



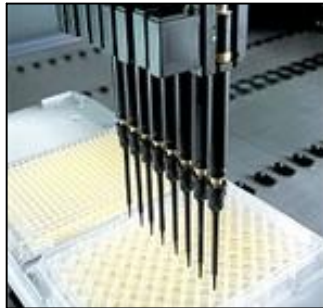
Array based targeting for massively parallel sequencing

Patrick Tarpey

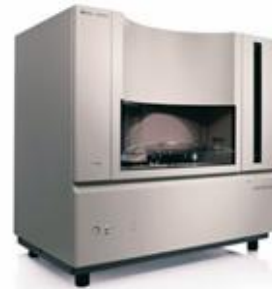
Team 78

High Throughput Sequencing in the CGP

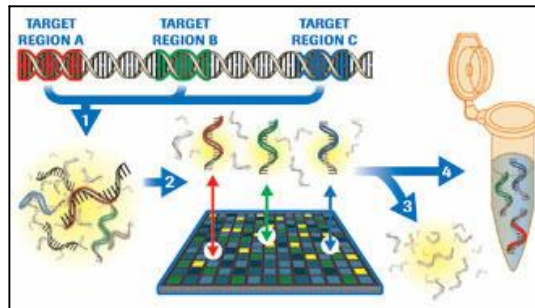
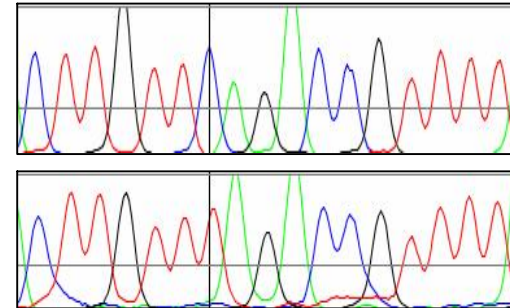
sample preparation



sequencing



data analysis

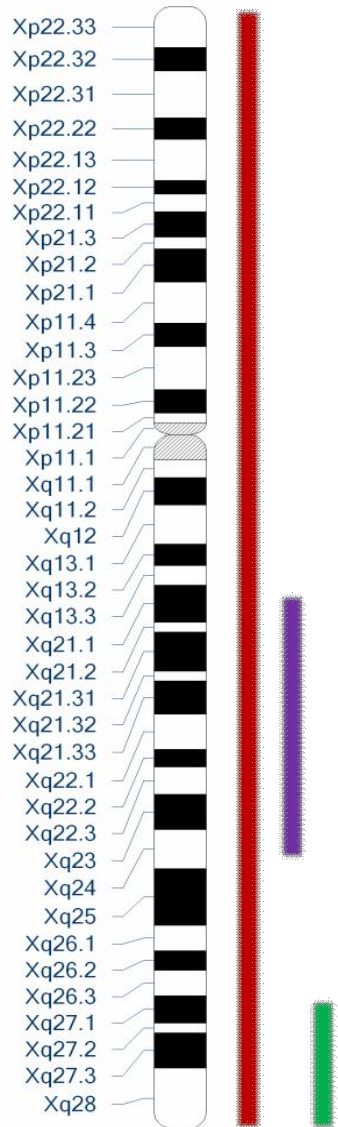


```
      20      30      40      50
CCTATTTCTCCAAAACTACCAAGACTATGAATATCTCATCAATG
CCTATTTCTCCAAAACTACCAAGACTATGAATATCTCATCAATG
.....
CTATTCTCCAAAACTACCAATACT.....
CCTATTCTCCAAAACTACCAATACT.....
CCTATTCTCCAAAACTACCAATACTgag.....
CTATTCTCCAAAACTACCAATACTATGAAT.....
tatttcaaaAAAACTACCAATACTATGAATATCTC.....
totTtCtccAaaAaaCTACCAATaCTATGAATATCTC.....
...TttcccAAAAaTaccAAATgCTCTGAGTATCTcAt.....
...tCtccAAAAaCTACCAaTACTaTgAGTATCTCATCA...
.....CCAAAaAACTACcAAATACTATGAgTatctcAaTt
.....AAAACTACCAATACTATGAATATCTCAtaaatg
.....aaaactaacAAATACTATGAATATCTCATCAATG
```

High Throughput Sequencing in the CGP



X-linked diseases under investigation



Aicardi

- 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

Arena

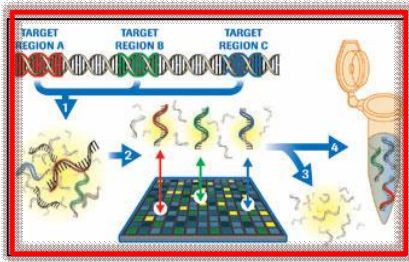
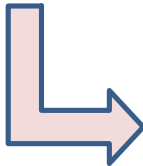
- Affects males
- X-linked recessive
- Linked to Xq22-25
- Mental retardation
- Spastic paraplegia

Lenz microphthalmia

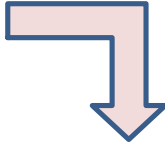
- Affects males
- X-linked recessive
- Linked to Xq27-28
- Microphthalmia
- Additional malformations
 - Ears
 - Teeth
 - Fingers
 - Skeleton
 - Genitourinary system

Exon Capture on Nimblegen Arrays

DNA samples

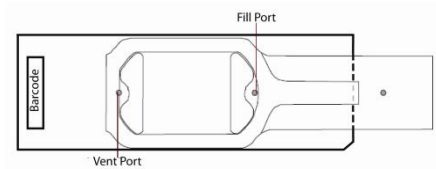
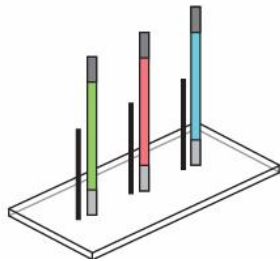
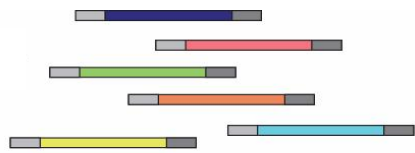


```
      20      30      40      50  
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG  
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG  
gTaTTTCTcAAAACTACCAAA.....  
CTATTTCTCCAAAACTACCAACT.....  
CTATTTCTCCAAAACTACCAACT.....  
CTATTTCTCCAAAACTACCAACTACT.....  
CTATTTCTCCAAAACTACCAACTACTATgag.....  
ctATTTCTCaaaaacTaocAAACTATGAGT.....  
-tatttoaaaaaCTACCAACTACTATGAGTATCTC.....  
-toTTCTcAAAACTACCAATACTATGAGTATCTC.....  
...TttcCAAAAoTaoCAATgCTCTGAGTATCToA.....  
...TCTcAAAACTACCAACTACTaTgAGTATCTCATA.....  
...CCAAaACTACAAACTACTATGAgTAtotoAToaaTT.....  
...AAAACTACCAACTACTATGAGTATCTCAToaaTg.....  
...aazocTAACAATACTATGAGTATCTCATCAATG
```

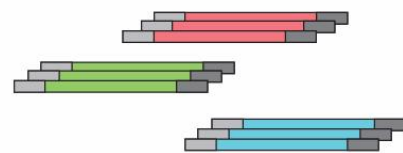


Data

Exon Capture on Nimblegen Arrays



8136 exons
1.5 Mb cds



- Shear genomic DNA (~500bp)



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



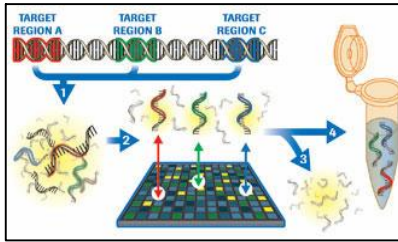
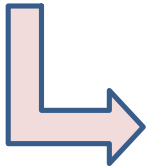
- Hybridise ~10ug adapted DNA
- Wash
- Elute



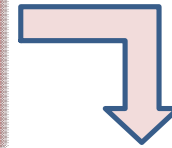
- PCR using Solexa primers
- Sequence

Data Analysis

DNA samples



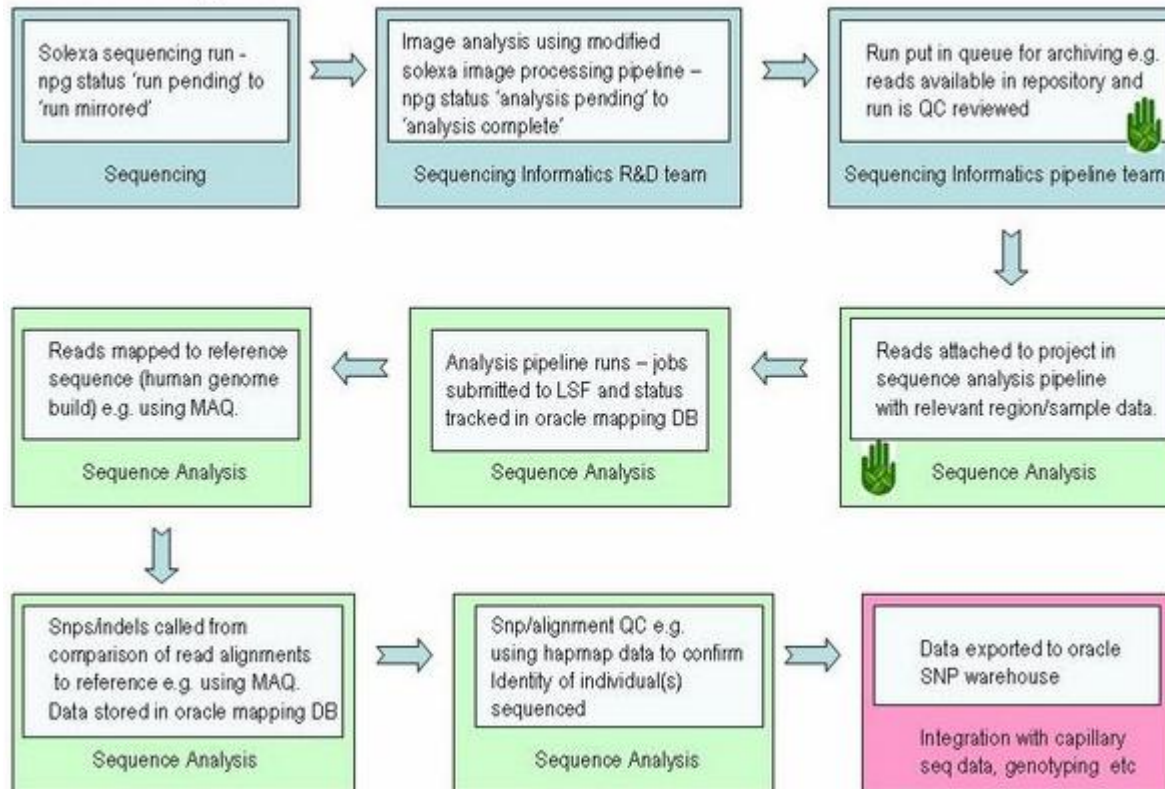
```
20 30 40 50
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
TATTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
CTATTTCTCCAAAACTACCAAGACTATGAGTATCTCATCAATG
```



Data

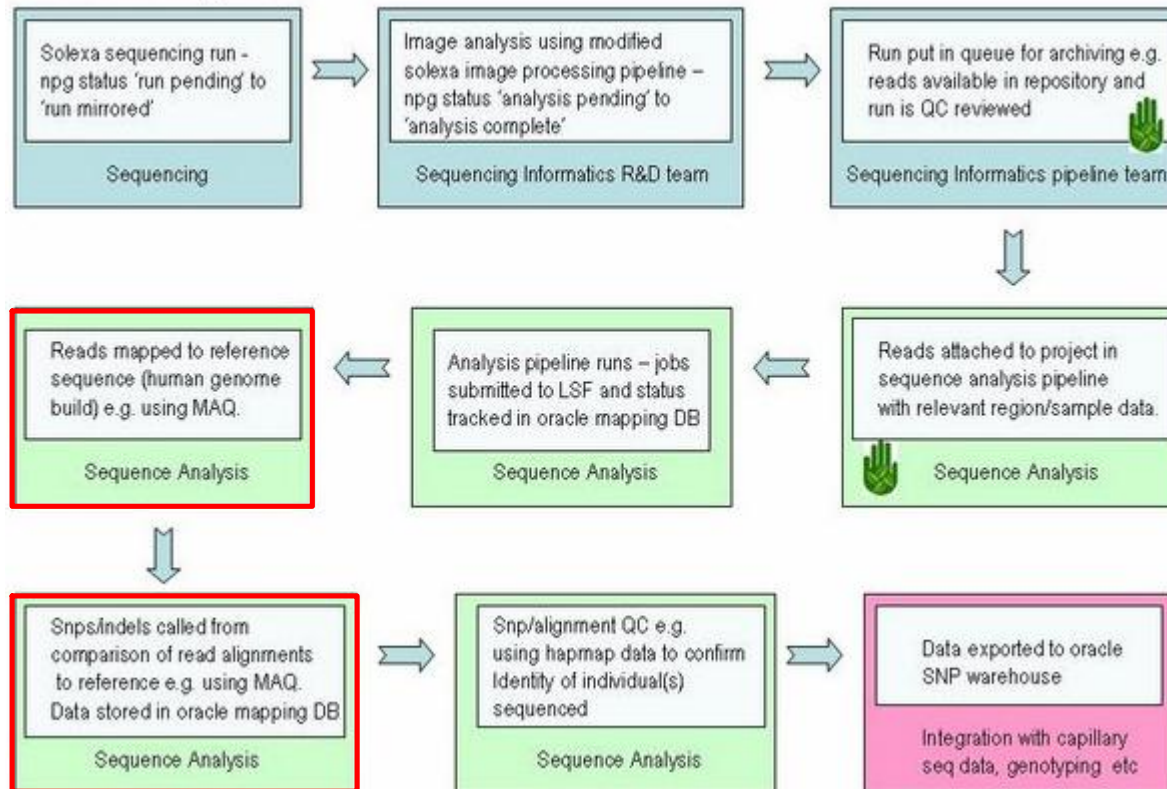
Data Analysis

Data flow diagram



Data Analysis

Data flow diagram



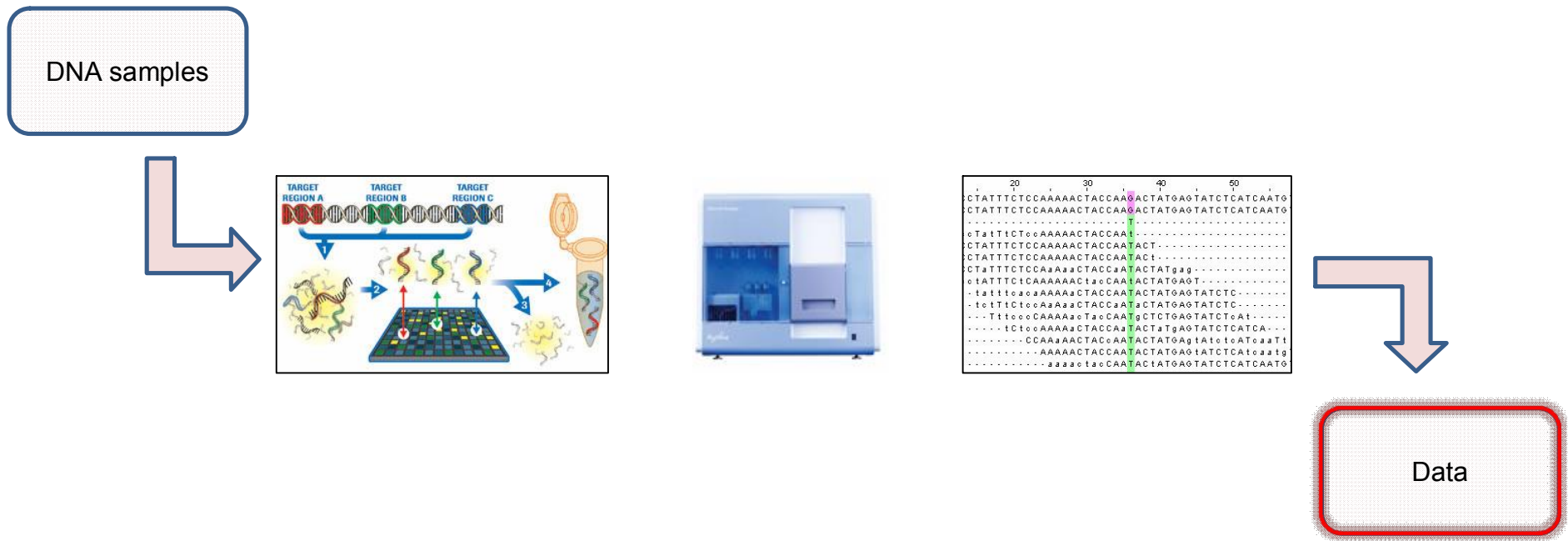
Data Analysis

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

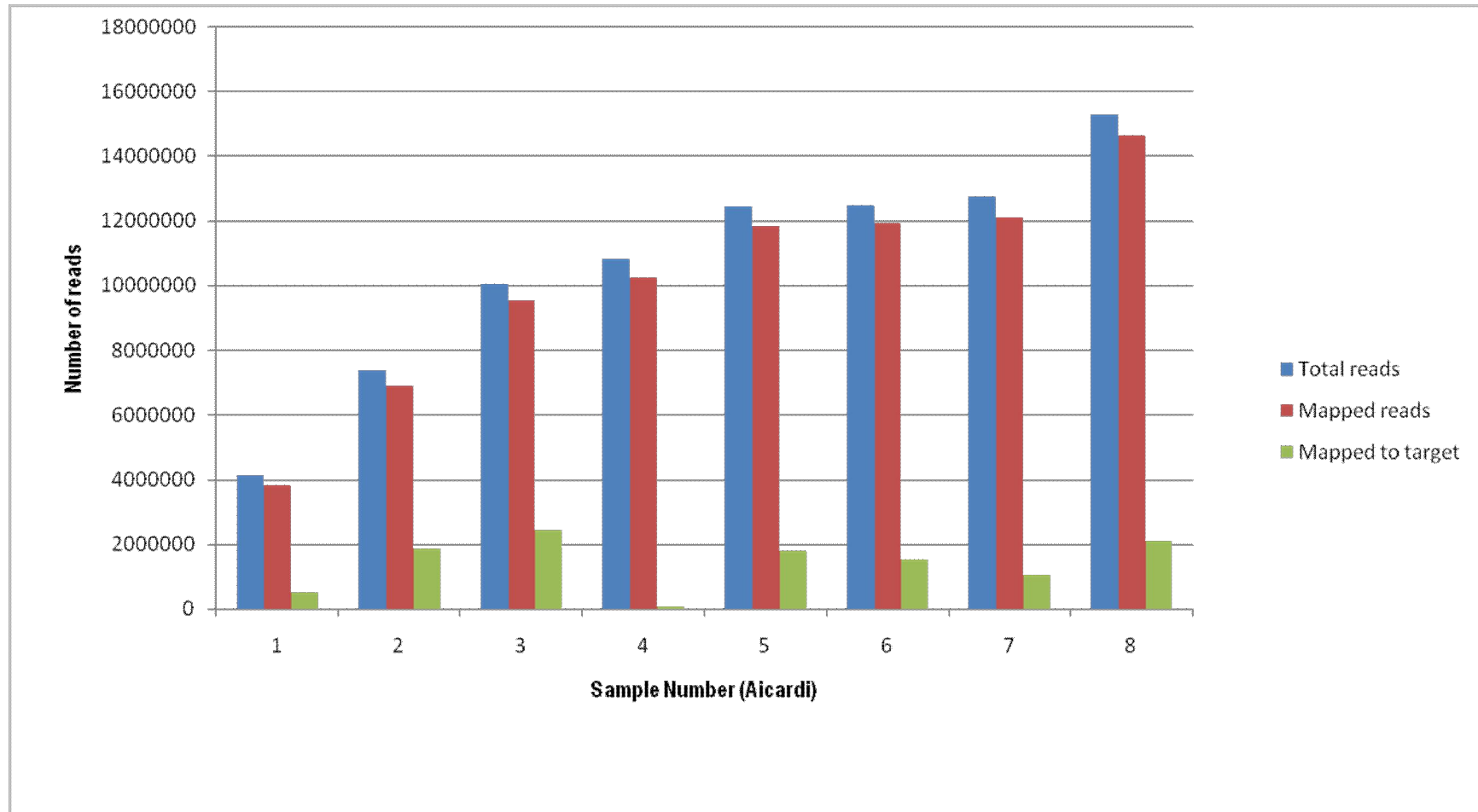
- **Variant ID > Webpage > Pileup Images**

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

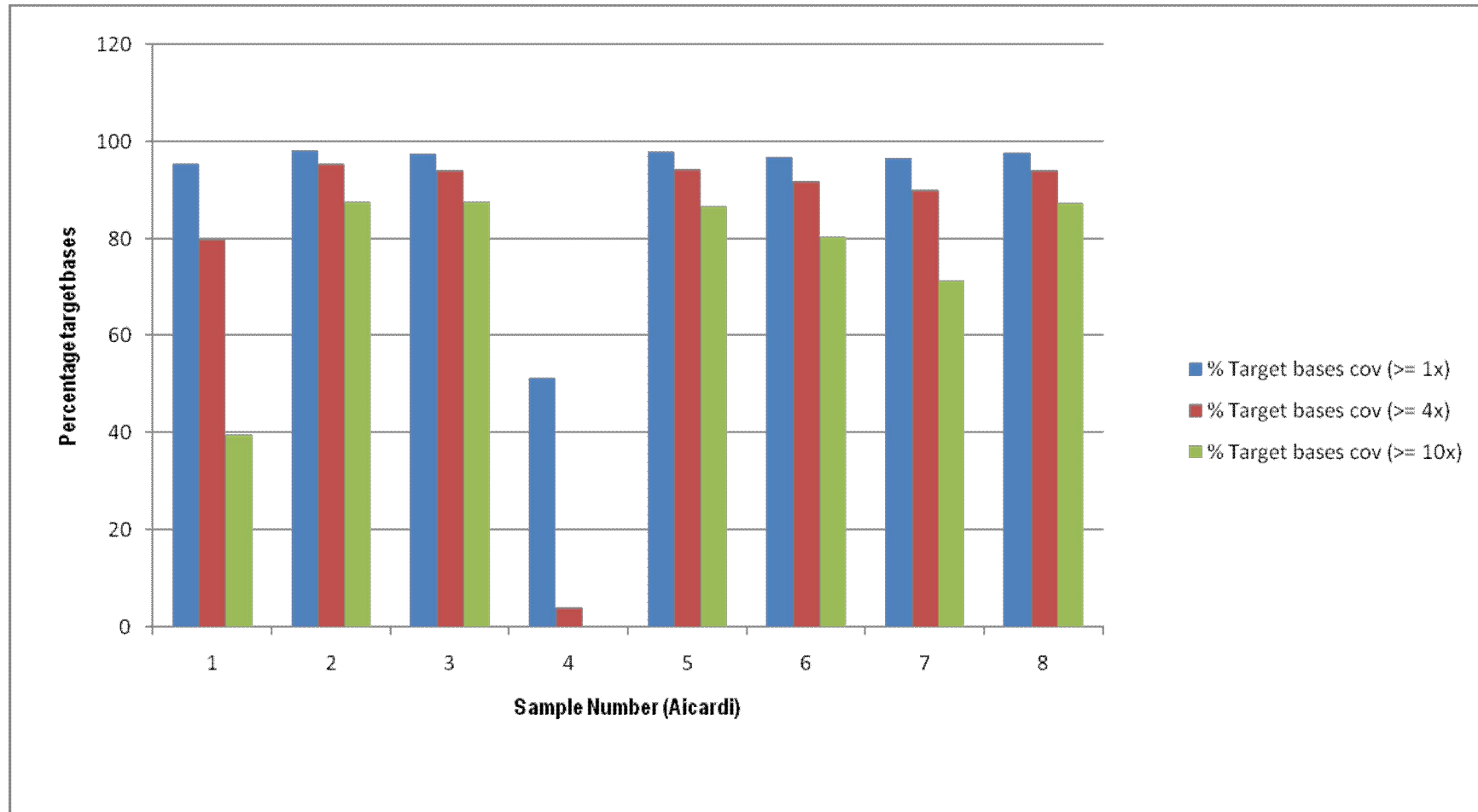
Results



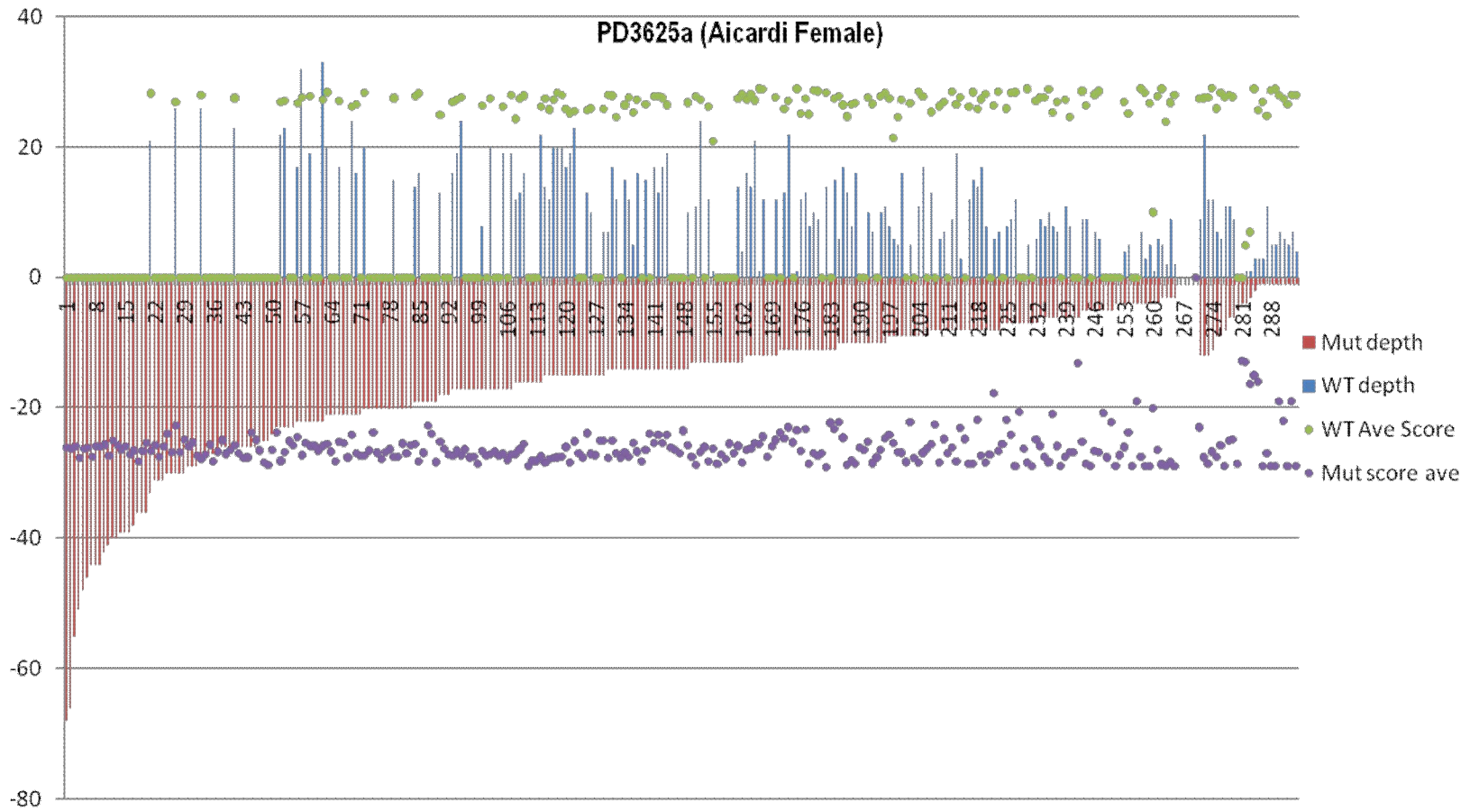
Exonic Capture Performance



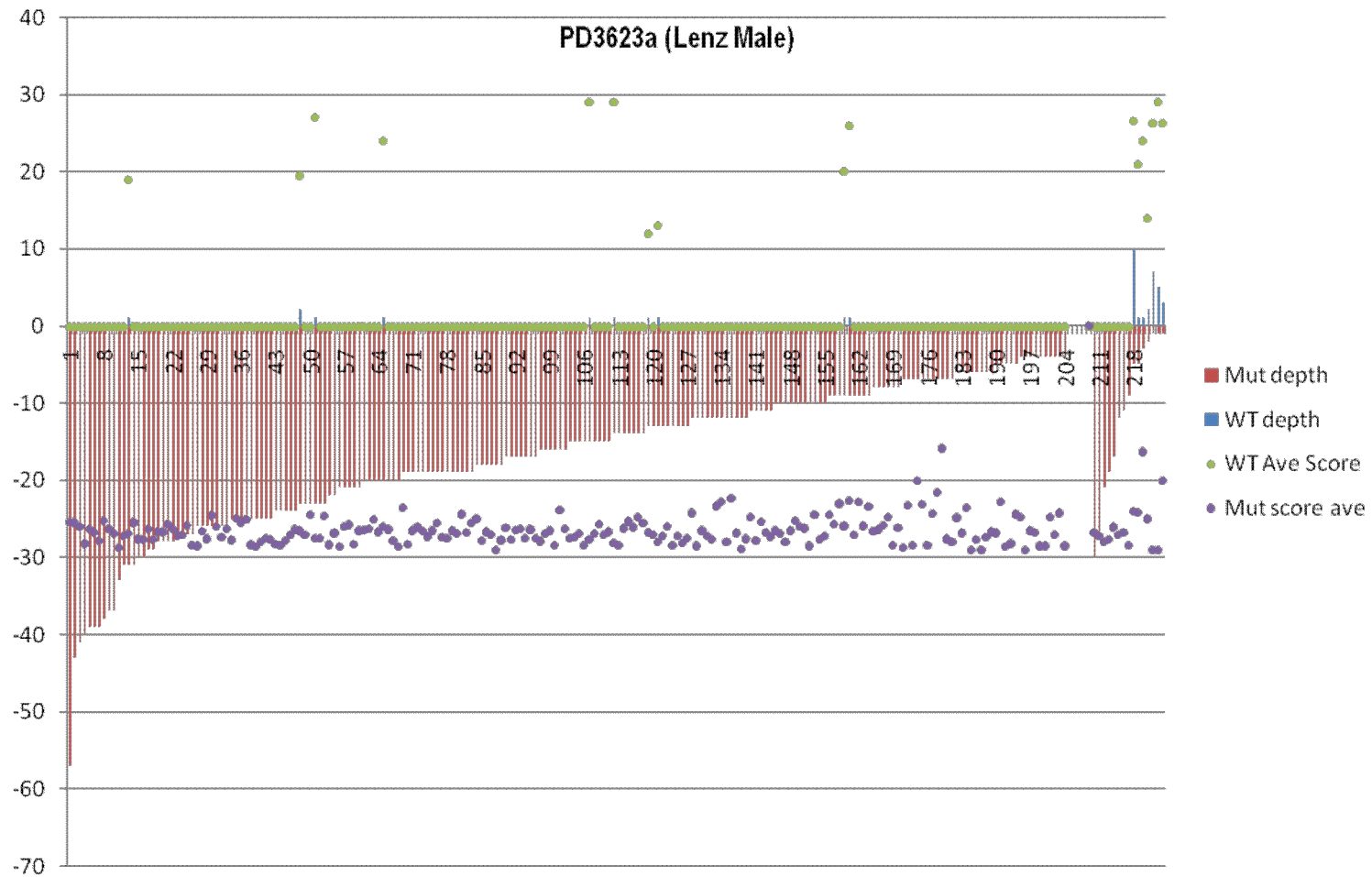
Exonic Capture Performance



Substitutions (overview)



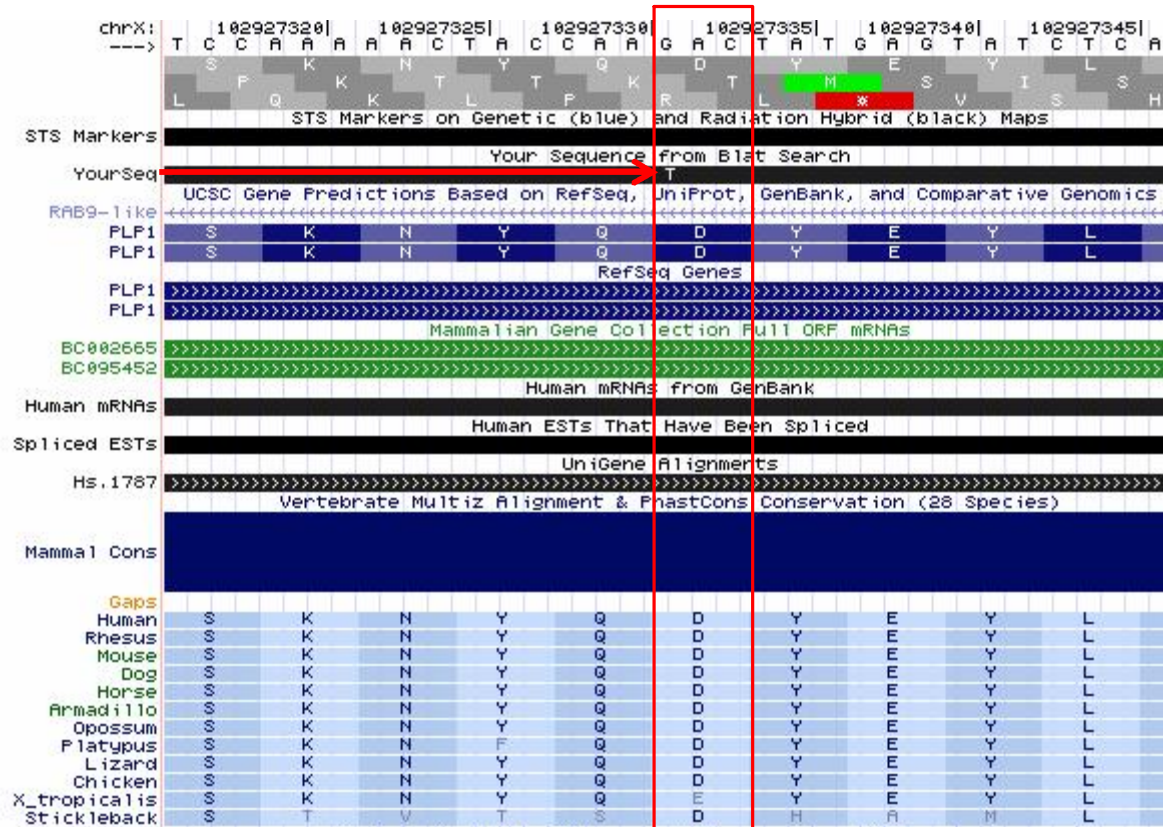
Substitutions (overview)



Arena (Missense)

```
AGCTAATTGAGACCTATTTCTCCAAAAACTACCAAGACTATGAGTATCTCATCAATGTGTAAGTACCTGC
. . . . . T . . . . .
a t c t a a T t g A g t c c T a t t t C t c c A A A A A C T A C C A A t . . . . .
. . . t A A t T g A g A C C T A T T T C T C C A A A A A C T A C C A A T A C T . . . . .
. . . T A A T T G A G A C C T A T T T C T C C A A A A A C T A C C A A T A C T . . . . .
3 . . . . . G a G A C C T a T T T C T C C A a A a a C T A C C a A T A C T A T g a g . . . . .
. . . . . a a a c c t A T T T C t C A A A A A A C t a c C A A t A C T A T G A G T . . . . .
3 . . . . . t a t t t c a c a A A A A a C T A C C A A T A C T A T G A G T A T C T C . . . . .
3 . . . . . t c t T t C t c c A a A a a C T A C C a A T a C T A T G A G T A T C T C . . . . .
. . . . . T t t c c c C A A A A a c T a c C A A T g C T C T G A G T A T C T c A t . . . . .
3 . . . . . t C t c c A A A A a C T A C C A a T A C T a T g A G T A T C T C A T C A . . . . .
3 . . . . . C C A A a A A C T A C c A A T A C T A T G A g t A t c t c A T c a a T t . . . . .
. . . . . A A A A A C T A C C A A T A C T A T G A G t A T C T C A t c a a t g T g . . . . .
3 . . . . . a a a a c t a c C A A T A C t A T G A G T A T C T C A T C A A T G T G T . . . . .
. . . . . a A t c T A c a A A T A c T a t G A G T A T C T C A T C A A T G T G T A . . . . .
. . . . . a c t a C c A A T A C T A T G A G T A T C T C A T C A A T G T G T A A G . . . . .
3 . . . . . C C A a t a c T a T g A g T A T C t c a t c a a t g t g t A a g t a c c . . . . .
. . . . . c a a t a c t a T g A g T a t C t c a T C A A t G T g T A A g T a c C g . . . . .
. . . . . A T A C T A T G A G T A T C t c a T C A A T G t G T a a G T a c c t g c . . . . .
```

Arena (Missense)



Arena (Missense)

#312920

SPASTIC PARAPLEGIA 2, X-LINKED; SPG2

GeneTests, Links

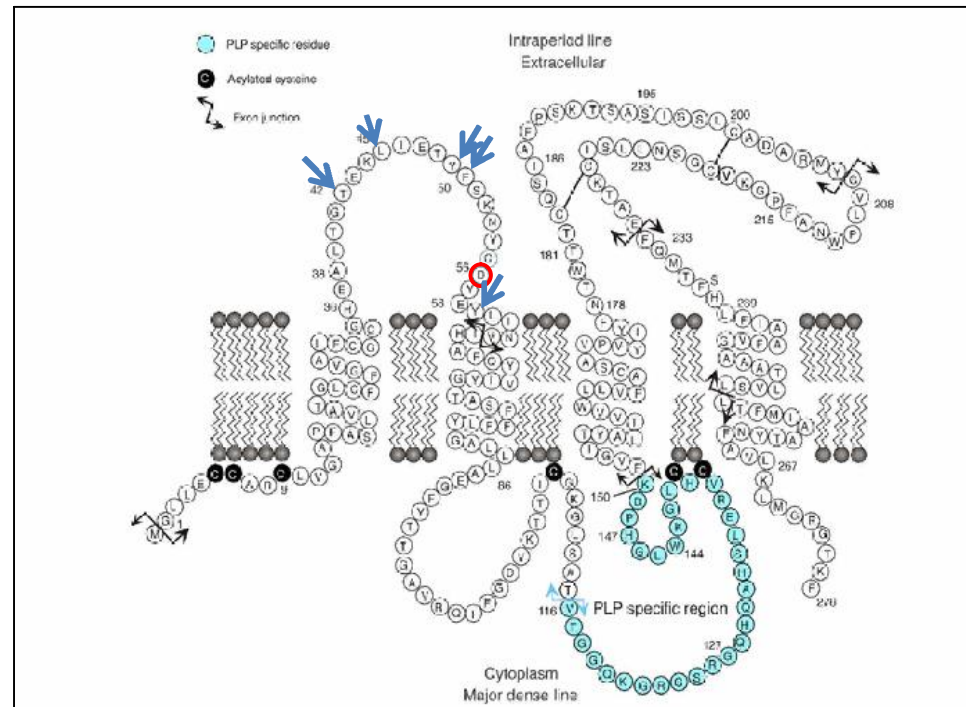
Alternative titles; symbols

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

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