



Building a next-generation sequencing pipeline for diagnostics

Bhaskar S¹, O'Sullivan J², Daly S², Ramsden S¹, Black G² and Wallace A¹

¹ Genetic Medicine, St Mary's Hospital, Manchester
² NIHR Biomedical Research Centre, Manchester

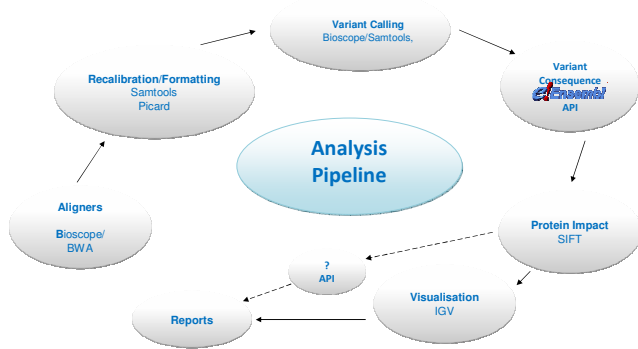
Introduction:

The advent of second generation sequencing has triggered a major effort to develop tools to manage and analyse large scale sequencing data. These new technologies have enabled many recent major genetic research discoveries. The scale and speed of these platforms mean genetic questions, previously impossible or too expensive to address, can now be investigated using next-generation sequencing. Unlike mature technologies like Sanger sequencing, next-generation sequencing is evolving at a rapid pace making diagnostic implementation challenging. Additionally, data generated by these platforms must conform to stringent QA/QC to permit their use for clinical diagnostics. A major area to address when implementing second generation sequencing technologies is handling, processing and tracking high volumes of data. We are currently using ABI's Bioscope amongst other software to build a bespoke informatics solution to address this need. Within the Regional Genetics Laboratories at St Mary's hospital, we work in close collaboration with research groups therefore effective delineation between research and diagnostic data must be maintained. We are currently engaged in research projects to investigate rare diseases with an unknown genetic basis using targeted or exome sequencing protocols on the ABI SOLiD 4 sequencing platform. This approach has uncovered disease causing mutations in previously unresolved cases. Although we have used this platform to screen for disease causing mutations in the research setting, we are developing these methods for use in the diagnostic setting.

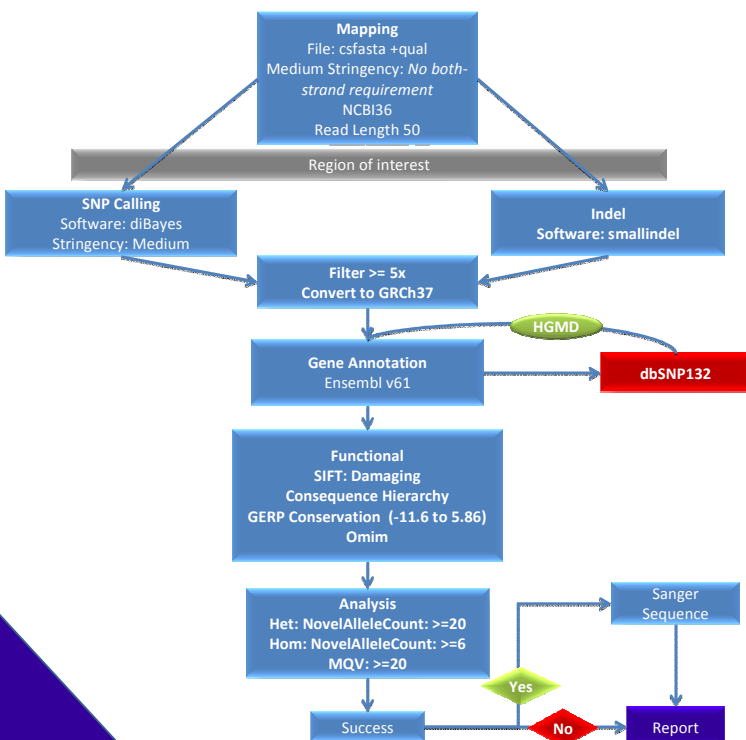
Aim of pipeline

To develop a fast accurate and robust system that provides:

- Efficiently processed next generation sequence data
- Fully traceable analysis process
- Modularised analysis system incorporating 3rd party software and resources
- Reports functionally relevant genetic variants
- Customisable pipeline which can be updated as user requirements change



Analysis workflow

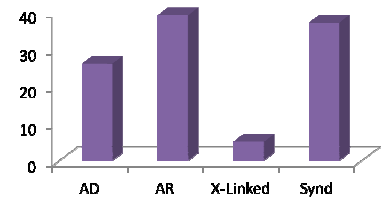


Case study – Retinal Dystrophy

Initially 10 samples were selected, 1 positive control sample and 9 patients with an Autosomal Dominant Retinitis Pigmentosa (ADRP) presentation. These patients had been through the screening programme for ADRP and provided by the NW Regional Genetics Lab which usually picks up mutations in approximately 50% of patients. A further 10 patients with an ADRP presentation were screened for the same panel of genes.

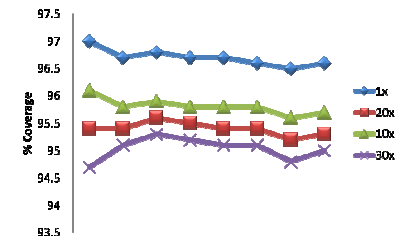
Array design

- 107 Genes known to be associated with Dominant, X-Linked, Recessive and Syndromic Retinal Dystrophy were selected for targeted enrichment using SureSelect from Agilent
- Gene classifications are summarised in the adjacent graph



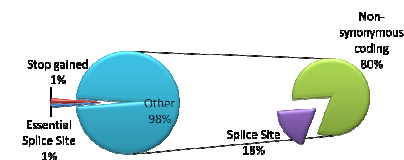
Sequencing summary statistics

- Average read depth at called SNPs was 831x
- Percentage of target regions covered at various read depths are presented in line graph



Functional consequence summary

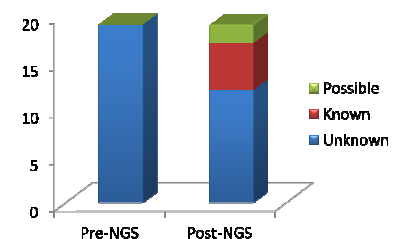
- Breakdown of high quality SNPs filtered to select those more likely to have a functional impact



Clinical impact

- 8 likely functional mutations were selected for confirmation by Sanger sequencing

• This pipeline allowed us to report disease causing mutations in 7 Retinal Dystrophy patients which would not have been possible using standard approaches. Graph show causative mutation status



Conclusions

- Using this approach we increase the pickup rate for the ADRP patients from 50% to 65%.
- With the recessive patient group we pick up 30% of disease causing mutations. This work provides us with a screening panel for recessive and simplex Retinal Dystrophy patients who previously didn't have access to genetic testing.