

# COSMIC GBrowse

## Visualising cancer mutations in genomic context

Dave Beare

[dmr@sanger.ac.uk](mailto:dmr@sanger.ac.uk)

Cancer Genome Project  
Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

# Introduction

- 2000: Cancer Genome Project (CGP)
- 2004: Catalogue Of Somatic Mutations In Cancer - COSMIC  
Oracle database and website  
<http://www.sanger.ac.uk/genetics/CGP/cosmic>

Sources of mutation data

1. Literature (curators)
2. Other database(s) eg TP53 (IARC)  
International Agency for Research on Cancer
3. Sequencing/mutation detection

- 2010: COSMIC GBrowse (22nd September??)  
<http://www.sanger.ac.uk/fgb2/gbrowse/cosmic> x

# GBrowse and CGP

- Q. How could we visualise the data deluge from next generation sequencing?
- A. Gbrowse

[Keiran Raine GMOD presentation in January 2010]  
A near instant solution to the problem (days/weeks, rather than months/years for an in house solution).
- Q. COSMIC was designed to be gene centric but what about sequencing whole cancer genomes and visualising mutations in genomic context?
- A. Gbrowse  
Again!

# GBrowse: Setup

- Hardware
  - 5 Virtual Machines [Debian Linux, 2G RAM) ]  
dev + master + renderfarm slaves (2) + PostgreSQL
- Software
  - apache 2.2.9
  - mod\_fastcgi 2.4.6
  - gbrowse 2.13 [perl 5.10.0 + bioperl 1.61 + bio::graphics 2.11]
- Databases
  - PostgreSQL
    - 2 databases: 'Reference' and 'Cosmic'
    - scripts to query/format/populate these databases

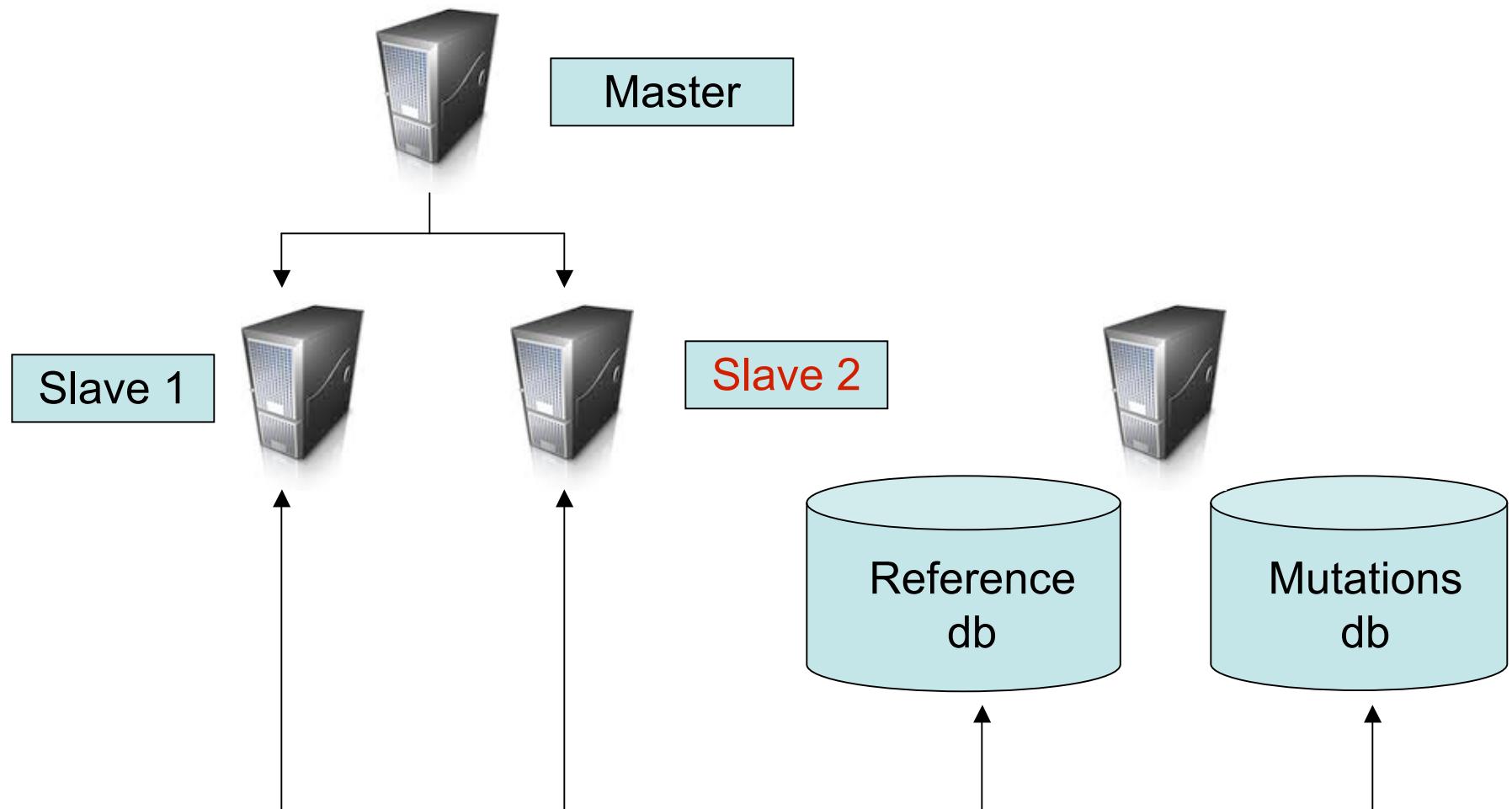
# GBrowse: Data

- Reference
  - Reference genome (GRCh37) + cytogenetic bands
  - Ensembl annotations (e! 58)
  - Cosmic Transcripts
- Cosmic
  - Mutations (substitutions, insertions/deletions)
  - Rearrangements
  - Copy Number Profiles
    - analysis of SNP6 microarray data over 800 cell lines
    - % samples which have copy number features (amplification, homozygous deletion, LOH, change)

# GBrowse: Configuration

- cosmic css/theme
- perl callbacks
  - glyphs
  - colours
  - hyperlinks
  - popups/tooltips
- renderfarm enabled

# GBrowse: Render Farm



# GBrowse: Select Tracks

<http://www.sanger.ac.uk/fgb2/gbrowse/cosmic>

**COSMIC** Catalogue of Somatic Mutations in Cancer

**COSMIC Cancer Genome Browser**  
Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK Tel:+44 (0)1223 834244

New to this site? Click [here](#) for help.

File Help

**Cosmic Release 48: 108.8 kbp from 10:89,622,870..89,731,687**

Human Genome [GRCh\_37 / Ensembl\_58]

Browser [Select Tracks](#) [Upload and Share Tracks](#) Preferences

<< Back to Browser

**Tracks**

- Overview  All on  All off  
 Chromosome
- Region  All on  All off  
 Genes
- 1. Genomic Features  All on  All off  
 Ensembl Protein Coding Genes  COSMIC Transcripts  non-coding RNA
- 2. Cosmic  All on  All off  
 Breakpoints  Amplification  LOH  
 Mutation density/details  Copy Number Change  Homozygous Deletion

[Back to Browser](#)

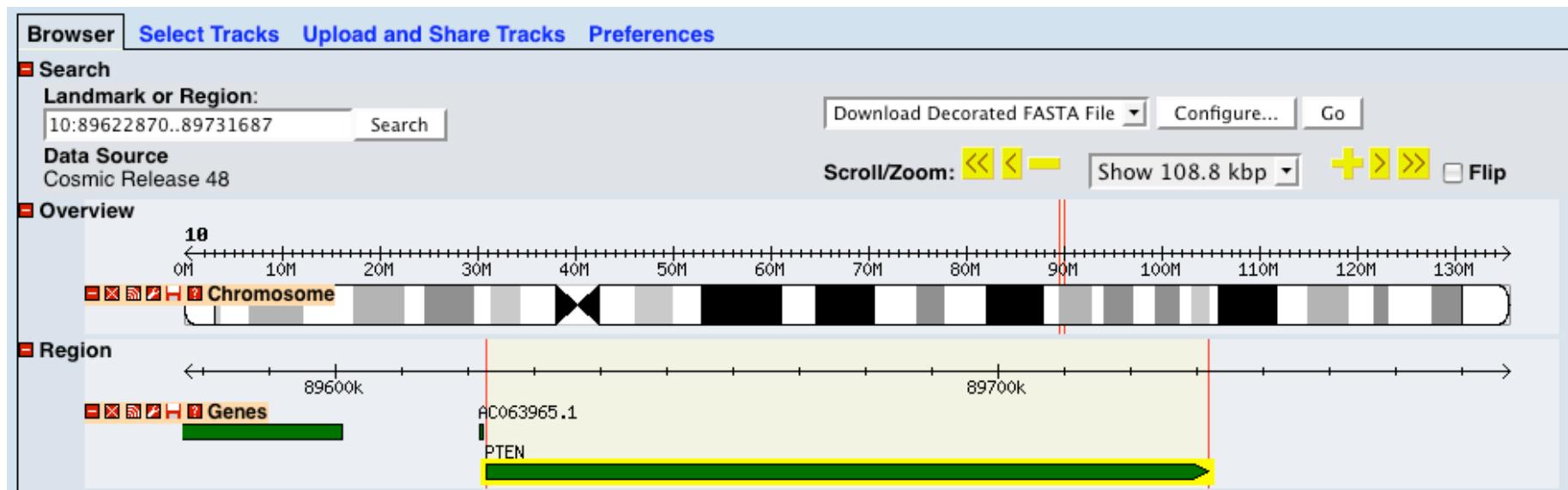
For questions about the data at this site, please contact [cosmic@sanger.ac.uk](mailto:cosmic@sanger.ac.uk).

The Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK Tel:+44 (0)1223 834244

Genome Research Limited is a charity registered in England with number 1021457

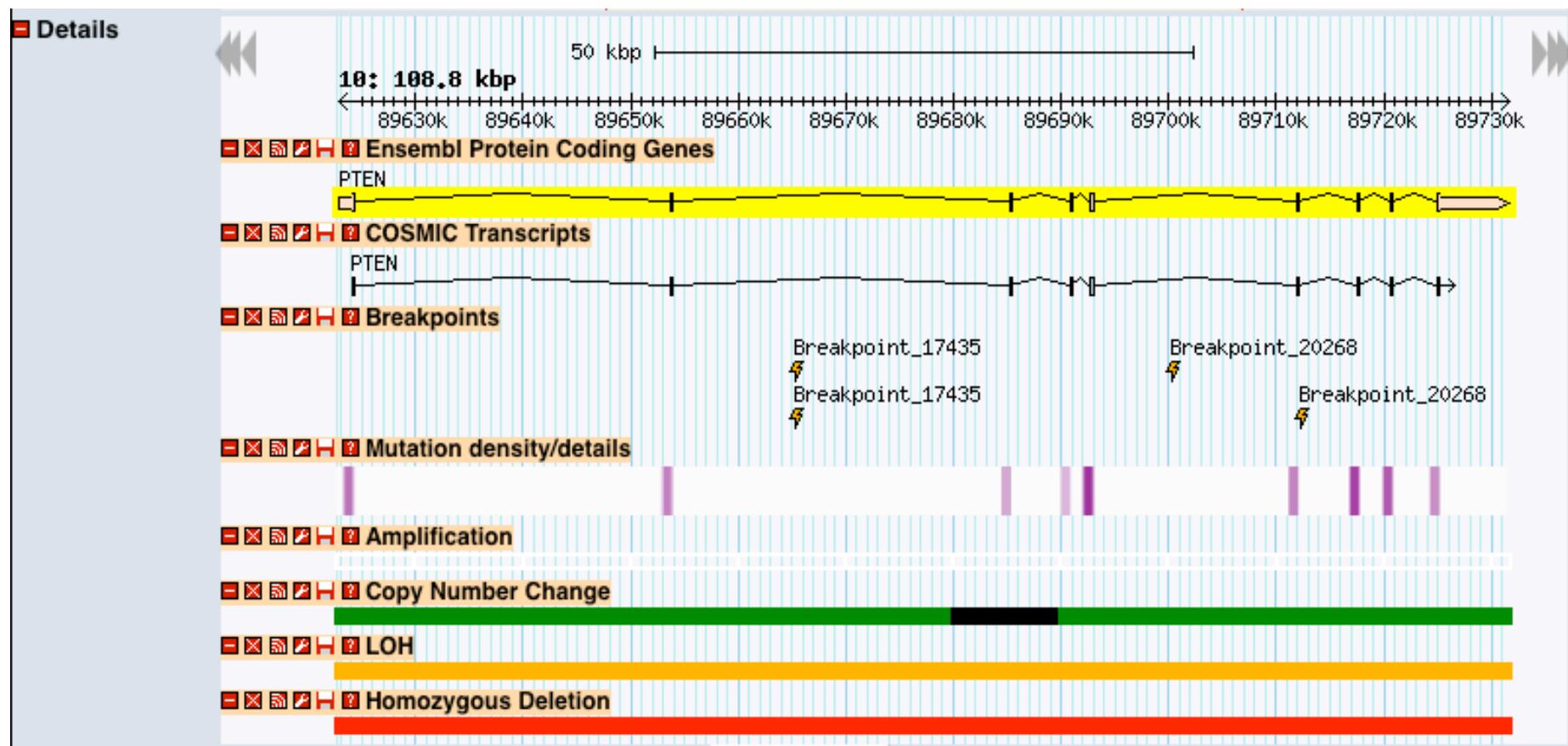
# GBrowse: Overview

<http://www.sanger.ac.uk/fgb2/gbrowse/cosmic>



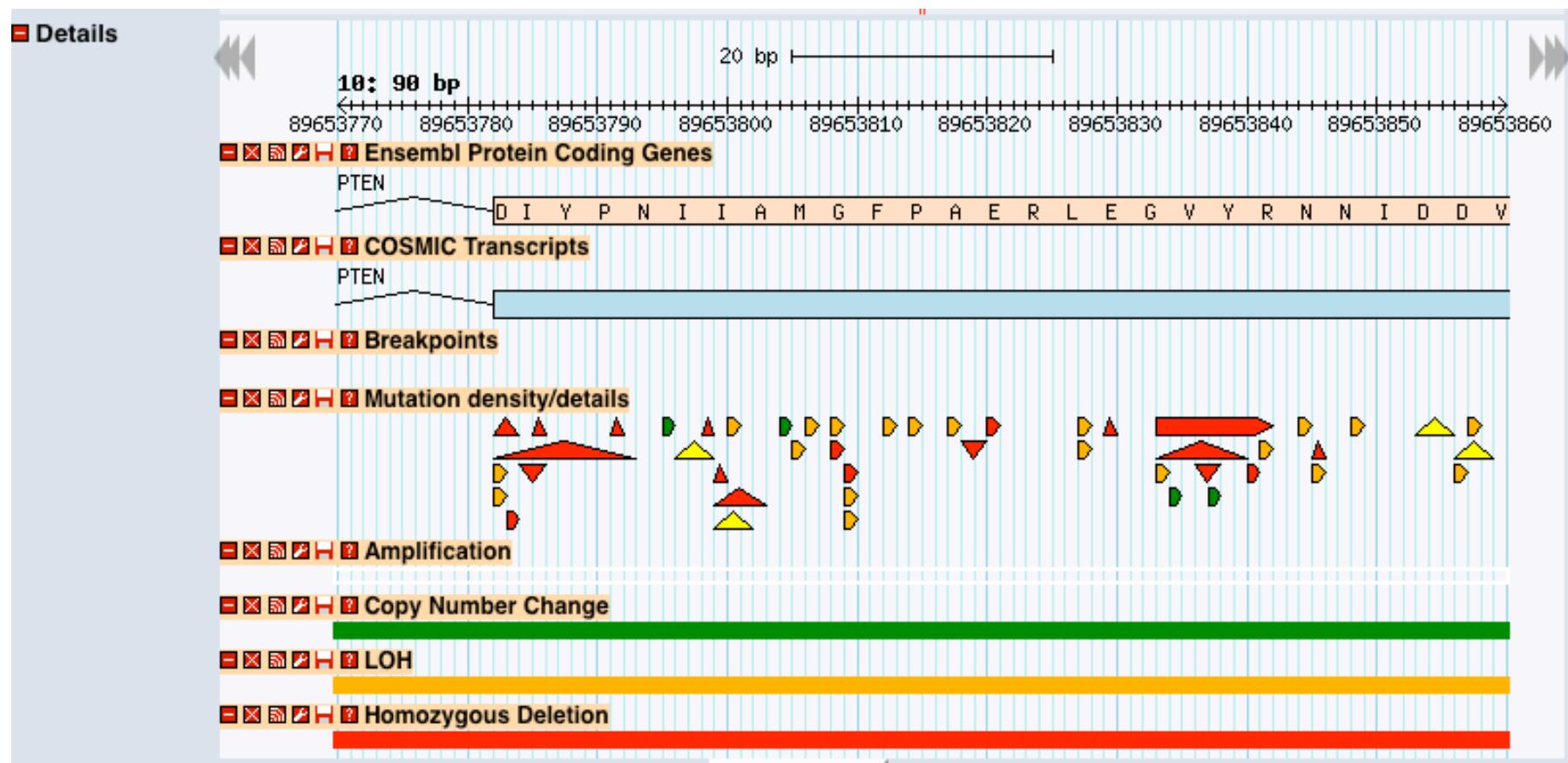
# GBrowse: Details

<http://www.sanger.ac.uk/fgb2/gbrowse/cosmic>



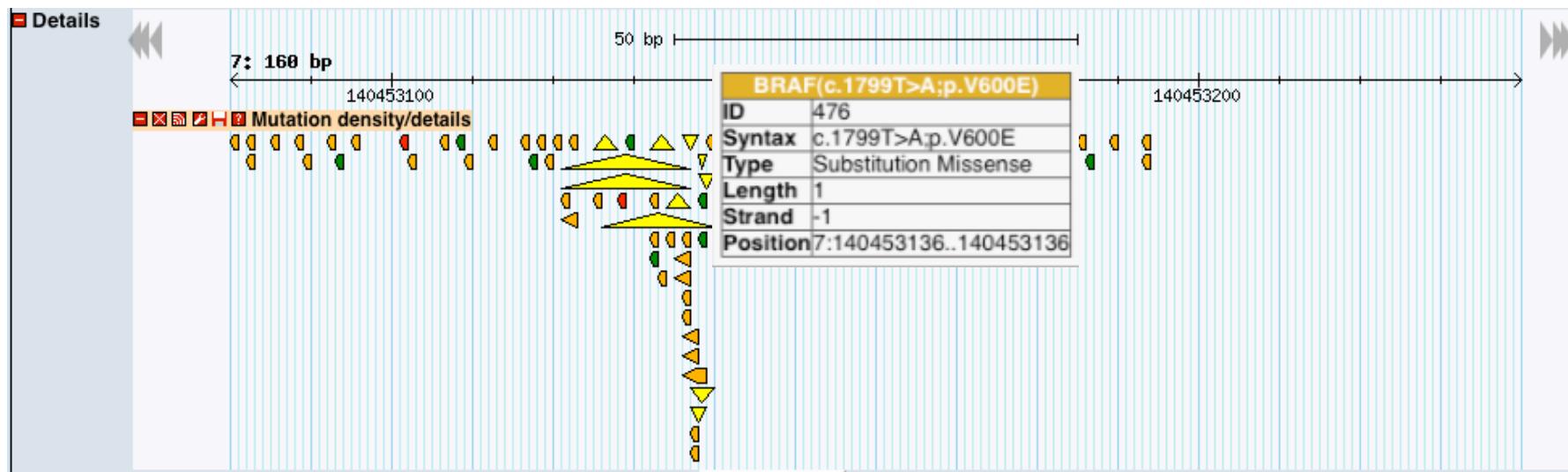
# GBrowse: Zoom

<http://www.sanger.ac.uk/fgb2/gbrowse/cosmic>



# GBrowse: Mutation Details

<http://www.sanger.ac.uk/fgb2/gbrowse/cosmic>



Key to track symbols and colours

Mutations:substitutions		Mutations:other		Structural Variants & Copy Number (% samples)				
Intronic		Frameshift		Breakpoint				
Nonsense		In Frame		LOH	>0-2%	2-20%	20-50%	>50%
Missense		Complex		Change	>0-0.2%	0.2-1%	1-2%	>2%
Silent		Deletion	▲ ▲	Amplification	>0-0.2%	0.2-1%	1-2%	>2%
Non-coding		Insertion	▼ ▼	Homozygous Deletion	>0-0.2%	0.2-1%	1-2%	>2%

# Cosmic: Breakpoints

Mutation id	17510																																																						
Mutation Type	intrachromosomal tandem duplication																																																						
Genomic Annotation	<p>NCBI 36 : <a href="#">Ensembl Contig View</a>  chr7:g.140244671_140248658dup  <small>Click here to switch on the COSMIC coding mutation annotation tracks if you have not previously done so.</small></p> <p>GRCh37 : <a href="#">Ensembl Contig View</a>  chr7:g.140598202_140602189dup  <small>Click here to switch on the COSMIC coding mutation annotation tracks if you have not previously done so.</small></p> <p>COSMIC GBrowse: Please click <a href="#">here</a> to see this data in COSMIC Cancer Genome Browser.</p>																																																						
Breakpoint Details	<table border="1"> <thead> <tr> <th colspan="9">NCBI 36 Annotations</th> </tr> <tr> <th>Mutation Order</th> <th>Chromosome From</th> <th>Breakpoint Position From</th> <th>Strand</th> <th>Chromosome To</th> <th>Breakpoint Position To</th> <th>Strand</th> <th>Non Templatd Inserted Sequence</th> <th>Remark</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>7</td> <td>140244671 </td> <td>-</td> <td>7</td> <td>140248658 </td> <td>-</td> <td>c</td> <td>-</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="9">GRCh37 Annotations</th> </tr> <tr> <th>Mutation Order</th> <th>Chromosome From</th> <th>Breakpoint Position From</th> <th>Strand</th> <th>Chromosome To</th> <th>Breakpoint Position To</th> <th>Strand</th> <th>Non Templatd Inserted Sequence</th> <th>Remark</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>7</td> <td>140598202 </td> <td>-</td> <td>7</td> <td>140602189 </td> <td>-</td> <td>c</td> <td>-</td> </tr> </tbody> </table>	NCBI 36 Annotations									Mutation Order	Chromosome From	Breakpoint Position From	Strand	Chromosome To	Breakpoint Position To	Strand	Non Templatd Inserted Sequence	Remark	0	7	140244671	-	7	140248658	-	c	-	GRCh37 Annotations									Mutation Order	Chromosome From	Breakpoint Position From	Strand	Chromosome To	Breakpoint Position To	Strand	Non Templatd Inserted Sequence	Remark	0	7	140598202	-	7	140602189	-	c	-
NCBI 36 Annotations																																																							
Mutation Order	Chromosome From	Breakpoint Position From	Strand	Chromosome To	Breakpoint Position To	Strand	Non Templatd Inserted Sequence	Remark																																															
0	7	140244671	-	7	140248658	-	c	-																																															
GRCh37 Annotations																																																							
Mutation Order	Chromosome From	Breakpoint Position From	Strand	Chromosome To	Breakpoint Position To	Strand	Non Templatd Inserted Sequence	Remark																																															
0	7	140598202	-	7	140602189	-	c	-																																															
Associated Samples	<table border="1"> <thead> <tr> <th>Sample Name</th> <th>Primary Tissue</th> </tr> </thead> <tbody> <tr> <td>PD3687a</td> <td>breast</td> </tr> </tbody> </table>	Sample Name	Primary Tissue	PD3687a	breast																																																		
Sample Name	Primary Tissue																																																						
PD3687a	breast																																																						



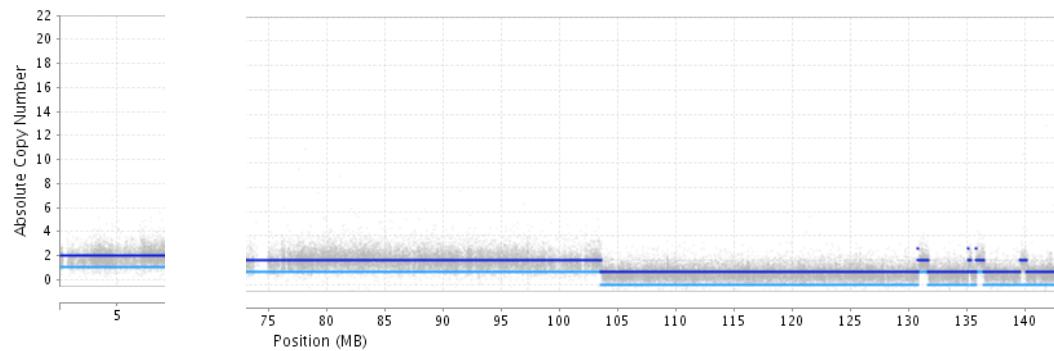
# Cosmic: Mutations

Mutation id	476
AA Mutation	p.V600E (Substitution - Missense)
CDS Mutation	c.1799T>A (Substitution)
Mutation Location	<p>Length (AA) ← → 767.00 AA</p> <p>Amino Acid</p> <p>The diagram shows a protein sequence from position 1 to 767. A red box highlights the V600 residue. An arrow points to the V600 residue, which is highlighted in red. The sequence is color-coded by amino acid type.</p> <p>Length (AA) ← → 41.00 AA</p> <p>Amino Acid Mutation</p> <p>NNIIFLHEDLTVKIGDFGEATVKSRWSGSHQFEQLSGSILWM</p> <p>Mutation: V600E (highlighted in red)</p>
Genomic View	<p>NCBI 36 : <a href="#">Ensembl Contig view</a> 7:140099605..140099605 <a href="#">Click here</a> to switch on the tracks if you have not previously used COSMIC DAS</p> <p>GRCh37 : <a href="#">Ensembl Contig view</a> 7:140453136..140453136 <a href="#">Click here</a> to switch on the tracks if you have not previously used COSMIC DAS</p> <p>COSMIC GBrowse: Please log in</p> <p>The genomic tracks show the BRAF gene structure across chromosomes 7 and 11. The top track displays gene models for NCBI 36 and GRCh37. Below these are tracks for various genomic features, including exons (red), introns (blue), and other regulatory elements. The bottom track shows mutation density and details, indicated by colored triangles (green, yellow, red) and vertical bars.</p>
Gene	<a href="#">BRAF</a>

# Cosmic: Genes

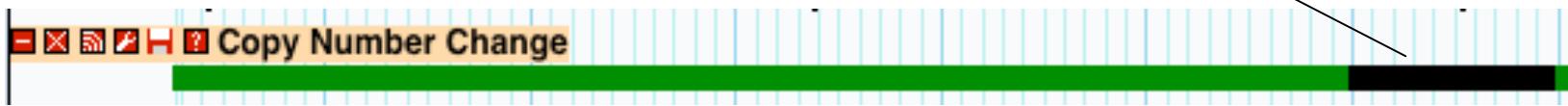
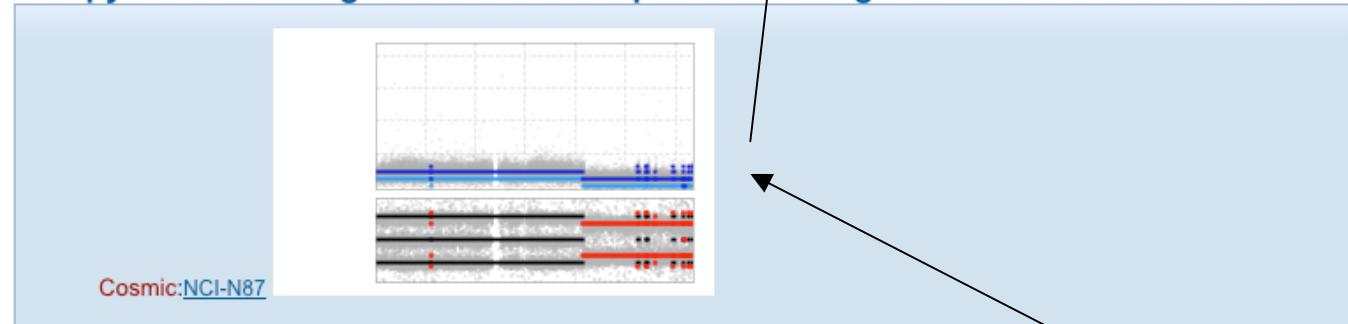
Gene Name	<a href="#">BRAF</a> (HGNC Symbol)
Synonyms: MGC126806, B-raf1, BRAF1, B-raf 1, RAFB1, MGC138284	
Small Intragenic Mutation Summary	<p>Mutations Insertions Nonsense Subs. Missense Subs. Deletions Complex Subs. Silent Subs. scale (AA)</p> <p>Histogram - Click for a histogram of the full gene sequence and mutation details.</p>
Rearrangement Mutation Summary	<p>Fused With:</p> <ul style="list-style-type: none"> <li><a href="#">AGTRAP</a> (1 mutations in 1 samples)</li> <li><a href="#">FCHSD1</a> (1 mutations in 2 samples)</li> <li><a href="#">KIAA1549</a> (130 mutations in 284 samples)</li> <li><a href="#">SLC45A3</a> (1 mutations in 6 samples)</li> </ul>
Additional Info	<p>Fasta Files: <a href="#">cDNA: NM_004333</a> <a href="#">Protein: BRAF</a></p> <p>Transcript and Protein: <a href="#">Aligned: NM_004333+BRAF</a></p> <p>Internal Databases: <a href="#">Array Copy Number Analysis: BRAF</a></p> <p>External Databases: <a href="#">OMIM: 164757</a> <a href="#">ENSEMBL: P15056</a>  <a href="#">NCBI Entrez Gene: 673</a> <a href="#">CCDS: CCDS5863.1</a>  <a href="#">Swiss-Prot: P15056</a> <a href="#">HGNC: 1097</a>  <a href="#">Atlas Genetics Oncology: BRAFID828</a></p> <p>DAS:  NCBI 36: <a href="#">Ensembl Contig View</a>  GRCh37: <a href="#">Ensembl Contig View</a>  Click here to switch on the tracks if you have not previously used COSMIC DAS</p> <p>COSMIC GBrowse: <a href="#">COSMIC Transcripts</a>  Please click       </p>

# Copy Number Profiles



**CONAN: Copy Number Analysis**

» Copy number changes in tumour samples across region: 7:140090001-140100000



# Future Development

1. Embed cosmic gbrowse in some cosmic web pages
  - replace old and slow drawing code
  - extend functionality
2. Current version is a summarised view of whole cosmic dataset but we need to be able to display subsets of data

How can we display all mutations for a specific sample or group of samples, or from a specific tissue or tumour type?

Too many for a static list of data sources, but there is a neat trick ..

Define data source in the URL, eg sample COLO-829

[http://www.sanger.ac.uk/fgb2/gbrowse/sample\\_COLO-829](http://www.sanger.ac.uk/fgb2/gbrowse/sample_COLO-829)

# Future Development

2. GBrowse.conf ... (need atleast 2.09)

see [http://gmod.org/wiki/GBrowse\\_2.0\\_HOWTO](http://gmod.org/wiki/GBrowse_2.0_HOWTO)

"Using Pipes in the GBrowse.conf Data Source Name"

→ [=~sample\_.+]

description = Cosmic Database v48 (sample filtered)

→ path = /gbrowse/bin/source\_config.pl -sample \$1 |

# path points to a script which generates the config

# sample name 'COLO-829' is passed to the script from regular expression

# track configuration generated for data source COLO-829 ...

[Mutations]

→ remote feature = http://.../cosmic\_export.cgi?sample=COLO-829

# cgi script returns COLO-829 mutation data from COSMIC

# GBrowse fixes/enhancements

1. remote feature  
perl callbacks cannot be used until Safe::World is fixed
2. init\_code  
perl callbacks defined with init\_code not accessible from slaves
3. BAM/SAM read sorting by similarity to reference
4. GC plots can give >100% values

# Summary

- CGP committed to using GBrowse
  - internal browser for next gen sequencing data
  - external browser for COSMIC data
    - genomic view of mutations, breakpoints and copy number data
    - COSMIC GBrowse to be released soon - 22/9/2010 ?
- CGP involvement in GBrowse development
  - new developer recruited
  - details still being discussed

# Credits

Sanger:

COSMIC Group

db - **Simon Forbes**, Mingming Jia, Rebecca Shepherd

web - Nidhi Bindal, [**Prasad Gunasekaran**]

Cancer IT Group:

**Kairan Raine**, Jon Teague, Adam Butler

Systems Support Group: Tim Cutts

DBA team: Tony Webb

Web Team: James Smith, Paul Bevan

GMOD:

**Gmod-gbrowse list**