

# Development of a service for hypertrophic cardiomyopathy using next generation sequencing – translation into a diagnostic laboratory

Oxford Molecular Genetics Laboratory  
Oxford Biomedical Research Centre

Jessica Woodley & Sarah Reid

# Aim/Strategy of translation

- Transfer the current diagnostic service for Hypertrophic cardiomyopathy (HCM) to next generation sequencing technology using the Roche 454 GS-FLX.
- Protocol designed and optimisation carried out by Oxford Biomedical Research Centre (results presented at CMGS Conference 2010)
- Further optimisation of protocol undertaken in the Oxford Molecular Genetics Laboratory over the last year.
- Aims of the work within the diagnostic laboratory have been to develop automation for laboratory work, investigate analysis methods and develop a cohesive workflow that meets CPA standards.

# Current Protocol

Gene	Protein	% of HCM	Exons screened
MYBPC3	Myosin-binding protein C	20-30	1-34
MYH7	Beta Myosin heavy chain	15-35	3-40
TNNT2	Troponin T	3-5	2-17
TNNI3	Troponin I	<5	1-8
MYL2	Essential myosin light chain	Rare	1-7
MYL3	Regulatory myosin light chain	<1	1-7
TPM1	Tropomyosin 1 alpha	<2	1-10
PLN	Phospholamban	Rare	2
ACTC1	Actin	Rare	2-7
CSRP3	Cysteine and glycine-rich protein 3	Rare	2-6
PRKAG2	5'-AMP-activated protein kinase subunit gamma-2	<1	7-16
GLA	$\alpha$ -Galactosidase	3	1-7
FHL1	Four and a half LIM domains 1	Rare	3-8
		<b>50-80</b>	<b>155</b>

- On each run 15 samples are analysed for 13 genes (155 exons) using 29 long-range PCR amplicons.

- Long range PCR products range from 3kb to 12kb

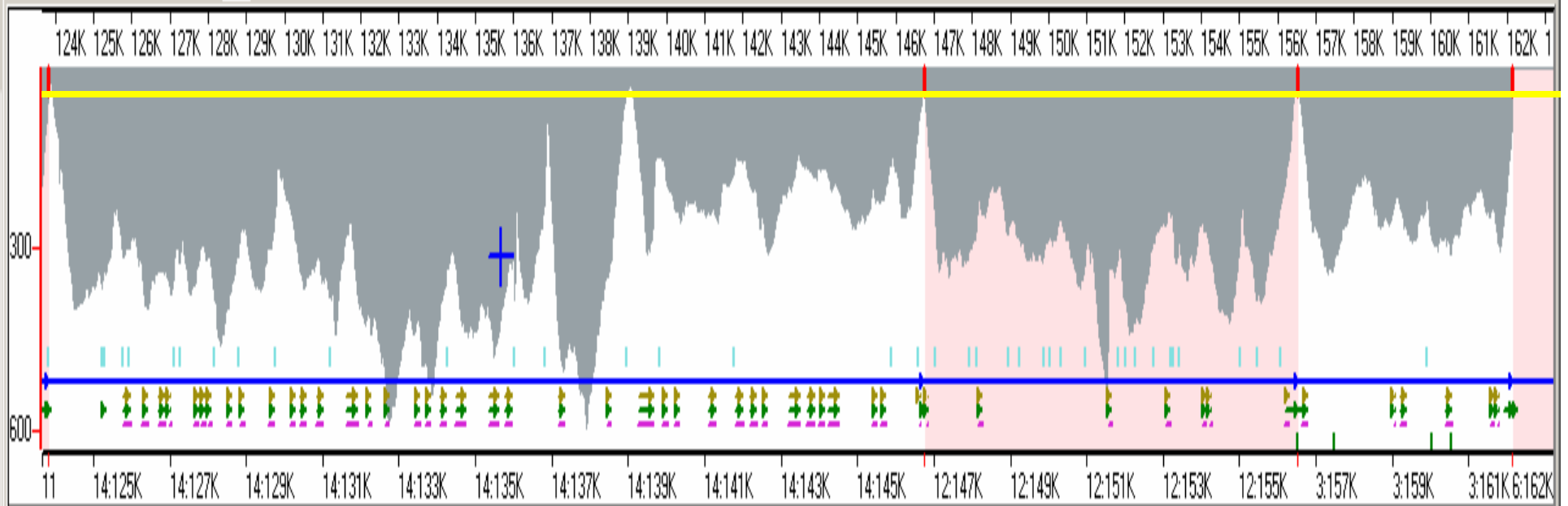
- ~1/3 primer sets include a 5' Amino Modifier C6 (NH<sub>2</sub>-blocked).\*

- Full capacity of 454 GS-FLX picotier plate is used; 15 samples are split between 2 gaskets.

- Flexibility of Long-range PCR approach enables genes to be easily added/removed.

\*Harismendy and Frazer (2010)  
*BioTechniques* 46:229-231

# Example of 454 data



*MYH7* -  
3 Amplicons

*MYL2* -  
2 Amplicons

*MYL3* -  
1 Amplicon

## Quality Checks

Day 1: Long Range PCR



Day 2: Quantitation of PCR products



Day 3: Pooling of PCR products



Day 4: Preparation of libraries



Day 5: Emulsion PCR



Day 6: Breaking emulsions



Day 7: Sequencing



Day 8: Analysis



Repeat PCR failures (1 Day)

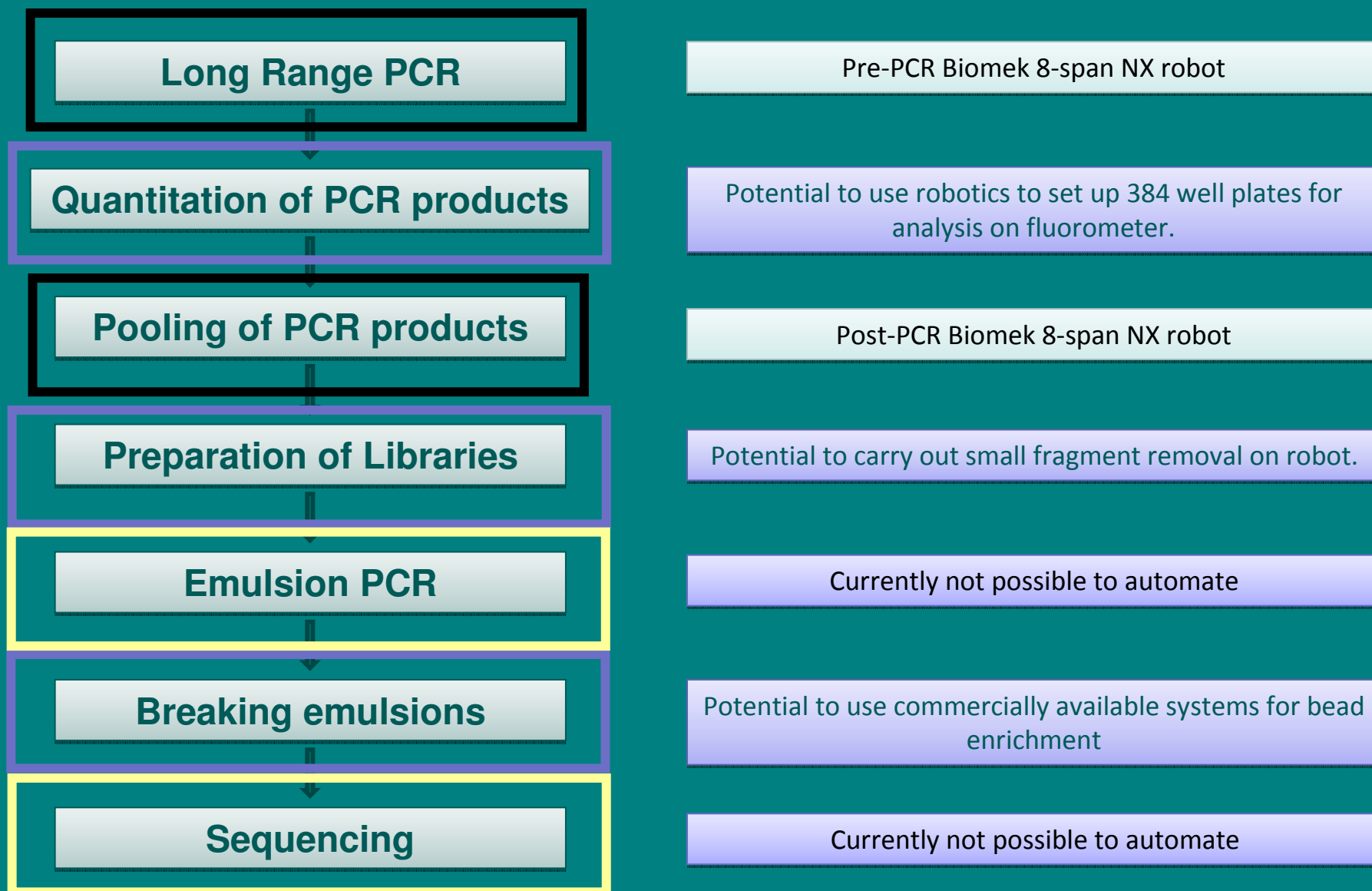


Repeat failed Library preparations (1 Day)



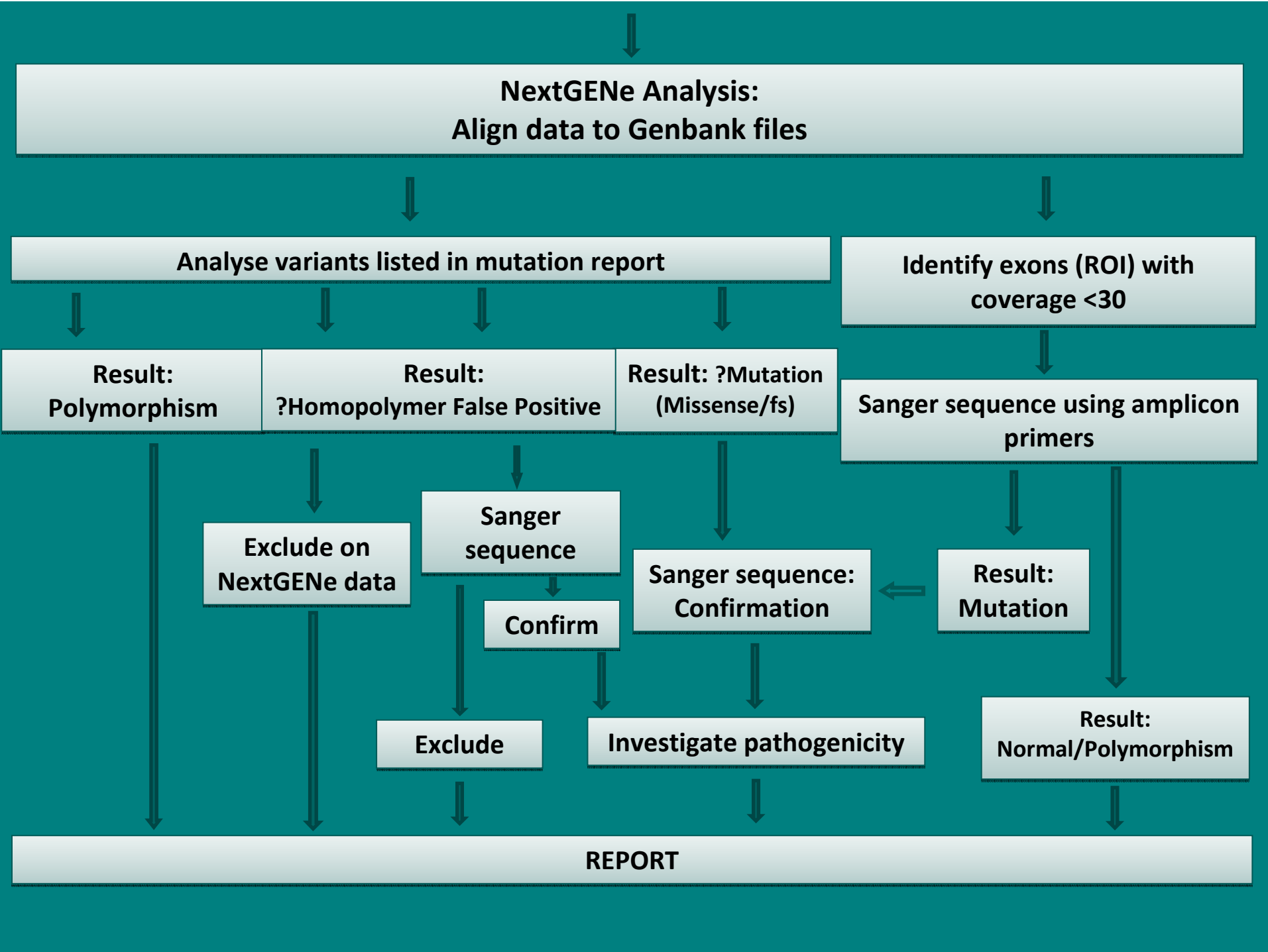
Repeat if failed enrichment (1 Day)

# Automation of 454 sequencing workflow



# Data Analysis

- Software for data analysis must be able to:
  - Accurately identify all true variants whilst limiting the number of false positives that require further investigation.
  - Accurately identify all exons/regions of interest where coverage is lower than minimum threshold.
- 454 software was not suitable for diagnostic laboratory for several reasons;
  - difficulties detecting frameshifts
  - unable to directly show regions below minimum coverage threshold
  - unable to directly give HGVS nomenclature for variants
- NextGENe software provides user-friendly data analysis;
  - No variants have been missed to date using the software (where coverage is >30)
  - Software can quickly identify regions that are below minimum coverage threshold
  - Ability to see overview of complete data for individual patient is useful for development of new services and improvement of existing services



**NextGENe Analysis:  
Align data to Genbank files**

**Analyze variants listed in mutation report**

**Identify exons (ROI) with  
coverage <30**

**Result:  
Polymorphism**

**Result:  
?Homopolymer False Positive**

**Result: ?Mutation  
(Missense/fs)**

**Sanger sequence using amplicon  
primers**

**Exclude on  
NextGENe data**

**Sanger  
sequence**

**Sanger sequence:  
Confirmation**

**Result:  
Mutation**

**Confirm**

**Exclude**

**Investigate pathogenicity**

**Result:  
Normal/Polymorphism**

**REPORT**

# Data Analysis

- The most significant factor affecting the specificity of data generated by the 454 are false positives in homopolymer regions.
- Higher coverage reduces their impact. However, across the 13 gene screen each individual is likely to show 6-12 false positives in homopolymer regions.
- False positives are almost exclusively restricted to the same locations in all samples.
- False positives in homopolymer tracts have a characteristic pattern in comparison to true positives.
- Across the 13 genes in our screen only *MYBPC3* shows a significant number of frameshift mutations (34). Only 2 of these occur in homopolymer tracts. Both mutations were successfully detected by the NextGENe software in positive controls.

# IT

- New server required for data storage and to allow efficient data processing.
- Custom cluster installed which matched the system used by the BRC.
- Purchasing, installation and methodology for data back-up required collaboration between the Trust IT department, Bioinformatics support at the BRC and lab's Bioinformatics manager.
- 454 uses Linux based operating system (RHEL6); setting up communication between the labs Windows based computers and 454 required new software (SSH) and training in Linux commands for users.
- Currently using 32-bit systems (standard in-house PC) for analysis. May need to transfer to 64-bit system in the future.

# Cost

- Major investment required to purchase machine, peripheral equipment and analysis software.
- Reagents expensive in comparison to LightScanner and Sanger sequencing costs.
- Expense of reagents should be balanced by reduced staff costs required for analysis.
- Micro-costing of process is currently being undertaken by the Health Economics Research Centre at the University of Oxford – results will be reported at a later date.
- 13-gene screen test price will be £1000; reduction compared to current 4-gene screen price (£1334).

# Further work

- Further develop the HCM protocol design:
  - To increase test robustness; particularly of the PCR stage.
  - To reduce the excess reads required to meet the minimum coverage across all genes.
- Investigate further automation to reduce laboratory time.
- Develop Starlims workflow to improve batch and result management.
- Transfer further in-house services to the 454 using long-range PCR method.
- Investigate other methods of sample preparation for further services.

# Conclusion

- A next-generation sequencing protocol has been successfully translated for use in the Oxford Molecular Genetics Laboratory for hypertrophic cardiomyopathy.
- Implementation has required major financial and time investment.
- To date the anticipated cost savings have not been achieved; advances in sample preparation methods are needed.

# Acknowledgements

Oxford Biomedical Research Centre



Oxford Biomedical Research Centre

Dr J Taylor

Dr K Kaur

Dr J Broxholme

Oxford Molecular Genetics Laboratory

Dr A Seller

K Thomson

G Dawson

Department of Cardiovascular Medicine

Prof H Watkins

Dr L Ormondroyd

Department of Clinical Genetics

Dr E Blair