Array based targeting for massively parallel sequencing

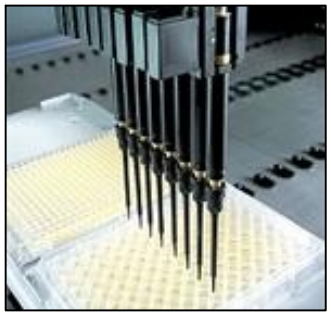
Patrick Tarpey
Team 78

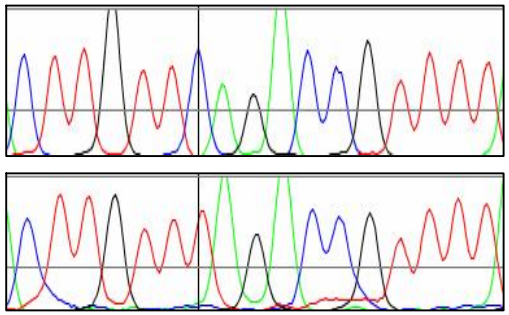
High Throughput Sequencing in the CGP

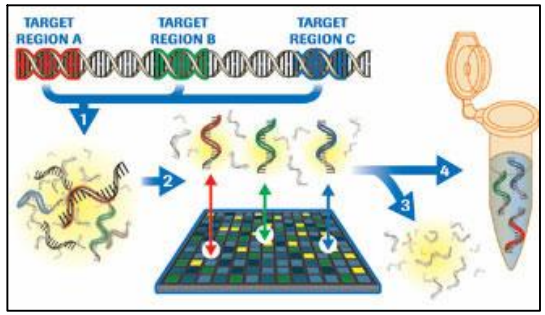
sample preparation

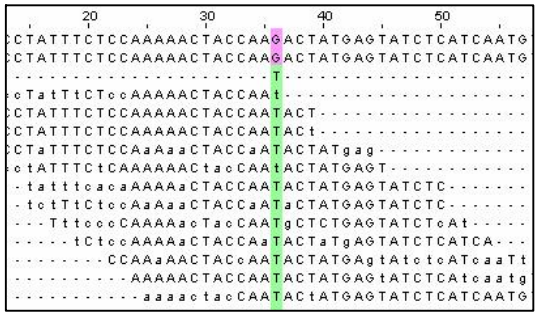
sequencing

data analysis









High Throughput Sequencing in the CGP

1 disease (XLMR)
200 samples
720 X chrom genes

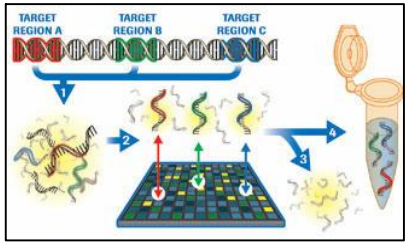
1858 mutations

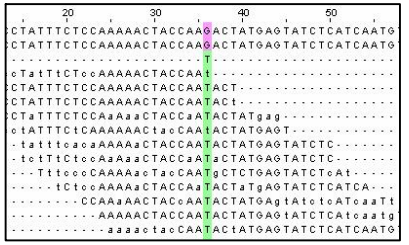
13 diseases
30 samples
920 genes

?

Overview

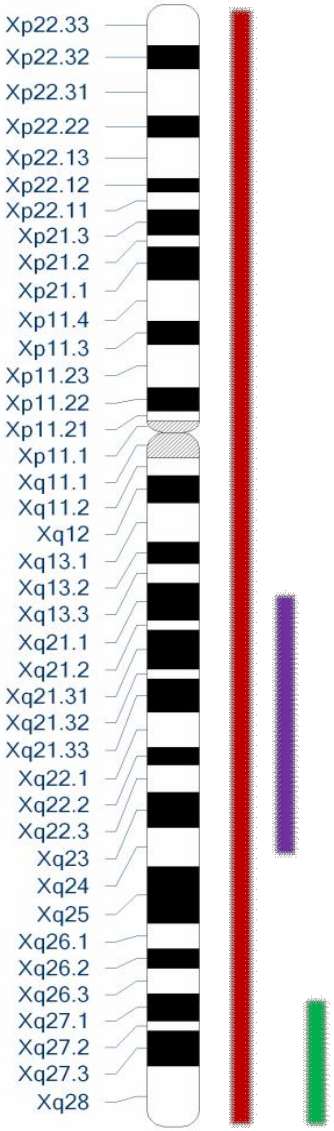
DNA samples





Data

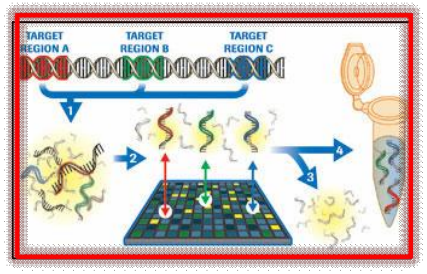
X-linked diseases under investigation

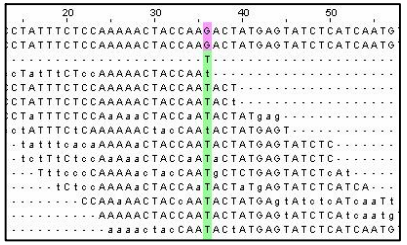


Aicardi	Arena	Lenz microphthalmia
•Affects females (two reported males both XXY) •X-linked dominant (male lethal) •No refined linkage interval •Agenesis of the corpus callosum •Infantile spasms •Lesions “lacunae” of the retina	•Affects males •X-linked recessive •Linked to Xq22-25 •Mental retardation •Spastic paraplegia	•Affects males •X-linked recessive •Linked to Xq27-28 •Microphthalmia •Additional malformations •Ears •Teeth •Fingers •Skeleton •Genitourinary system

Exon Capture on Nimblegen Arrays

DNA samples

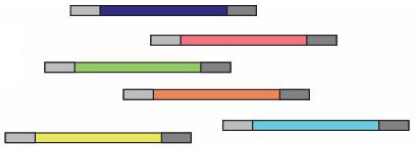




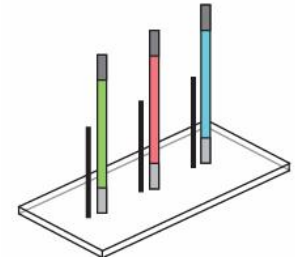
Data

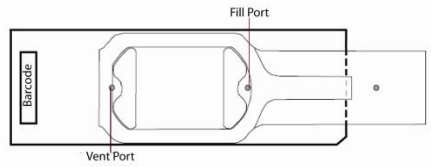
Exon Capture on Nimblegen Arrays

- Shear genomic DNA (~500bp)



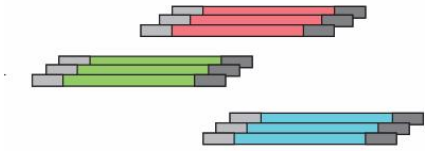
- End repair
- Ligate Solexa adaptors
- Size select (~200bp on gel)





8136 exons
1.5 Mb cds

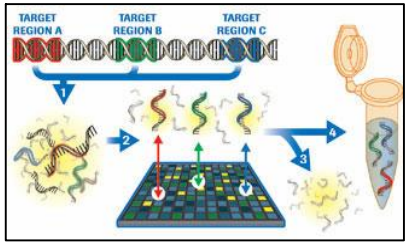
- Hybridise ~10ug adapted DNA
- Wash
- Elute

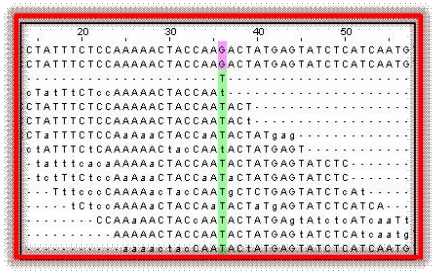


- PCR using Solexa primers
- Sequence

Data Analysis

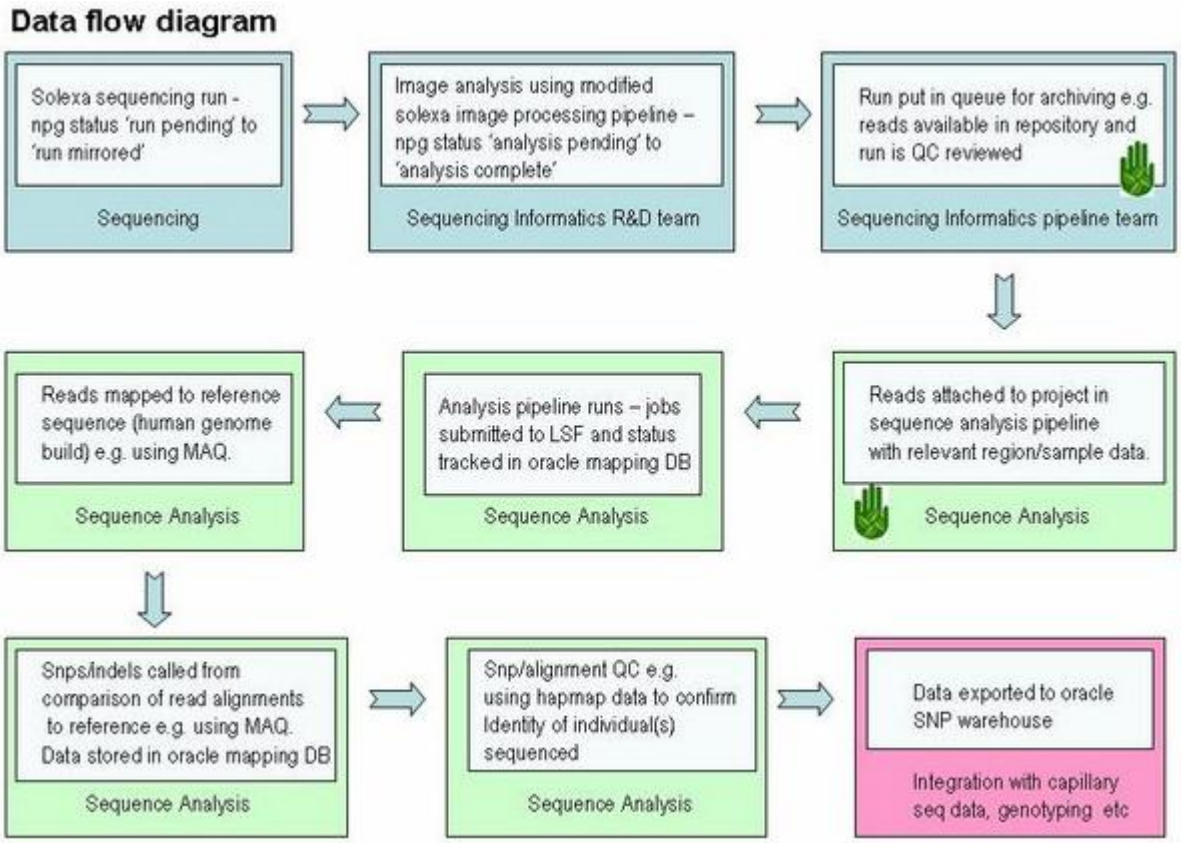
DNA samples



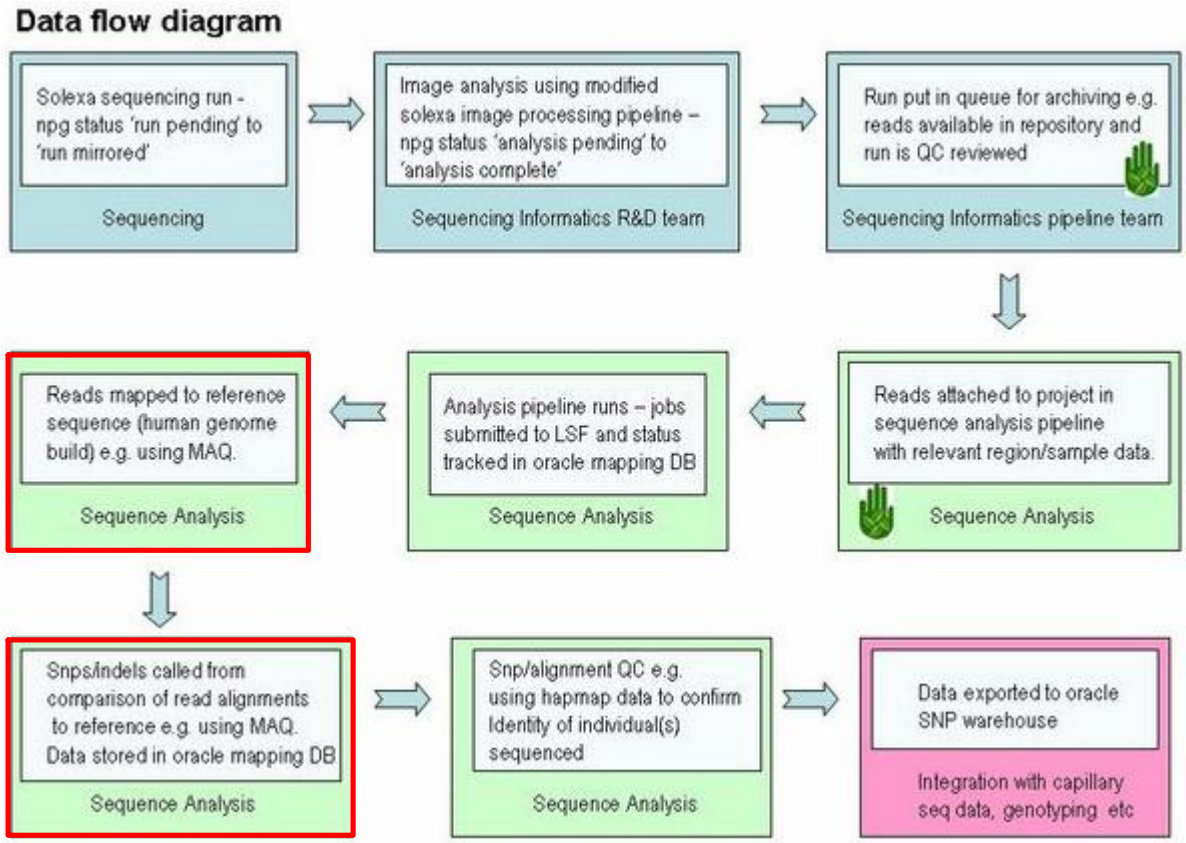


Data

Data Analysis



Data Analysis



Data Analysis

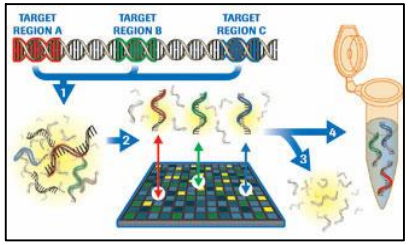
- **Annotated variant**
  nonsense, insdel, splice, missense, silent
- **Read depth**
  Each allele (zygosity)
- **Quality score**
  Each allele
- **Known/unknown (SNPdb)**
- **Flanking sequence window (blat)**

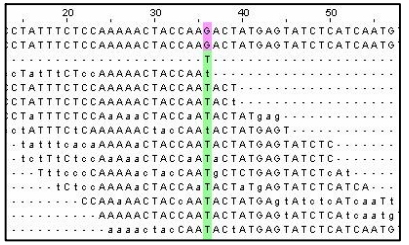
- **Variant ID > Webpage > Pileup Images**

- **Coverage & depth of reads in each gene in each sample**
  - **(%bases >10X depth)**

Results

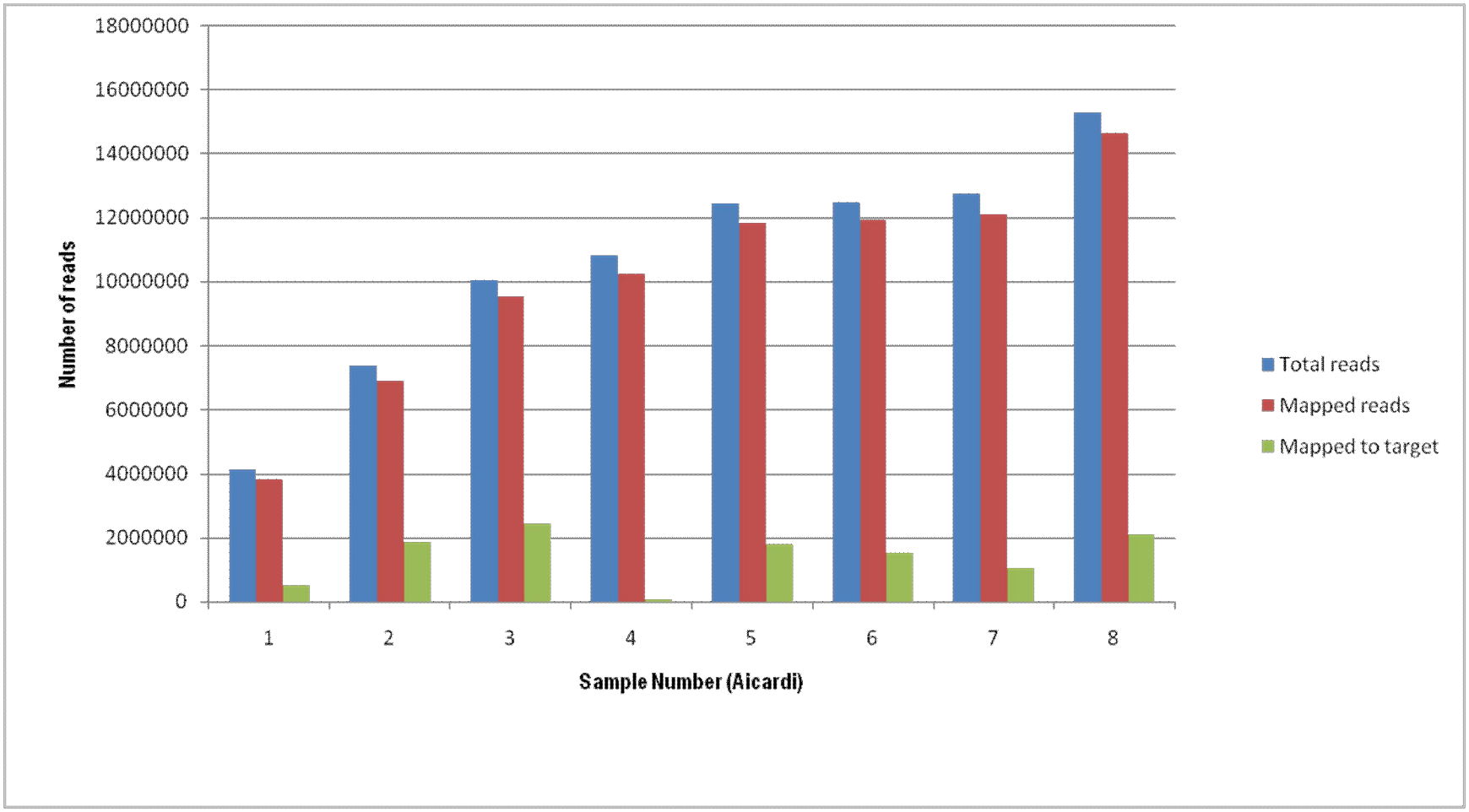
DNA samples



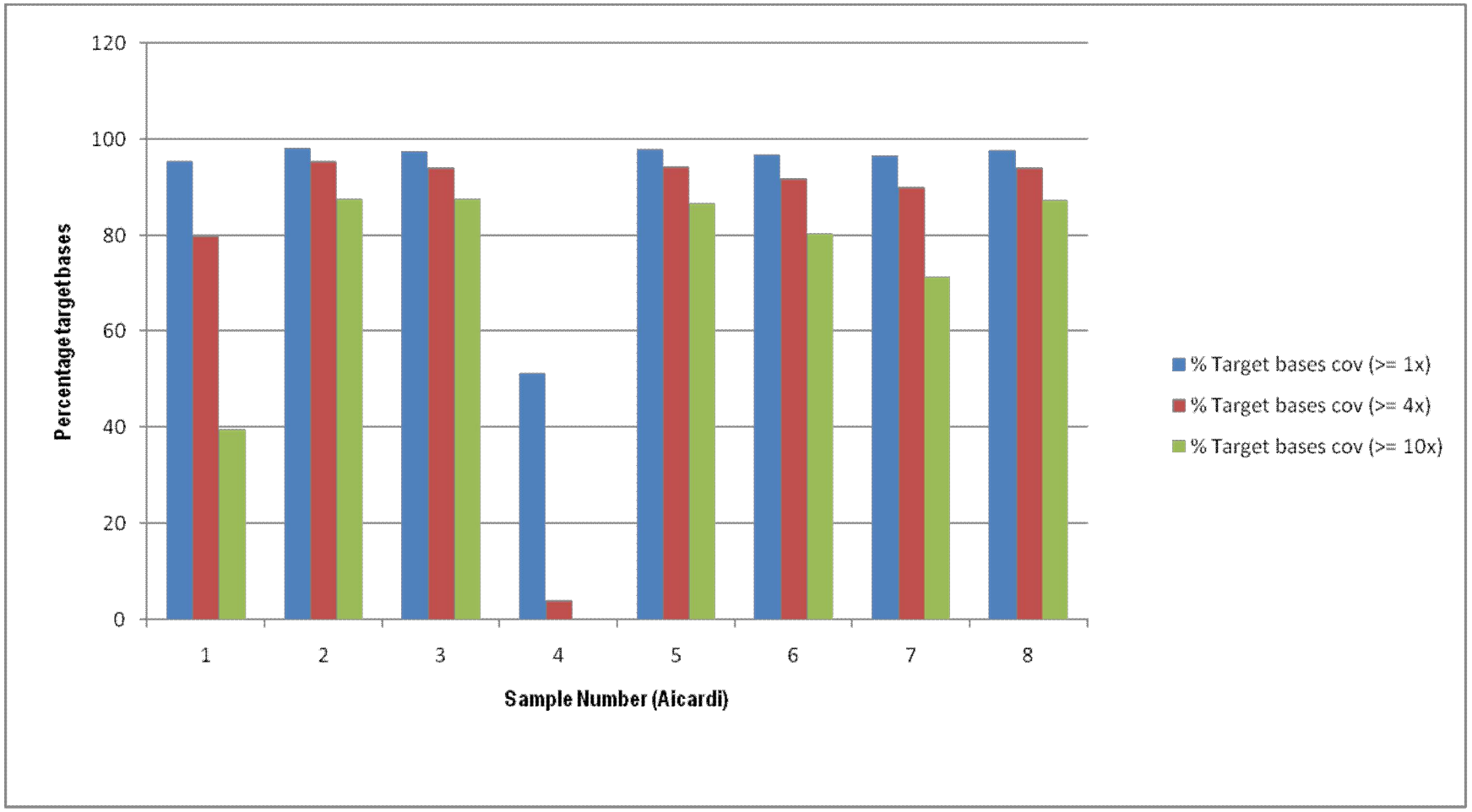


Data

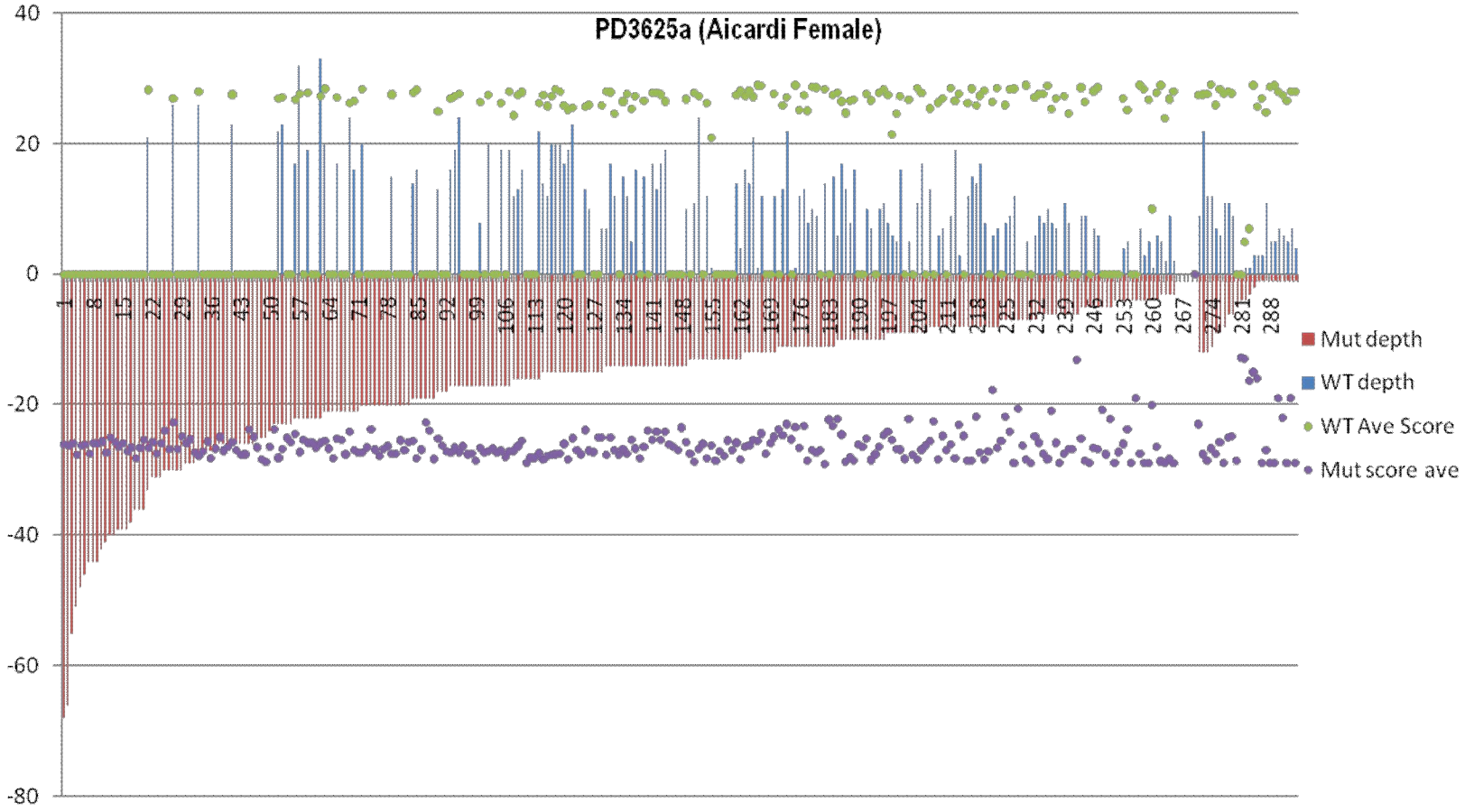
Exonic Capture Performance



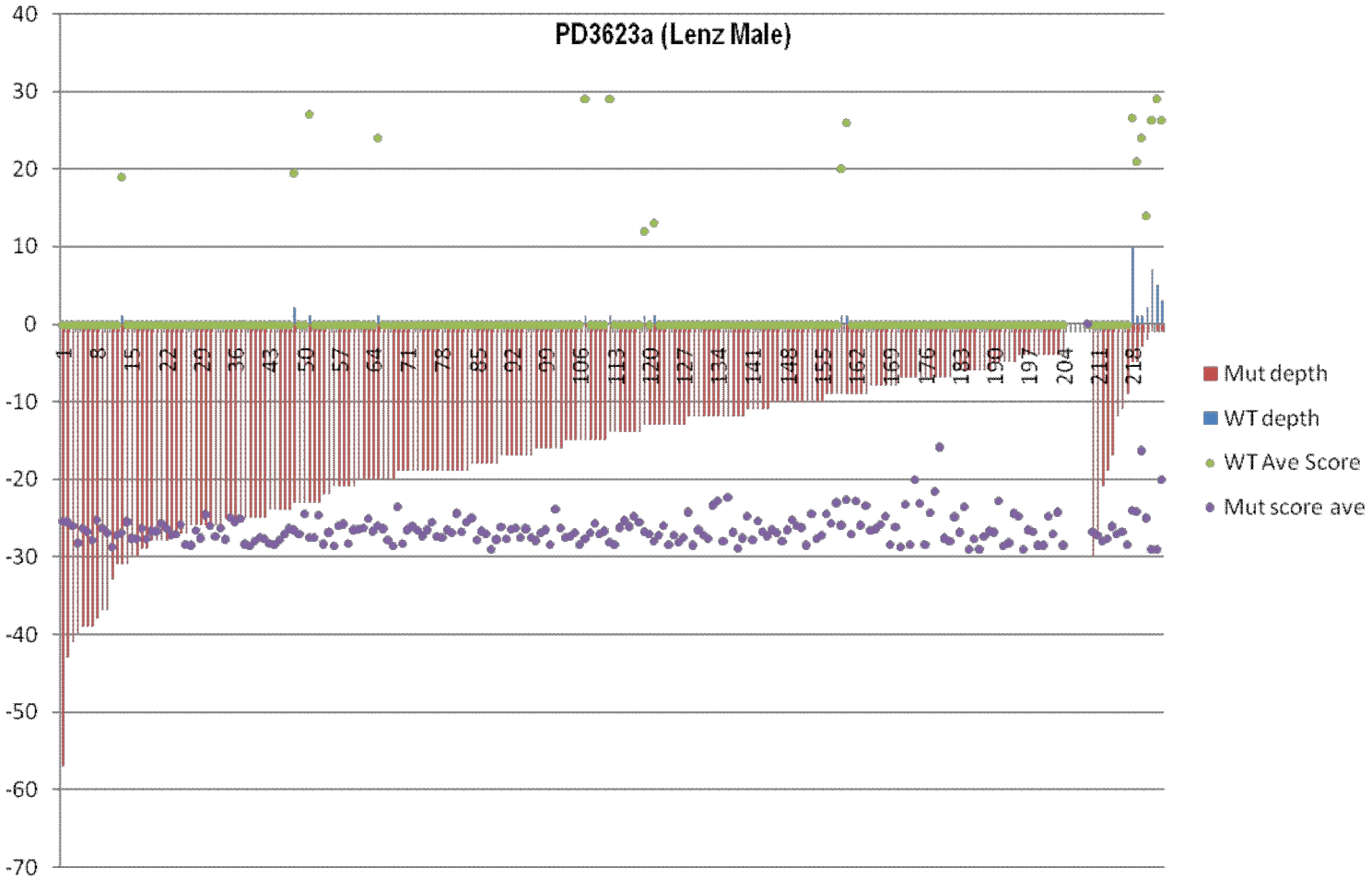
Exonic Capture Performance



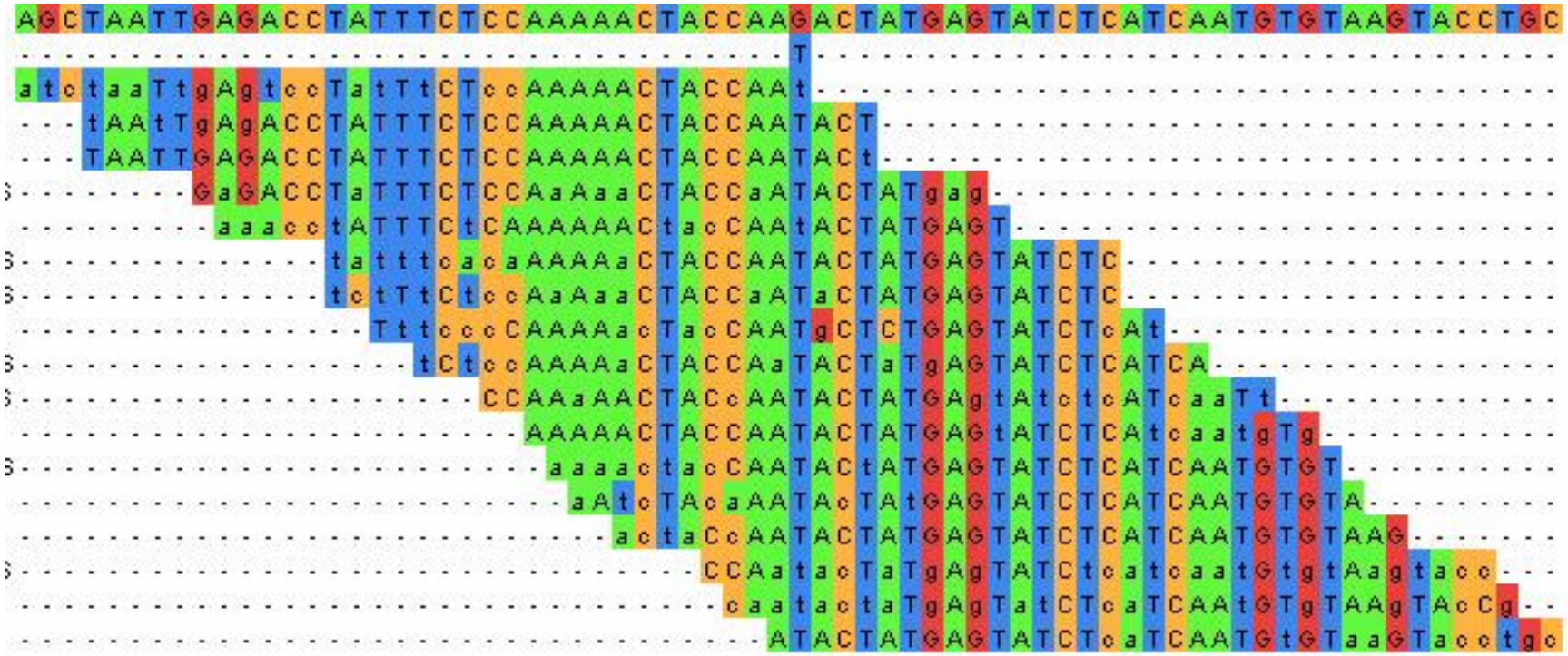
Substitutions (overview)



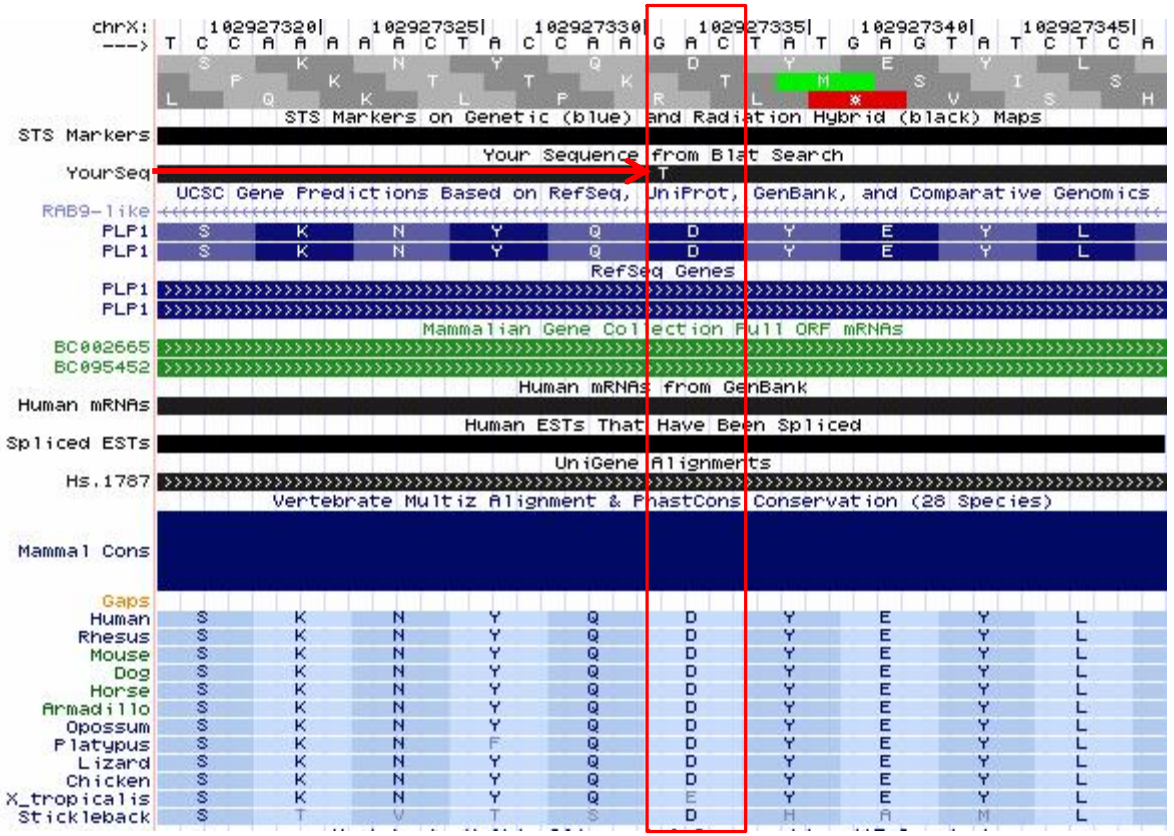
Substitutions (overview)



Arena (Missense)



Arena (Missense)



Arena (Missense)

#312920 GeneTests, Links

SPASTIC PARAPLEGIA 2, X-LINKED; SPG2

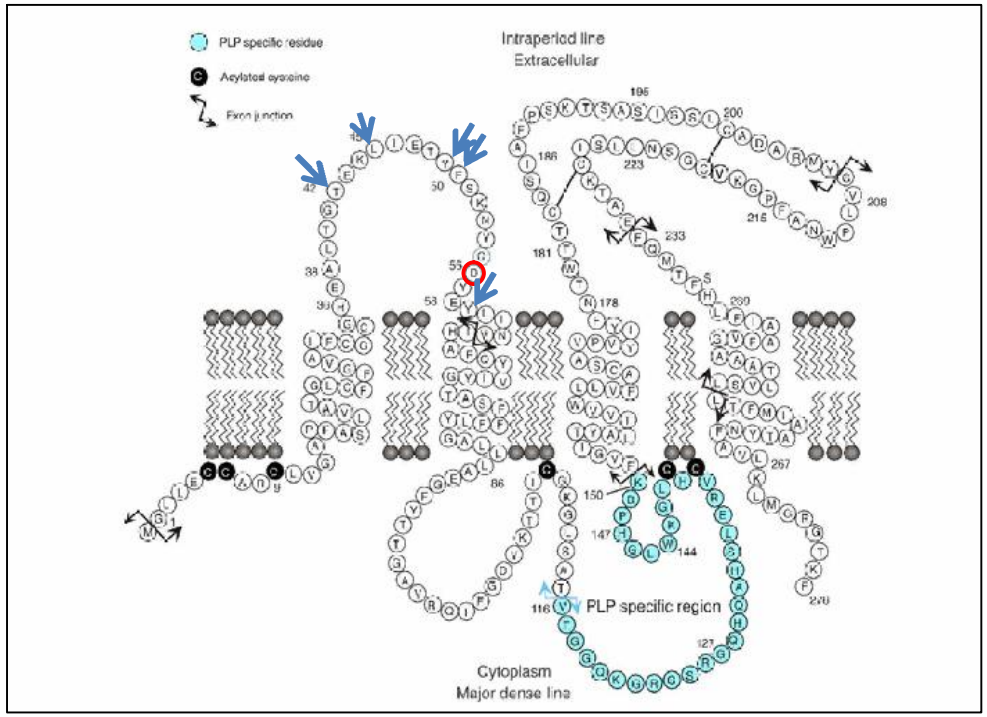
Alternative titles; symbols

SPPX2

Gene map locus Xq22

TEXT

A number sign (#) is used with this entry because spastic paraplegia-2 is caused by mutation in the myelin proteolipid protein gene (PLP1; 300401) and is therefore allelic to Pelizaeus-Merzbacher disease (PMD; 312080).



Summary

Protocol

- **Enrichment variable, however >80% target bases >10X coverage in most samples.**
- **Protocol under continued development, other options available soon (Agilent).**

Analysis

- **Known substitutions identified with high confidence.**
- **False positives relatively low, recognisable.**
- **Small insertions/deletions presently have a high false positive call rate.**

Compared to XLMR screen

- **Analysis easier (with bioinformatics support!)**
- **All data available simultaneously.**
- **Developmental delay good candidate disease to utilise this technology (heterogeneous).**

Team 78: Cancer Genome Project, XLMR Project

Mike Stratton, Andy Futreal
RaffeallaSmith

Bamford, Sally
Barthorpe, Andrew
Beare, David
Bhamra, Gurpreet
Bignell, Graham
Buck, Gemma
Butler, Adam
Campbell, Peter
Chen, Lina
Cole, Charlotte
Cole, Jennifer
Dalgleish, Gillian
Davies, Helen
Dawson, Elisabeth
Dicks, Ed
Edkins, Sarah
Erin Pleasence
Forbes, Simon
Greenman, Chris
Halliday, Kelly
Hardy, Claire
Hinton, Jonathan
Hunter, Chris
Jia, Mingming
Jones, David
Kok, Joyce
Lee, Rebecca
Leroy, Catherine
Maddison, Mark
Mahoney, Claire
McLaughlin, Wendy
Menzies, Andrew
Mironenko, Tatiana
O'Meara, Sarah
Pleasance, Erin
Santarius, Thomas
Shepherd, Rebecca
Stebbings, Lucy
Stephens, Philip
Teague, Jon
Tofts, Calli
Turner, Rachel
Van Haaften, Gijs
Webb, Jenny
West, Sofie
Widaa, Sara
Wray, Paul

Sequencing Technology R&D

Daniel Turner, Mike Quail: Ian Good head, Iwanka Kozarewa, Lira Mamanova, Elizabeth Sheridan

Variation Informatics

Carol Scott

Contribution of Samples

Lucy Raymond, Jozef Gecz, Charles Schwartz, Nazneen Rahman, Graeme Black, Leslie Biesecker

wellcome trust sanger institute