the variation requirements will be passed through a more stringent homogenisation procedure before re-testing. This process should be repeated a limited number of times until acceptable readings are achieved or the sample is deemed to not meet quantity requirements. Quantification results will be logged against the sample number in the HTSF database and the average thereof used for the normalisation and preparation of working dilutions (see PCR setup).

The use of suitable sample quality checks will be finalised once an operational system is in place. However, we envisage using the control gene PCR as described by van Dongen et al. 2003 as well as running a small aliquot of the sample (e.g. 5µl) on 0.8% agarose with a Hind III digest under defined electrophoresis conditions (e.g. 1 hour @ 100v) to check sample quality.

All the data described in this section will be logged against the relevant sample in the database. Repeat samples will be requested for any samples failing to meet quantity or quality requirements. These requirements will be determined empirically once a working system is in place.

5. Sample storage

Samples will be stored in 96 position re-configurable arrays (e.g. Matrix system) in the plate and position allocated at sample log-in and will always be returned to their this position after usage. Short term storage (e.g. ≤2 weeks) will be at 4°C to allow live samples to be batched for testing before freezing thus minimizing any freeze/thaw effect on the sample quality. Long term storage (>2 weeks) at -20°C.

Different tests or groups of tests will be recapped with appropriate colour coded caps (example system: BRCA - red, HNPCC - blue, Marfans - green, FAP - yellow, other - black). It is anticipated that sample tube will never be manually relocated from their storage position on a storage rack (see batching). However, since sample tubes will not be labeled way other than a 2D bar-code, this will provide a simple check that batching has been carried out correctly.

6. Batching

6.1. Sample collection

To maximise efficiency of the primary screen the number of samples in each batch for testing should be maximised. The target is to report results in eight weeks; therefore, to allow for handling times not associated with the HTSF (both pre and post testing) as well as an allowance for contingency, the aim of the HTSF will be to report results within four weeks of sample receipt (half the allocated time). This time can be split into sample collection time and batch processing time which, is further sub-divided into primary screen (1°), secondary screen (2°), and reporting (figure 1).

Figure 1: Simplified process schedule for a single test (e.g. BRCA full screen).
Two weeks will be allowed for sample collection. After this time all samples will be arranged in a single batch and tested in parallel for all the fragments required for the screen (e.g. 85 fragments for a full BRCA screen). One week will be allowed for the primary screen (CSCE and analysis) after which all fragments with a positive result will be re-batched and passed to the secondary screen. One week will be allowed for the secondary screen, analysis, checking and reporting. The two weeks allowed for the processing phase of batch 1 will constitute the collection phase for batch 2.

Batch	**Week**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	Sample collection		1°screen	2°screen + report				
2			Sample collection		1°screen	2°screen + report		
3					Sample collection		1°screen	2°screen + report
4							Sample collection	

6.2. Primary screen batching

Under the batching scheme the number of samples in each batch to be tested will vary; in most cases more than one plate will be required to complete the primary screen. To keep the process simple, batching schemes will be arranged so that all plates required for the primary screen will have the same sample lay out. This will allow a single master plate to be prepared, which can then be replicated as many times as required to complete the screen. Importantly it will also mean that, within a screen, one sample will always be in the same plate position (e.g. C06) thus simplifying sample tracking.

The variables that define the plate layouts required for the primary screen are: number of samples in the batch, number of fragments to be screened for each sample, and the number of controls required for each fragment. In a simple scheme three controls will be used for each fragment tested: a mutation positive plasmid control, a mutation negative plasmid control, and a water control. With this number of controls there are nine batch sizes that fit on a 96 well plate with no empty wells: 1, 3, 5, 9, 13, 21, 29, 45 and 93 (figure 2). These batch sizes are subsequently be referred to as 'standard batches'. With different numbers of controls different standard batches will be possible.

Samples for each successive batch will be collected for two weeks. At the end of this time the largest possible standard batch will be compiled starting from the oldest sample and selecting samples in date order. Any remaining samples will be included in the next batch with the start of the two week collection period being set at the date of receipt of the oldest remaining sample. These operations should be performed automatically by the LIMS (or associated add-on) and the relevant milestones flagged (i.e. start of batch collection, end of batch collection).

The LIMS will provide a list of storage plates on which the samples for batching are stored. This list will comprise both manual numbering and a bar code for that plate. These plates will be manually retrieved from the store (fridge or freezer depending on the length of storage) and their identity checked by scanning the bar-codes. All storage plates will be briefly centrifuged before the re-arraying procedure is commenced.

Batches will be compiled from the retrieved storage plates using an automated tube handling system (BioMicroLab XL9). This provides eight input rack locations, one output rack location and a bar-code reader. Each location on the deck is bar-coded so that the robot knows where any rack is located at any time. Location of the storage racks on the deck will be done in conjunction with the rack ID check described above.

The samples for testing will be selected from their storage locations and temporarily re-arrayed in a new 96 well rack in the output rack location, starting with the oldest in position A1, the next oldest in position B1 and continuing in date order down successive columns; the array will be completed with a mutation positive plasmid control and a mutation negative plasmid control. A work list for

this re-arraying will be generated by the LIMS (or associated add-on). During the batching process on the automated tube handling system all sample bar-codes will be checked to ensure tubes are positioned correctly. These data will be fed back to the LIMS and stored in the database against the appropriate sample.

The completed batch of sample tubes will be transferred to a cooling block to allow the liquid wax to solidify. Once this has happened it will be possible to remove the tube caps without the risk of aerosol contamination. The batch rack will then be transferred to the pre-PCR robot where a master plate for the batch will be made.

The master DNA plate will be made by taking an aliquot of each sample and diluting it to 2.5ng/µl in a suitable volume according to its concentration as stored on the HTSF database. For example, in the standard batch of nine samples illustrated in figure 2b, 85 tests will be required on each sample. However, eight tests can be performed on a single plate, so only 11 replicates of the master plate will be required (85 ÷ 8, rounded up). Each PCR will require 5ng of DNA or 2µl of the final dilution. Therefore each well of the master plate will require a final volume of 11 replicates x 2µl = 22µl plus an allowance for wastage (~20%) giving a total of 27µl. The minimum pipetting volume should be 2µl with a wet dispense to maintain accuracy (i.e. the diluent should be dispensed first).

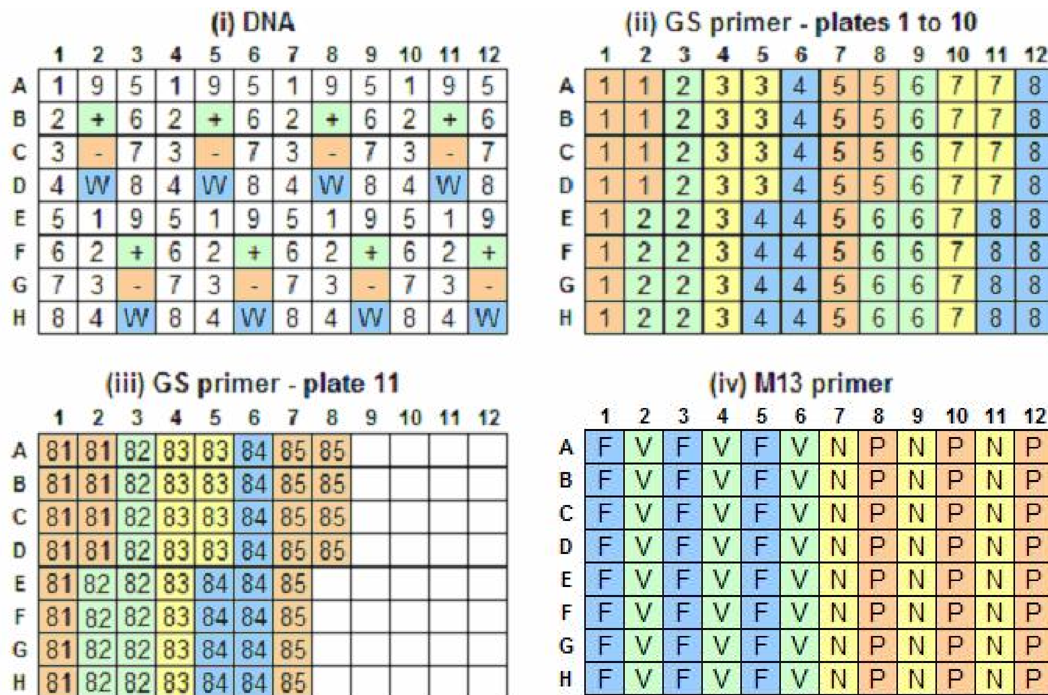
The dilution process will need to be performed in two stages to allow dilution over two orders of magnitude. The intervening dilutions will require vigorous mixing and / or incubation to ensure complete homogenisation before the final dilution is made. Final details will be derived experimentally. A similar incubation / agitation will be required for the completed master plate before it is replicated onto the PCR plates. After completion of the master plate the original samples will be replaced in their allocated storage locations on the automated tube handling system.

Figure 2: Standard batching schemes.

(a) Characteristics of the nine standard batch sizes using three controls that fill a 96 well plate with no empty wells as applied to a full BRCA screen requiring investigation of 85 fragments. Each plate used for the screen will use the same sample layout (see figure 2b) except the last plate, which in most cases (except a batch of 93) will be only partially used.

Batch size	No. of controls	Tests per fragment	Fragments per plate	Plates per screen (85 fragments)	Empty wells on last plate
1	3	4	24	4	44
3	3	6	16	6	66
5	3	8	12	8	88
9	3	12	8	11	36
13	3	16	6	15	80
21	3	24	4	22	72
29	3	32	3	29	64
45	3	48	2	43	48
93	3	96	1	85	0

(b) The plate layouts for a batch of nine samples (1 to 9) with three controls (+ mutation positive, - mutation negative, W water) in a full BRCA screen requiring 85 fragments (highlighted in figure 2a). The screen requires a total of 11 plates. (i) Shows the DNA layout; (ii) shows the layout of gene specific (GS) primer pairs on plates 1 to 10 (plate 2 will use primers 9 to 16 inclusive and so on); (iii) shows GS primer pair layout for plate 11 and; (iv) shows layout of M13 primer pairs (M13F in each case is fluorescently labelled: F=FAM, V-VIC, N=NED, P=PET). The distribution of M13 primers will be constant regardless of batch size. This layout facilitates the mixing of the PCR products before detection on the capillary analyser.



All calculations regarding batching and plate layouts will be performed in a utility attached to the HTSF database / LIMS. This utility will generate work lists that will direct the robotics to perform the correct sequence of applications as required. The system should be capable of handling 384 well plates in a similar manner. This utility will also control sample collection and flag critical dates for sample testing to ensure timely reporting. It should be possible to override the work scheduling system in order to fit in with the day to day running of the lab. However this should be kept to a minimum and will only be permitted when authorised by a senior member of staff.

6.3. Secondary screen batching

All fragments with a positive result for the primary screen will need to be sequenced to confirm and characterise the mutation present. PCRs will be setup from a fresh aliquot of the original sample as a final precaution against sample mix-up: information regarding which samples and which fragments are to be tested will be logged on the LIMS during analysis of the primary screen. This information will be used to generate work lists for the pre PCR robot.

As with the primary screen, the possibility of using automated sample arraying to select the samples for the secondary screen will be considered.

7. PCR setup

7.1. PCR reagents

PCRs will be setup according to the SPODS protocol (Appendix A) in 10µl reactions. The components of each reaction are detailed in table 1. Exact concentrations of master stocks and working stocks are subject to ongoing optimisation and will be finalised during implementation. Similarly sealing options will need to be investigated to establish the most suitable system. Ideally working stocks of primers will be stored in 96 well plate format (one plate per test) and sealed using liquid wax as this will allow the greatest flexibility.

All reagents will be stored in bar-coded tubes or plates. Details of the current working stock and all master stocks will be held on the LIMS. During batch setup the user will print off a work list that will contain bar-codes for all suggested reagents to be used. Usage will be confirmed during deck setup by means of scanning the appropriate reagent on the work list; the LIMS will then indicate what the reagent is and where it should be located on the deck. It will be possible to override this system to deal with situations where reagents run out. In these instances both the suggested reagent and the reagent actually used will be logged against the appropriate process in the LIMS.

7.2. Primary screen PCR setup

Fragments that are known to contain common polymorphisms will require screening to exclude them from the secondary screen at the post PCR stage. These fragments will be flagged in the LIMS and this information used to organise the fragments that require polymorphism screening on to the same plate[s]. Thus The PCR primer layout described in figure 2(b) will not be in numerical order but arranged to facilitate the polymorphism screen (see 8.2 below).

PCRs will be setup in maximum batches of eight plates regardless of the number of plates required for the entire screen. This will allow direct manual transfer to the PCR cycling blocks. Input stacks to handle plates and tips should, at minimum, be capable of dealing with this throughput. Substantially higher throughput than this would be desirable to cope with future expansion but this will be cost dependant. A cooled (4°C) output stack for completed plates would also be desirable.

Reagent dispensation patterns will be controlled by work lists generated by the LIMS as described in figure 2. All aspirations and wet dispensations will be a minimum of 2µl to maintain accuracy. Dispensations should be in the order; largest to smallest. In practice it is likely that the master mix will be dispensed first followed by the water/glycerol followed by the primers and then the DNA. All actions performed by the robot will be stored in the LIMS against the appropriate plate to provide an audit trail.

The use of coloured dye[s] in various reagents and / or DNA samples to facilitate visible checking that the correct additions have been made will be investigated. This may or may not be necessary depending on the accuracy of the system and may, in fact, not be possible due to interference with fluorescent detection.

Sealing options for the PCR plates will be finalised during implementation. It would be preferable to use liquid wax as this will facilitate the most automated and flexible system. This would also prevent open plates standing in the output stack for any period of time thus guarding against cross-contamination. However manual and automated sealing using foils will also be investigated.

All plates will be agitated and pulse centrifuged before loading on to the PCR blocks. The plate bar-codes will be manually scanned when they are removed from the robot, before PCR cycling. User and time information will be stored on the LIMS against the plate record to build a comprehensive audit trail.

7.3. Secondary screen PCR setup

As with the primary screen the patterns of reagent dispensation will be calculated using a LIMS utility that will generate the appropriate work lists for the robot. In this instance on the unmodified M13 primers will be used (M13FU and M13RU).

Table 1: PCR reagents

Reagent	**Comments**
Template DNA	Handling of DNA is detailed in section 6 batching
Gene specific (GS) primers	Working stocks of GS primers will be stored in pair mixtures to simplify addition to the reactions. In cases where they are supplied as single primers working stocks will be prepared by mixing appropriate pairs. The concentration of GS primers used in the PCR is likely to be too low to allow reliable storage at an appropriate dilution for any period of time. Dispensation of these primers will therefore need to be done in two stages to achieve the desired concentration in the final reaction. Thorough homogenisation of the intervening dilution will be required before final dispensation.
Universal primers (M13)	Working stocks of M13 primers will be stored in pair mixtures to simplify addition to the reactions. The forward primer in each pair will be fluorescently labelled with one of four fluorophores (FAM, VIC, NED or PET). In cases where they are supplied as single primers working stocks will be prepared by mixing appropriate pairs. Single un-paired stocks of the two unmodified primers will also be required for use in the secondary screen (sequencing). Thus in total six M13 primer stocks will be required.
Amplitaq Gold Master mix	This contains polymerase and will need to be located in a cooled reagent rack when on the robot deck.
Glycerol (this reagent may not be used)	Working stock of glycerol will be as dilute as possible given the reaction size and the final concentration required. This is likely to be about 50% which will require very careful pipetting.
H_2O	MilliQ water will be used. It would be advantageous to combine this addition with the glycerol to decrease its viscosity and reduce the number of additions required to set up the PCR.

8. Post PCR processing

On completion of the PCR cycling, plates will be removed from the blocks, pulse centrifuged and the bar-codes scanned. Regular scanning of the bar-codes and logging of data will allow tight control of where any plate is in the process at any time.

8.1. Primary screen

Several post PCR processes will need to be optimised during implementation. PCR products will be diluted by one or two orders of magnitude and then mixed together in a four-plex containing the four differently labelled products (FAM, VIC, NED and PET) before capillary analysis. Multiplexing makes more efficient use of time on the capillary analyser and dilution reduces the relative salt concentration in the reactions facilitating more effective electro-kinetic loading.

One of the greatest potential causes of failure will be measurements on the capillary analyser falling outside the readable quality window (currently 1000 to 25000 rfu). Therefore, it would be desirable to determine the concentration of PCR products before the dilution so that the concentrations of the component PCR products in the four-plex could be normalised to an appropriate level for good quality results. This process will be done before sample mixing and could be carried out using a fluorescent (Pico Green) or optical density based assay on a spectrofluorometer integrated with the post PCR robot. Alternatively quantification could be achieved using a kit based gel electrophoresis system such as the E-Gel system from Invitrogen.

Dilutions may require a two stage process, although this will be avoided if possible. Thorough homogenisation of any intervening dilutions will be required. CSCE will be performed on the capillary analyser using a water loading protocol; no addition of formamide will be necessary for this process.

Mixing of PCR products will always be done by combining the same wells within a single plate with each other. However, since this mixing allows four PCR plates to be compressed onto a single loading plate, samples originating in one well may ultimately be analysed in any one of four wells (figure 3).

To facilitate maximum automation, input and output stacks for plate and tip box storage will be required. A capacity of about 20 plates should be sufficient but more would be desirable depending on budget. Because plates are multiplexed by four during the processing the capacity of the output stack, in terms of plates, need only be a quarter of the input stack. Because the layout of M13 labelled primers is standardised (figure 2b), it is not essential to mix plates in a particular order in the multiplexing process; all that is required is to log which fragments are in which wells. This will require an integrated bar-code reader on the post PCR robot to identify plates as they are processed.

8.2. Polymorphism screen

It is intended that an assay to type known common polymorphisms will be introduced in order to exclude them from the secondary screen. Without such an assay the expected sequencing rate will be about between 10% and 20% of the fragments screened by CSCE (0.2% pathogenic mutations, 8% common polymorhisms with the remainder being false positives or failures). Thus a polymorphism screen should reduce the required throughput of the secondary screen (sequencing) by >90%. To be viable this assay needs to be fast and cost effective compared to sequencing. It is not critical that this assay be fully operational in the early stages as sample numbers are likely to be relatively low. This assay will therefore be the subject of ongoing evaluation during implementation of the HTSF.

Three options for the polymorphism screen are currently being considered:

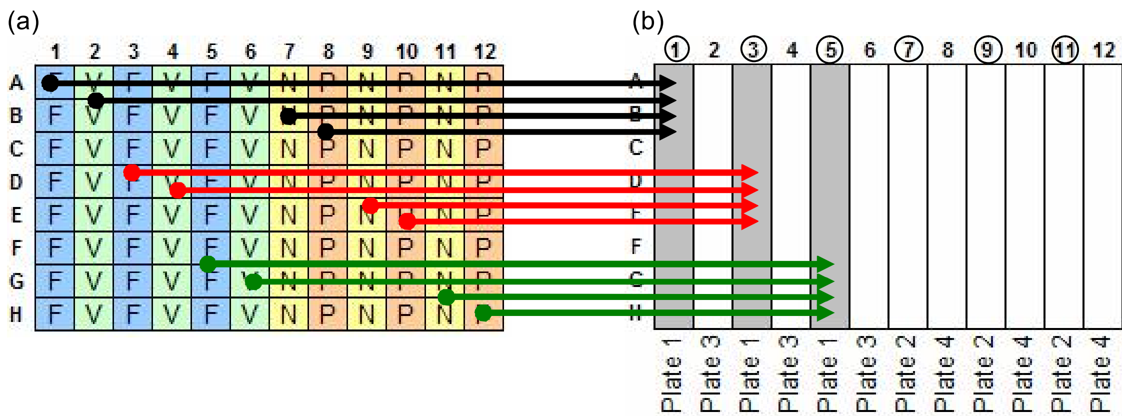
- Template directed incorporation (or primer extension) assay with detection by fluorescence polarization (FP-TDI) – AcryloPrime system from Perkin Elmer.
- Oligo ligation assay – SNPlex from Applied biosystems.
- Multiplex ligation dependant probe amplification (MLPA).

MLPA would appear to be the most suitable method since the polymorphism screen could be incorporated into the deletion screen that will be required for full diagnostic testing.

Results from the polymorphism screen will be read in conjunction with CSCE results to determine which fragments require full sequencing and which can be excluded on the basis that they only contain a non pathogenic polymorphism.

Figure 3: Sample mixing

Columns 1, 2, 7 and 8 on the PCR plate (a) are combined into column 1 of the loading plate (b) indicated by black arrows. The fluorescent labels on the PCR products in the PCR plate are indicated by F=FAM, V=VIC, N=NED and P=PET. Thus each well in column 1 on the loading plate contains one PCR product labelled with each of the fluorescent dyes. The two other mixing combinations are indicated by the red and green arrows respectively. The loading plate can accommodate four complete PCR plates when mixed. The offset pattern in which the plates are arranged on the loading plate is arranged to suit the loading pattern of an ABI 3730 48 capillary analyser so that two PCR plates worth of data are collected in a single run; ringed columns are loaded in run 1 whilst the remaining columns are loaded in run 2.



8.3. Secondary screen

In addition to sample dilution and mixing the post PCR robot should be capable of setting up sequencing reactions for the secondary screen. This will involve a PCR clean-up; most likely by filtration plate or magnetic bead separation as these provide the most automated solution but a SAP / EXO1 protocol will also be considered. We envisage that the necessary equipment for these trials would be fitted to the robot on a trial basis and only purchased once the final decision has been made.

Sequencing reactions will be setup in 5 to 10µl volumes (depending on dispensation constraints). Since all PCR products will be tagged with M13 sequences at either end the sequencing reaction setup will be standardised to two protocols; one using M13F primer and one using M13R. Sequencing reagents will require a cooled reagent block.

Again sealing options for the PCR plates will be finalised during implementation but use of liquid wax would be preferable. All plates will be agitated, pulse centrifuged and the bar-codes manually scanned before cycling. User and time information will be stored on the LIMS against the appropriate plate.

As with the PCR cleanup the method used for sequencing clean up will be subject to investigation. The preferred method is filtration plate but bead based methods and ethanol precipitation will be considered. Because of the cost associated with

filter plate cleanups these are likely to be performed on 384 well plates if this method is chosen. This will require re-formating of the plate layouts to ensure that data for each plate is collected into its own file during capillary electrophoresis. The use of the 384 well option becomes more critical with higher throughput and will be integral if the fall back position of sequencing for the primary screen becomes necessary. The investigation of sequencing cleanup will be performed in conjunction with the PCR cleanup. Compatible methods will be chosen to limit the number of integrated applications required on the post PCR robot.

It is likely addition of formamide to 50% will be required to prevent evaporation of the sequencing reactions in the capillary analyser before loading. However, no plate sealing will be required at this stage.

The throughput for the secondary screen is envisaged to be considerably lower than that for the primary screen. Thus no extra provision in terms of plate stackers will be required. Currently the contingency fall back analysis method is to use sequencing for the primary screen. Again no extra capacity provision is envisaged for this eventuality although this should be borne in mind when designing the cleanup protocols.

9. Capillary electrophoresis

Capillary electrophoresis for the primary screen will be carried out on an ABI Prism 3730 48 capillary DNA analyser. This instrument requires the input of a sample sheet for each run which is linked to a plate bar-code reference. To save complications with LIMS integration it is anticipated that this instrument will be run from the associated workstation as installed by ABI. For each plate the LIMS will generate appropriate sample sheet(s) with all the associated run and analysis modules and an embedded plate bar-code reference in tab delimited format. These sample sheets will be uploaded to the 3730 work station from the HTSF network.

During operation the 3730 reads the bar-code on each plate in the input stack and automatically associates the correct sample sheet with the run. Any plates that do not have an appropriate sample sheet defined in the system will be passed directly to the output stack without analysis. This eventuality will need to be flagged either in the ABI system software or, preferably, by the LIMS. Run records can be output to a defined location on the HTSF network where they can be uploaded by the LIMS for incorporation into the audit trail. Similarly output data can be saved to a defined, run specific location for analysis.

It is anticipated that the secondary screen will be performed on a second, smaller capillary analyser (probably an ABI 3130xl). Data flow and sample tracking will be as described above for the primary screen.

10. Sample and reagent tracking

The primary identifier will be the sample number allocated by the originating lab during sample preparation. This number will be stored in the HTSF database as a new test. When a sample is activated (i.e. included in a testing batch) a batch number, a 96 well plate location, and a 384 well plate location will be logged against the sample. At the same time a batch record will be initiated and designated live. It will be possible to have more than one batch record logged against a single sample reference regardless of the status of the batch (either live or complete). All details of plates and processes appropriate to the batch will be logged against this batch number. The status of a particular sample within the system at any time will be accessible by reference to the batch number logged against the sample and the status of that batch.

Particular milestones will logged against each sample: for example; master plate complete; all fragments amplified by PCR; all fragments analysed by capillary electrophoresis; all fragments analysed; report complete. Time deadlines for each of these milestones, based on the date of sample receipt, will be automatically generated at sample login. This will allow the progress of screening to be monitored for each sample / batch with any potential problems highlighted at an early stage enabling timely remedial action.

In a similar way stocks of reagents will be tracked and, on the basis of numbers of samples received, future reagent requirements will be predicted to ensure appropriate supply. In addition the batch identity of each reagent used for a particular test will be logged. This should highlight any defective or out of date batches of reagents.

11. Analysis

11.1. Primary screen

Initially analysis of the primary screen will be performed using the generic software provided by ABI (*GeneMapper*). Currently *GeneMapper* is not ideally configured for the analysis of CSCE although it can be satisfactorily achieved. It is hoped that ABI will take on support of this methodology and provide software specifically designed for the purpose; however this cannot be guaranteed.

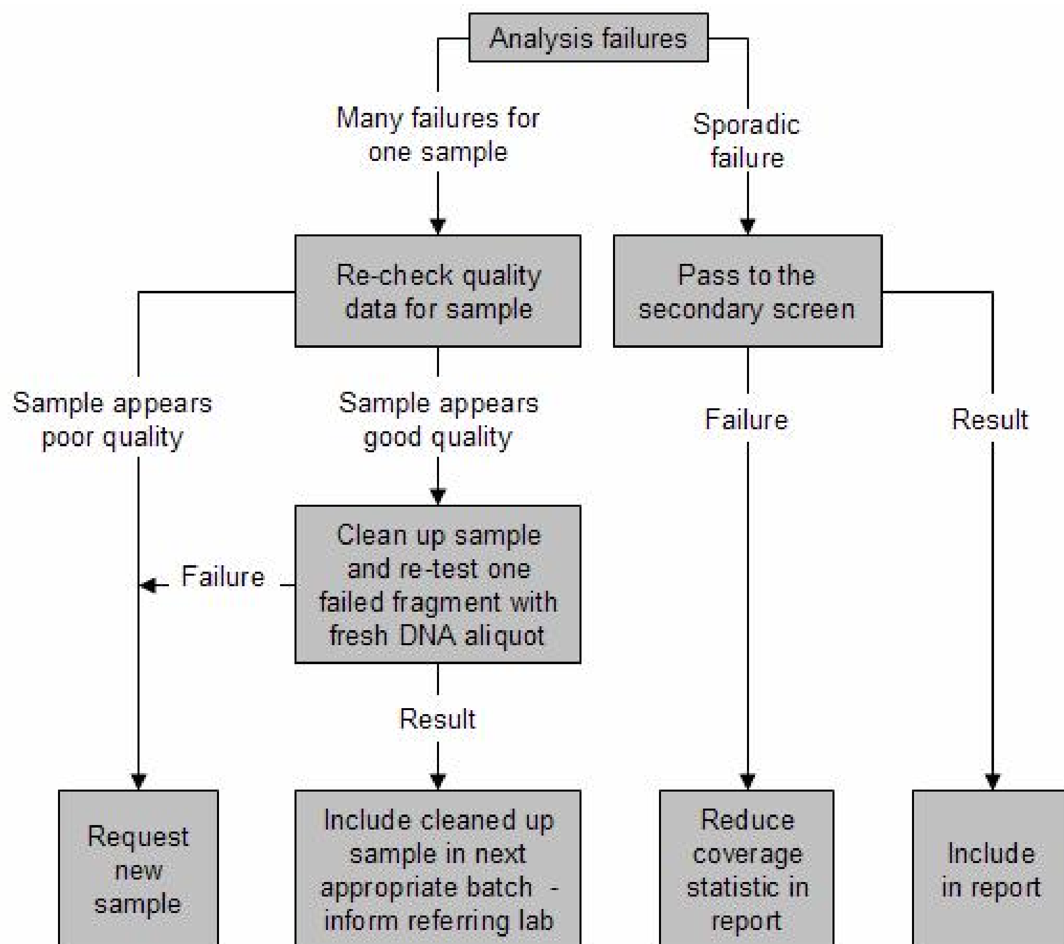
Another option being investigated is for a bespoke utility to be written in conjunction with a software developer. Soft Genetics, who have developed *Mutation Surveyor,* have expressed a tentative interest in such a scheme. An ideal software would output a numerical expression of the difference between a defined pair of reference and test electropherograms. The analysis should accommodate both split peaks and shouldering and should allow quality thresholds to be defined.

Ideally electropherograms and, at a later stage, when analysis can be more automated, quality data and calls, should be available via the LIMS. This will allow data to be called up for a particular sample and a decision to be made without the risk of entering a result against the wrong sample. Any pre-screening data relating to the presence of known non-pathogenic polymorphisms should also be available at this stage. At this point of the process the decision will be binary; [a] the fragment analysed is normal (or only contains a known non-pathogenic polymorphism) and no further testing is required or [b] the fragment should be passed to the secondary screen for further analysis. This information will be logged against the sample reference and used to compile a batching schedule for the secondary screen. Clearly each batch record for each sample will need to be able to accommodate multiple results (one for each fragment tested – at least 100).

11.2. Failures

Any fragments failing to yield a satisfactory result in the primary screen will be flagged in the LIMS during analysis. On completion of the primary screen analysis of the frequency of failures for each sample will be assessed.

Figure 5: Failure handling algorithm.



Where failures are considered to be assay specific (i.e. infrequent and sporadic) the fragment will be passed to the secondary screen as if they were mutation positive. If it is clear that the failure is sample specific (i.e. a high proportion of the fragments for a particular sample fail), the quality data for that sample will be re-examined. If the quality of the original sample appears good the sample will be cleaned up and a repeat assay using a fresh aliquot of the sample will be setup using one of the failed fragments. If this is successful the sample will be entirely re-tested using a fresh aliquot in the next appropriate batch and the referring lab will be notified. If the quality of the original sample is poor a repeat sample will be requested from the referring lab.

Failures in the secondary screen will be repeated until they pass unless the secondary screen is a result of a failure in the primary screen in which case the analysis will be terminated and the coverage statistic included in the final report will be reduced to allow for the number of bases not screened. An algorithm describing the failure handing process is given in figure 5.

11.3. Secondary screen

The choice of programme to handle sequencing analysis for the secondary screen is currently the subject of a detailed programme comparison being organised jointly by the reference labs. However, in the short term the programme of choice is likely to be *Mutation Surveyor* from Soft Genetics. Even with a false positive rate of 20%, the primary screen will not generate more than two plates of sequencing per week. Given this low anticipated throughput for sequencing in the HTSF, the numbers of analyses is unlikely to be problematic.

As with the primary screen, electropherograms should be available for access via the LIMS allowing direct input and confirmation of results. A link will be provided between the LIMS and a database which will store sequence, primer and variation data, to facilitate the interpretation of variations. An Excel base database performing these functions already exists.

12. Reporting

Once analysis of all fragments is complete the LIMS will provide utility to review all results from a single sample. Using this information all significant variations will be compiled into a final report, which will also contain the sample reference and details of the tests carried out with the coverage achieved. Since no patient or clinician information will be held by the HTSF database the report will be compiled from a database of standard texts with variable fields inserted to suit the findings.

Essentially the report will comprise a simple text file that can be merged with the clinician details by the referring lab. Since the report will contain no de-codable details identifying the patient, it should be possible to send it by email. Fine details

of the precise format of the report and the data transfer procedure will be finalised during implementation.

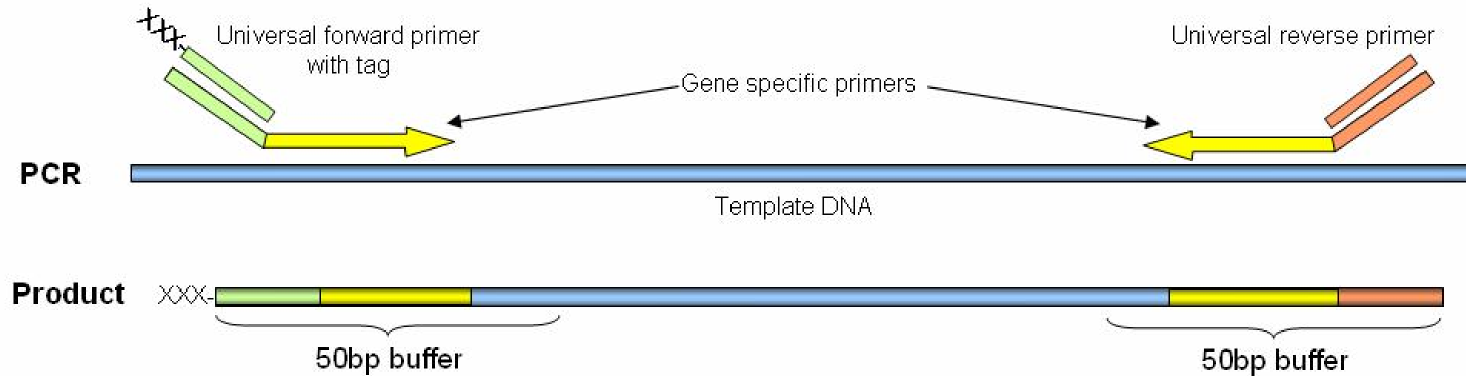
13. System Control

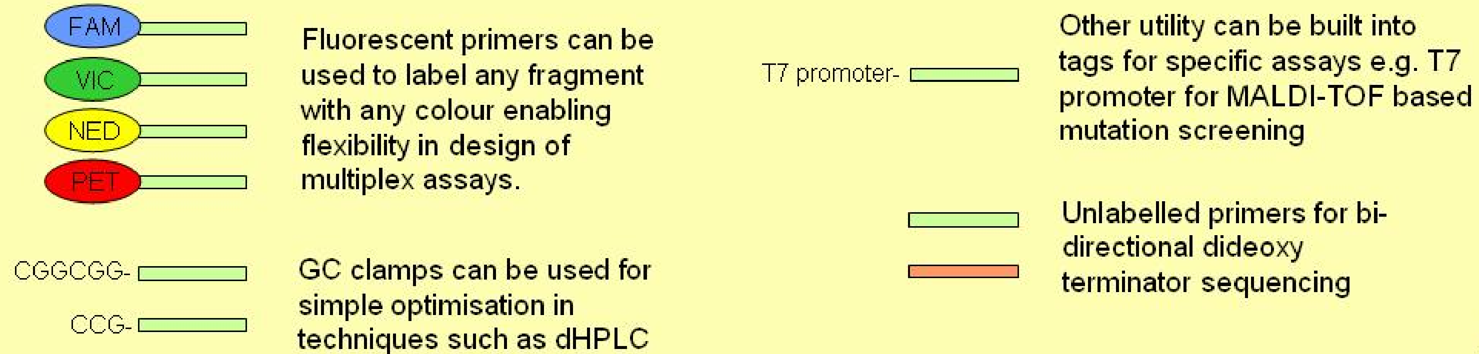
It is intended that the system be controlled centrally to provide a direct link between data held on the LIMS and the setup of processes carried out by the various instruments. Work lists will be generated by the LIMS and these will be used to control the various processes. A fully integrated system is not envisaged. That is sample flow through the process will be interspersed by manual relocation of plated between instruments. This should allow the greatest amount of flexibility in the system. At each location in the system where a manual move is made (e.g. pre-PCR robot to PCR machine) a bar code reading of the plate being moved will be taken to log the parameters of the movement (date, time, operator, from, to etc) on the LIMS.

Each instrument will be started from the local workstation but strict controls over who will be authorised to edit protocols will be in place. It is envisaged that the protocols for the robots will be written in a modular format. This will allow different modules to be called depending on the test being performed thus compiling a complete protocol specific to the current requirements. We think this format will provide the greatest flexibility and be the easiest to retrieve after a system failure although advice will be sought on this matter.

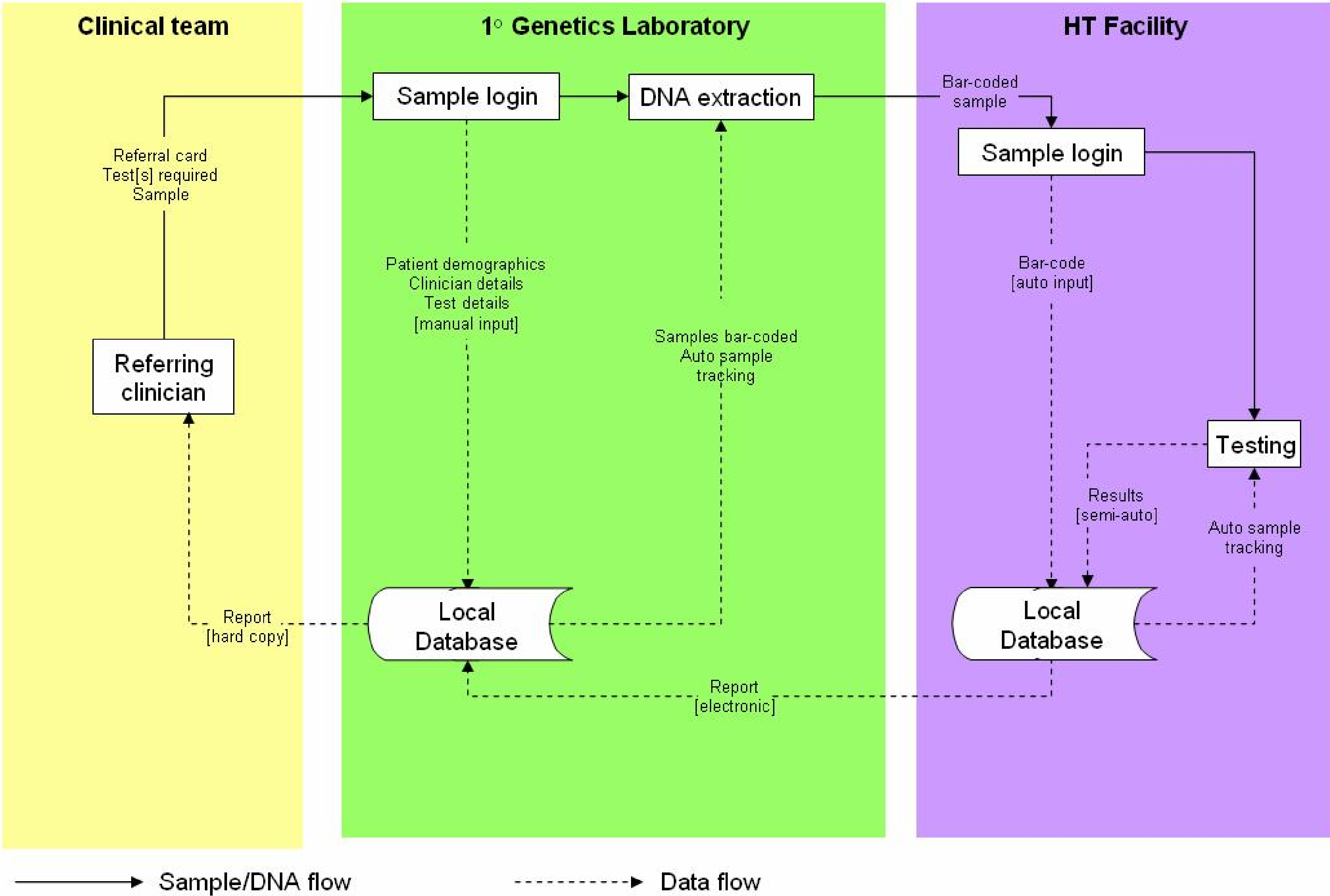
Appendix A: Standardised primer design and optimisation (SPODS)

All gene specific primer pairs (yellow) are designed with one of two universal tags; each PCR having one forward and one reverse tag. PCR is performed in a single reaction containing both gene specific primers and universal primers (red and green) using standard thermal cycling conditions. A panel of 5' modified universal primers can be used to adapt the PCR to a wide range of analysis techniques (further details available at http://www.ngrl.co.uk/Wessex/spods.htm)

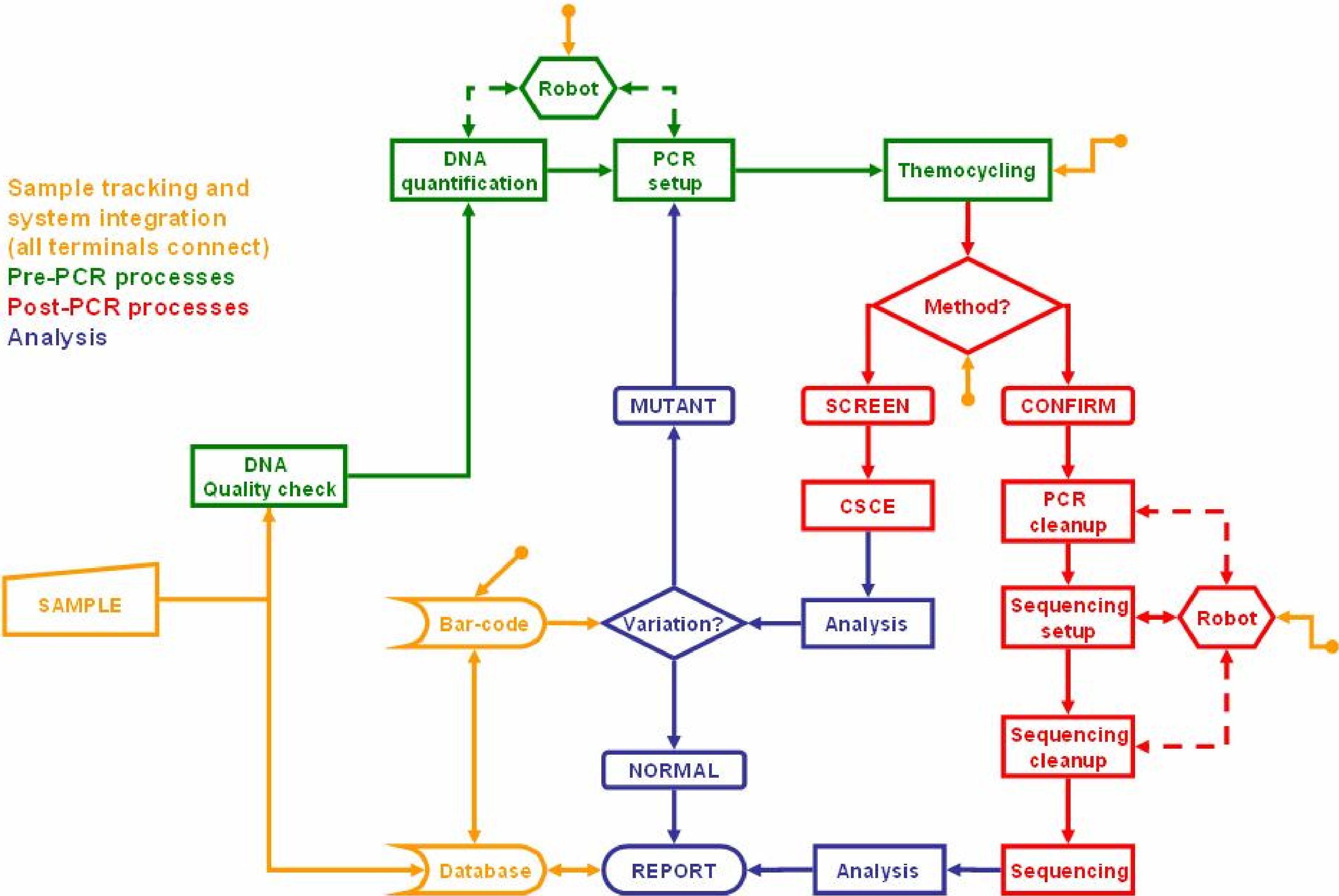




Appendix B: Outline of proposed usage of the HTSF



Appendix C: Simplified process diagram for the HTSF



Appendix D: Operational policy for high throughput mutation scanning within the SCOBEC cluster

1. Aims

1.1. To maximise efficient use of instrumentation

1.2. To maximise efficient use of scientific expertise

1.3. To redistribute work load to appropriate level

1.4. To minimise sample turn around time

1.5. To maintain skill mix (technical and theoretical) at local level

1.6. To improve equality of service across disorders – more time to focus on minority disorders.

1.7. To free time to allow development of local research and development links (eg with universities)

1.8. To improve QA by standardisation of process

2. Outline remit

2.1. Sample screening only – predictives / pre-symptomatic testing will not be undertaken. For tests that are carried out by the HT facility primers and protocols will be available for other labs to use as required

2.2. Tests to be SPODS compatible (see appendix A).

2.3. Samples received and processed by barcode only

2.4. Sample tracking and result allocation will be entirely in-silico.

2.5. Samples will be processed and all sequence variations characterised at DNA level and protein level as appropriate.

2.6. Reports will be returned by electronic transfer with results associated to the barcode.

2.7. The report text will be in a format suitable for inclusion in a clinical report for the referring clinician. The report will deal with the patient in isolation with no reference to any other family members.

2.8. Reports will include coverage/confidence score for 'normals'

2.9. Data on all 'normal' fragments archived

2.10. Each centre to retain clinical control and testing management on local referrals. Other arrangements will be made for dealing with any pre-symptomatic / predictive testing that is required.

2.11. Initially focus will be on large cancer genes, in particular BRCA1 and BRCA2, but other disorders including multi-factorial diseases would be considered. This will be based on group opinion and demand.

2.12. Specific projects will be undertaken in collaboration with individual labs.

3. General principles

(See appendix B for outline proposal)

3.1. DNA will be extracted and bar-coded at the source lab.

3.2. Bar codes will contain three pieces of information; referring lab, unique patient ID and test required (in this case all samples will be for BRCA full screen).

3.3. To ensure all tube transfers are correctly carried out and recorded the following protocol will be adopted:

3.3.1. For each tube transfer (regardless of what process is being carried out) the main database will generate a work list for the appropriate robot (pre/post-PCR)

3.3.2. The user will scan pre-bar coded labware which will then be allocated to the work list.

3.3.3. Before the robot performs the allocate task it will first check both the source and destination labware for the barcodes in the work list.

3.3.4. If there is any discrepancy the process will be halted and details of the error fed back to the main database.

3.3.5. The task will only be performed if the deck of the robot is correctly set up.

3.4. All instrumentation will be set up to feed back information regarding the process carried out to the main database. This information will be added to each sample record as appropriate to build up a complete audit trail for the sample.

3.5. Currently we are pursuing CSCE as the method of choice for screening. In the event that CSCE is unsatisfactory in any way a sequencing strategy will be adopted. This will cost in the order of 5x more than a CSCE approach. We have also investigated a MALDI-TOF approach but this is not currently considered to be fully operational although it is a very promising approach for the future.

3.6. It is envisaged that final analysis of the primary and secondary screens will be performed on different instruments (3730/3100 as appropriate). Each instrument will be dedicated to its purpose and overseen by a single operator.

3.7. Reports will contain identifying bar code number, the complete ascertained genotype for the sample to DNA level (including polymorphisms) and a complete audit trail for work carried out on the sample including relevant sample files . An agreed list of mutations and polymorphisms could be used to highlight those variations with definitely known function at this stage.

3.8. Before implementation of the system a full validation of the sample tracking will be performed. This will probably be by analysis of a plate of DNAs by identity PCR but suggestions would be welcome.

4. Backlog screening.

4.1. Primary screen

4.1.1. Upon receipt samples will be scanned to enter the patient ID onto the database. Samples will be sorted according to the 'test required' portion of the bar code.

4.1.2. The database will allocate each sample a position on a master plate. Each master plate will accommodate a suitable selection of controls (to be agreed). When a master plate is full the database will prepare a work list for the pre-PCR robot and add the master plate to the outstanding work list.

4.1.3. To set up a DNA master plate the user will select a plate from the outstanding work list and scan a pre-bar coded deep well plate to receive the samples. The database will assign the plate bar-code to the prepared work list. A list of samples for the plate will be printed out.

4.1.4. At the workstation the user will load all the samples on to the liquid handler along with the master plate assigned to those samples.

4.1.5. The liquid handler will use the information on the pre-prepared work list to perform the following functions:

4.1.5.1. Bar-codes on the samples and the master will be read to ensure they match the work list. If they do not the process will be stopped and a list of discrepancies will be fed back to the database as an error report.

4.1.5.2. Assuming samples and master plate identity are correct samples will be quantified and a suitable dilution made in the master plate.

4.1.6. A process report for each sample including quantification and dilution details will be fed back the main database.

4.1.7. The database will remove the plate setup from the outstanding work list and add all the tests required for that plate according to the 'test required' portion of the individual bar-codes (in this case ~100 fragments for BRCA screening will be added).

4.1.8. The SPODS system (appendix A) will be used for amplification. Fragments can be analysed in any order. To set up a PCR for a particular fragment the user will select the appropriate fragment from the outstanding work list and assign a pre-bar coded plate to the PCR by scanning. A work list will be prepared and sent to the pre-PCR robot. As with the master plate setup both master plate and PCR plate will be scanned to ensure they match the work list.

4.1.9. Primers will be arrayed in deep well plate for ease of access. Reaction components (including primers) will also be bar coded and checked in a similar way to ensure the correct components are added. Any discrepancies in the deck layout of the liquid handler will call a halt to the process and be reported by the system. A full audit trail of additions will be fed back to the main database. As each fragment is setup it will be removed from the outstanding work list.

4.1.10. Thermocycling should be integrated into the system. Currently we have not decided on the most appropriate instrumentation. However the Biogene instrument is capable of feeding back information regarding the exact thermocycling that took place through individual well measurement (as opposed to recording the program that was used). Some form of scan will need to be introduced at this stage (either manual or automatic) to alert the database that thermocycling has taken place.

4.1.11. Four plates of PCR can be screened by CSCE at once on the 3730 by appropriately mixing differently labelled products. Up to four plates would be selected from the completed PCR list. The database will check to ensure the products are differently labelled. A pre-bar coded loading plate will be assigned to the analysis by scanning the bar code. A work list will be generated and sent to the post-PCR robot. The database will also generate a sample sheet for the 3730 and a deck setup for the user.

4.1.12. Assuming the post-PCR robot deck layout is in agreement with the work list the robot will make the appropriate dilution of each sample and mix them into the loading plate along with a size standard. Again this process will be logged into the database.

4.1.13. Plates can be loaded onto the 3730 in any order. The instrument will read the bar codes on the loading plates and assign the correct sample sheet for that plate.

4.1.14. On completion of the run sample files will be automatically analysed and assigned to the appropriate sample in the database. Currently it is envisaged that analysis will be carried out using GeneMapper software but other options are being investigated. However analysis is performed the data will only be viewed via the main database. Each sample will have a list of completed analyses linked to it (including normal controls so that comparisons can be made). These will be viewed one at a time and a decision regarding the mutation status will be entered at the point of viewing. Controls will be in place as to who is able to access this utility.

4.1.15. Normal fragments will be reported and mutation positive samples will be flagged for secondary screening.

4.2. *Secondary screen (Confirmation)*

4.2.1. Samples will be grouped by the fragment and arranged on a plate. A work list generated and sent to the pre-PCR robot. A new PCR plate will be allocated to the work list by scanning a pre-bar coded plate.

4.2.2. Secondary screening will be performed starting from a fresh aliquot of the original DNA received. After checking the robot deck is correctly set up the pre-PCR robot will set up the PCR. The work list in this case will be more complicated as different primers will be required at different locations on the plate.

4.2.3. All primer sets will be optimised to work using the same thermocycling profile. Thermocycling will be performed as per the primary screen.

4.2.4. The method for PCR clean up has not yet been decided. Main contenders are SAP/EXO and Millipore plate clean up. Regardless of method the same mechanism used in all previous stages will be used to ensure accurate and traceable transfer of samples from plate to plate. All details of processing will be fed back to the database.

4.2.5. Confirmatory sequencing will be set up in both orientations. Thus for each plate passing through the secondary screen two plates of sequencing will be generated. Again work lists will be generated and pre-bar coded plates allocated to the work list by scanning.

4.2.6. Sequencing clean up will most likely be performed on 384 well Millipore plates. Sample transfer and audit data will be dealt with in the standard way.

4.2.7. Completed reactions will be run on a 3730 or 3100 as appropriate. As with the primary screen sample file will be analysed automatically and assigned to the correct sample in the database. Viewing of results will be performed via the database to ensure results are assigned to the correct sample.

5. Routine screening

5.1. This will be performed in much the same way as backlog screening. The main difference will be that samples will be dealt with in small batches using pre-aliquoted plates of primers and reaction mix. In the early stages, for simplicity of set up and optimal turn around times, samples will be analysed one by one (i.e. one sample per batch). Although this significantly increases liquid handling (due to multiplication of controls) analysis will not be significantly increased. Machine capacity can easily cope with the required throughput. At a later stage other batch sizes could be introduced (if required) to reduce costs. Examples might be one batch per column or one batch per row of a plate.

5.2. The numbers of controls will need to be carefully considered. Each fragment will probably require one water control one positive control and one negative control. Consideration needs to be given to the course of action should any one of the controls fail. The PCRs are being optimised in order to minimise failures (in particular low DNA concentration) but this will be a key area of research.