COSMIC GBrowse

Visualising cancer mutations in genomic context

Dave Beare

dmb@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 ×

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

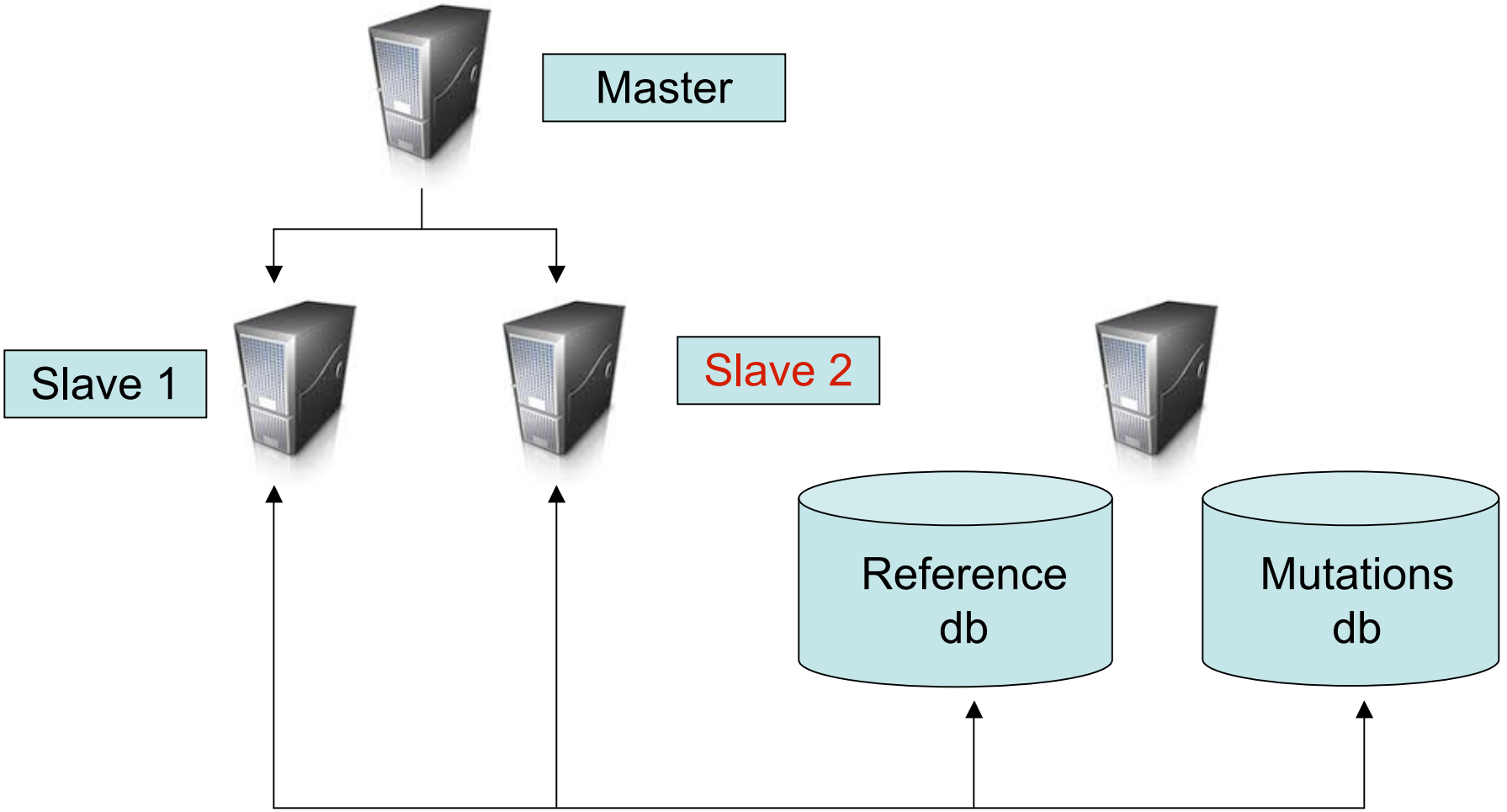
- Reference
  - -- Reference genome (GRCh37) + cytogenetic bands
  - -- Ensembl annotations (e! 58)
  - -- Cosmic Transcripts

- Cosmic
  - -- Mutations (subsitutions, 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

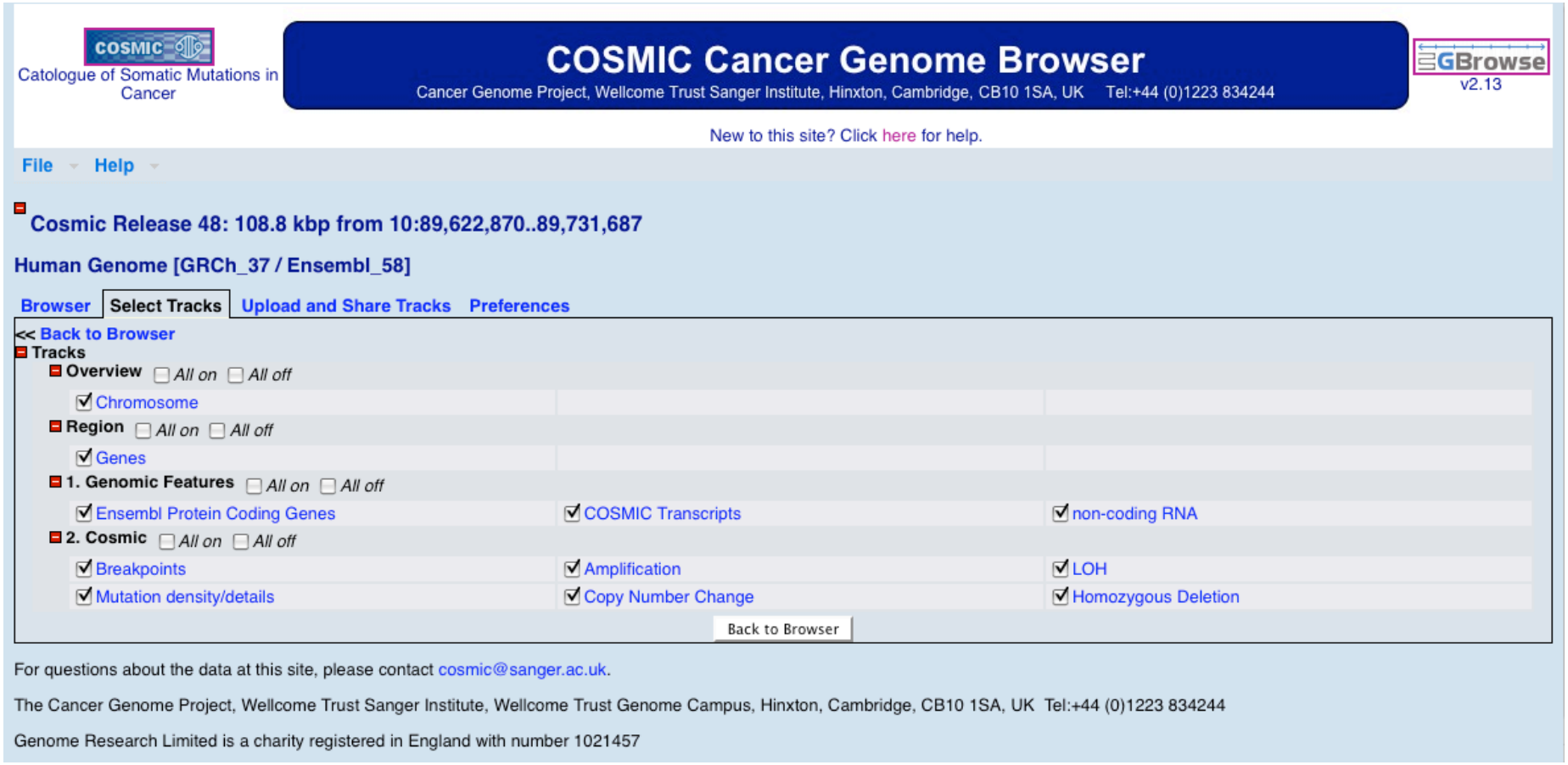
- cosmic css/theme
- perl callbacks
  - -- glyphs
  - -- colours
  - -- hyperlinks
  - -- popups/tooltips
- renderfarm enabled

GBrowse: Render Farm



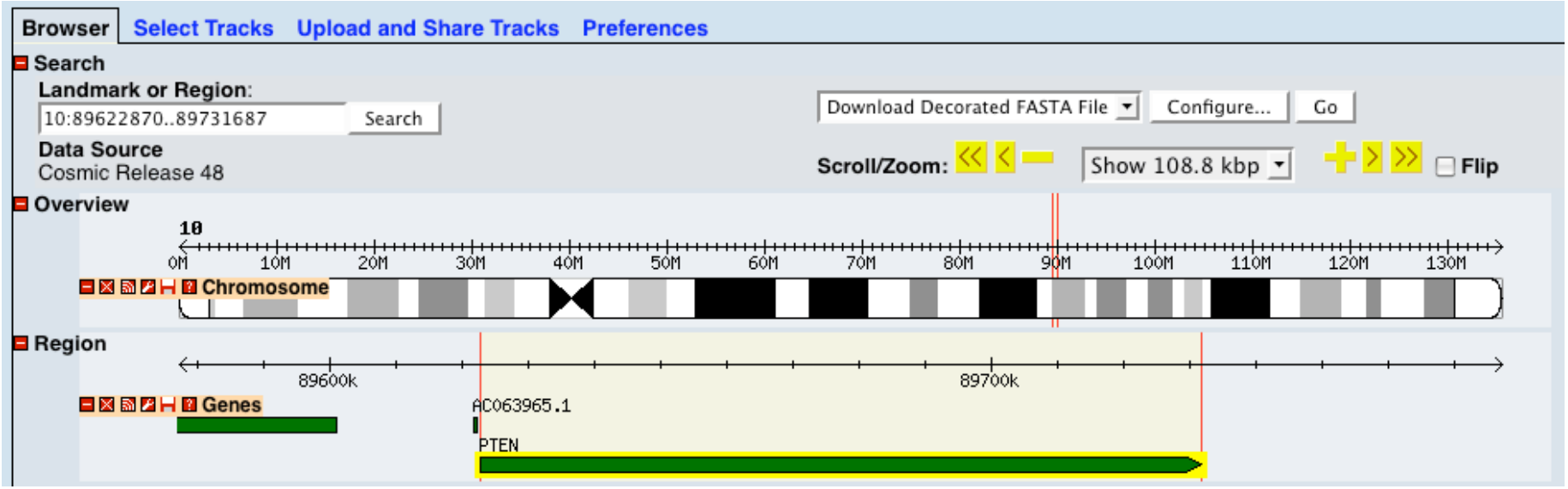
GBrowse: Select Tracks

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



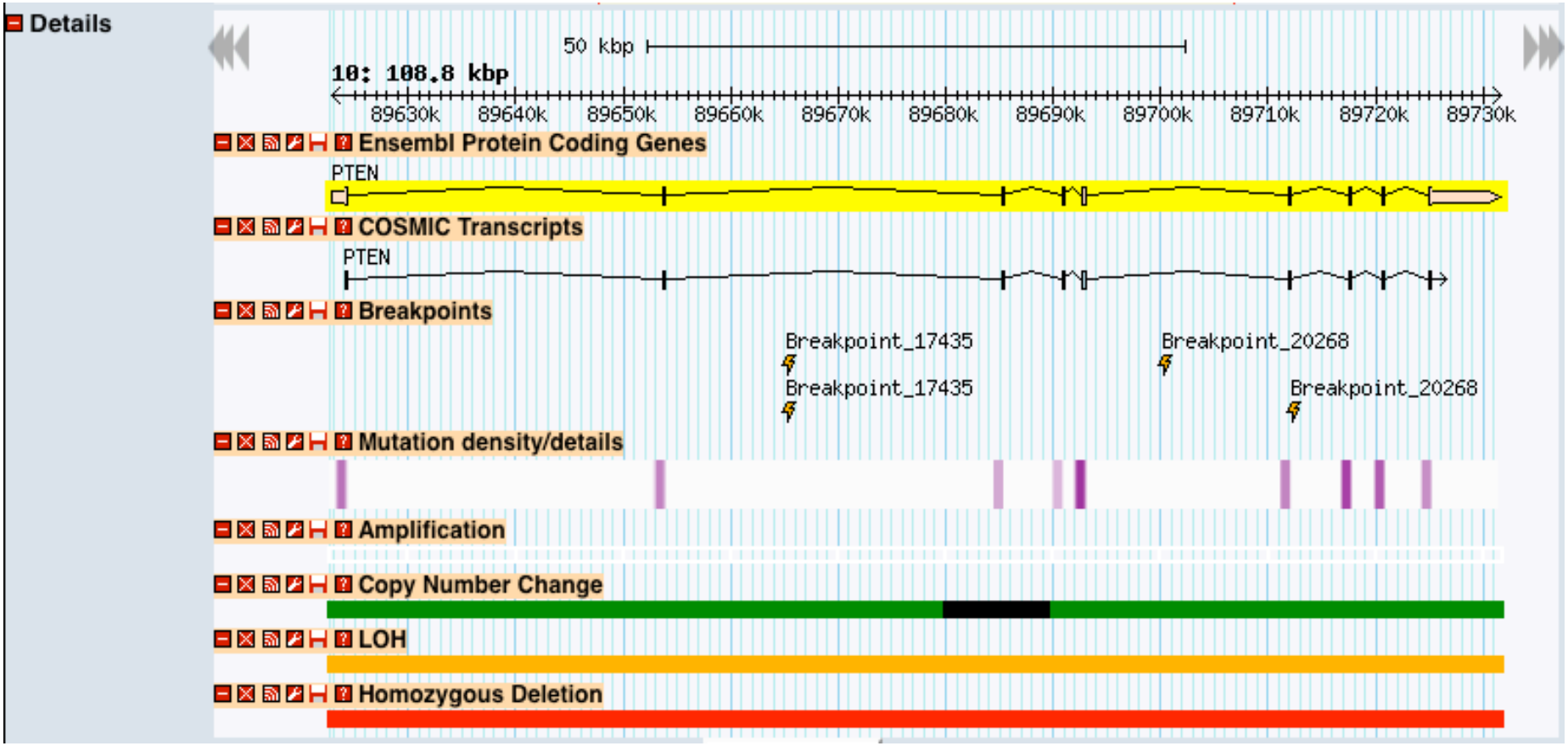
GBrowse: Overview

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



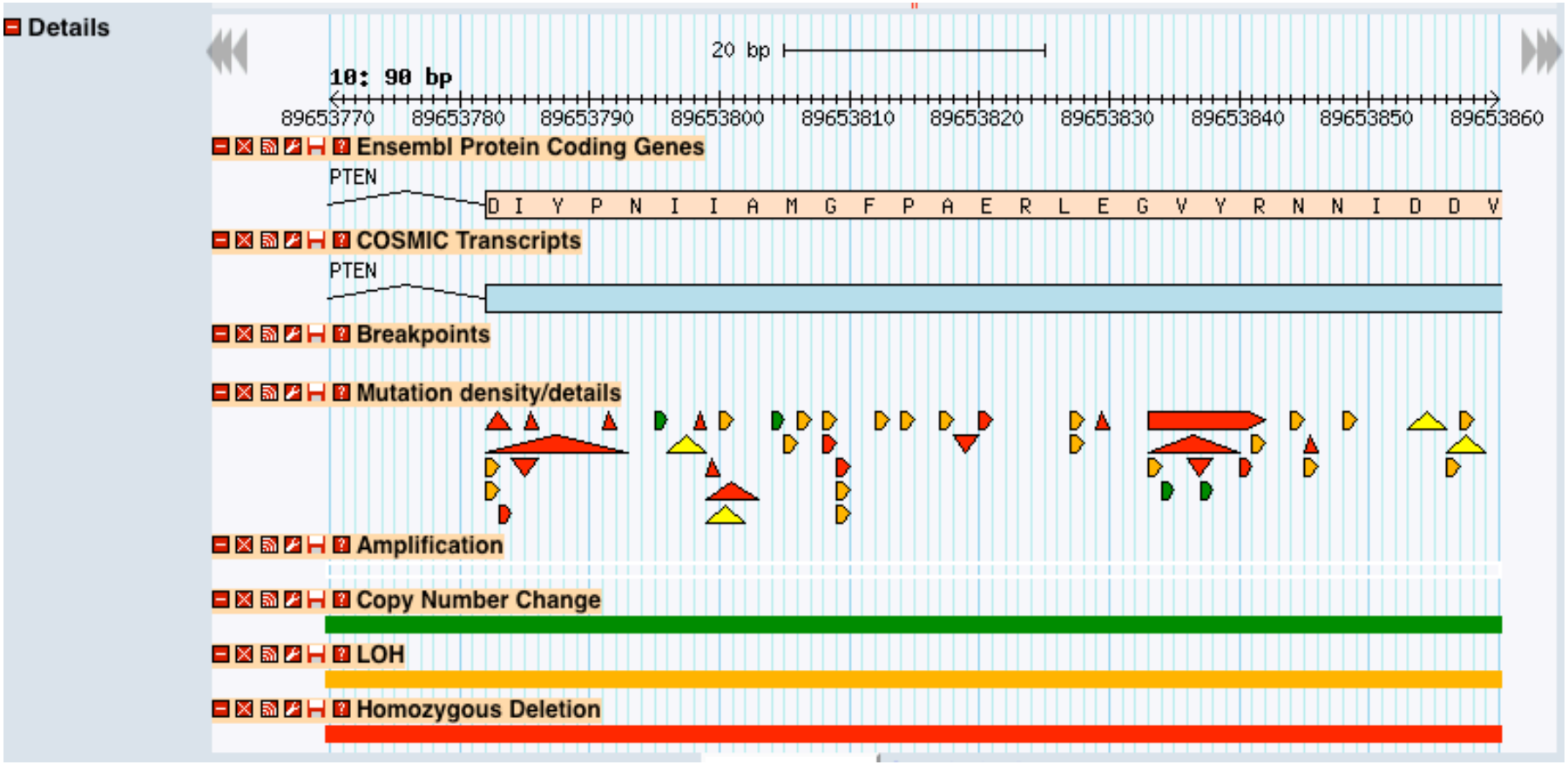
GBrowse: Details

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



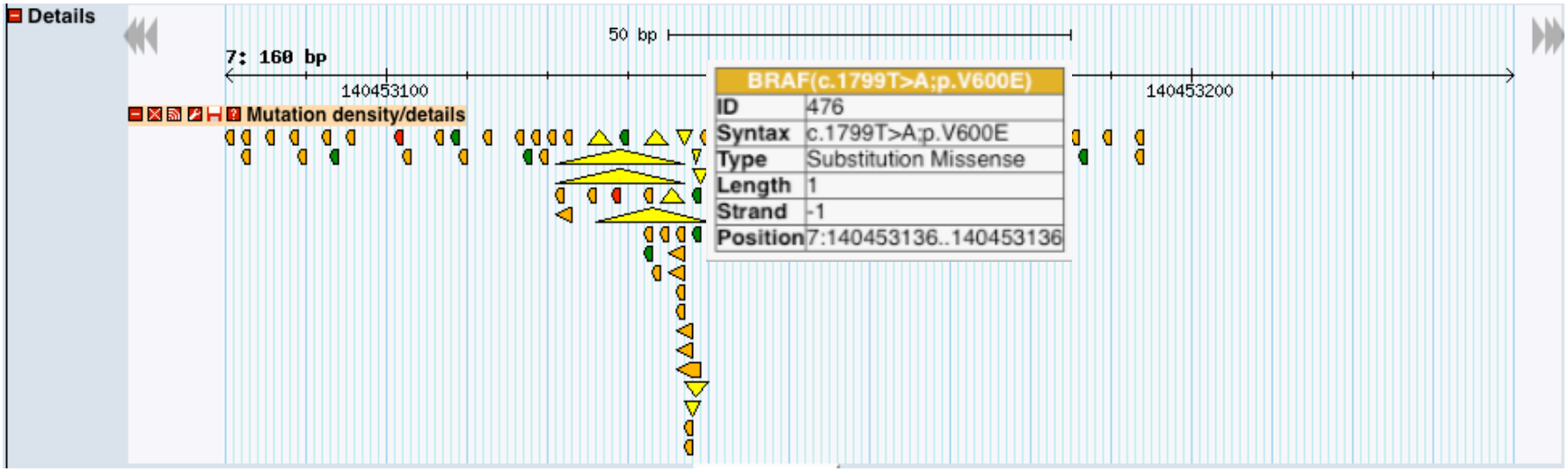
GBrowse: Zoom

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	**NCBI 36 :** Ensembl Contig View chr7:g.140244671_140248658dup Click here to switch on the COSMIC coding mutation annotation tracks if you have not previously done so. **GRCh37 :** Ensembl Contig View chr7:g.140598202_140602189dup Click here to switch on the COSMIC coding mutation annotation tracks if you have not previously done so. **COSMIC GBrowse:** Please click here to see this data in COSMIC Cancer Genome Browser.

Breakpoint Details

NCBI 36 Annotations

Mutation Order	Chromsome From	Breakpoint Position From	Strand	Chromosome To	Breakpoint Position To	Strand	Non Templated Inserted Sequence	Remark
0	7	140244671	-	7	140248658	-	c	-

GRCh37 Annotations

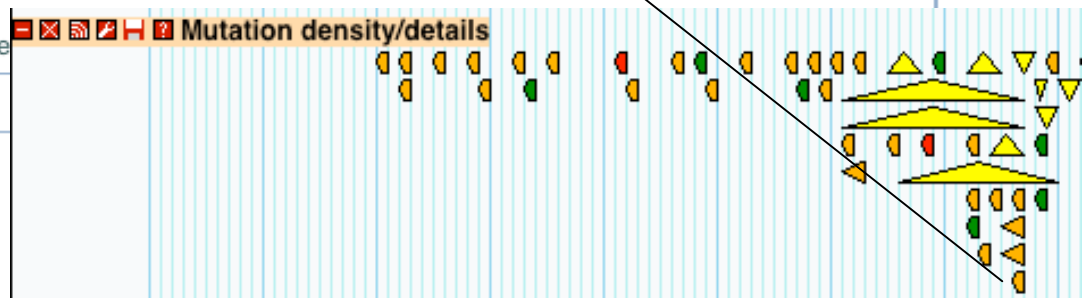
Mutation Order	Chromsome From	Breakpoint Position From	Strand	Chromosome To	Breakpoint Position To	Strand	Non Templated Inserted Sequence	Remark
0	7	140598202	-	7	140602189	-	c	-

Associated Samples

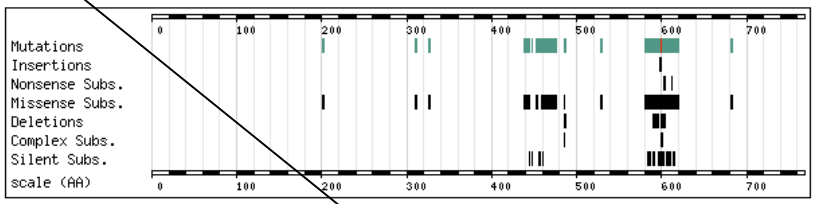
Sample Name	Primary Tissue
PD3687a	breast

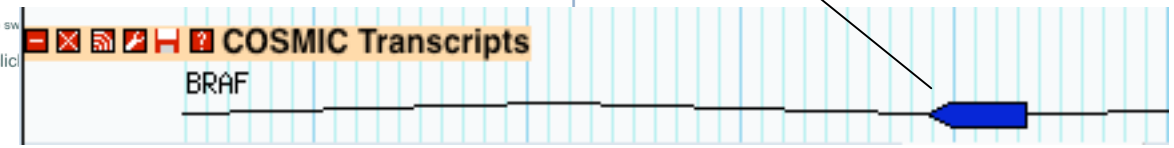
Cosmic: Mutations

Mutation id	476
AA Mutation	p.V600E (Substitution - Missense)
CDS Mutation	c.1799T>A (Substitution)
Mutation Location	Length (AA) 767.00 AA Amino Acid Length (AA) 41.00 AA Amino Acid Mutation N N I F L H E D L T V K I G D F G L A T V K S R W S G S H Q F E Q L S G S I L W M
Genomic View	**NCBI 36 :** Ensembl Contig view 7:140099605..140099605 Click here to switch on the tracks if you have not previously used COSMIC DAS **GRCh37 :** Ensembl Contig view 7:140453136..140453136 Click here to switch on the tracks if you have not previously used COSMIC DAS **COSMIC GBrowse:** Ple
Gene	BRAF

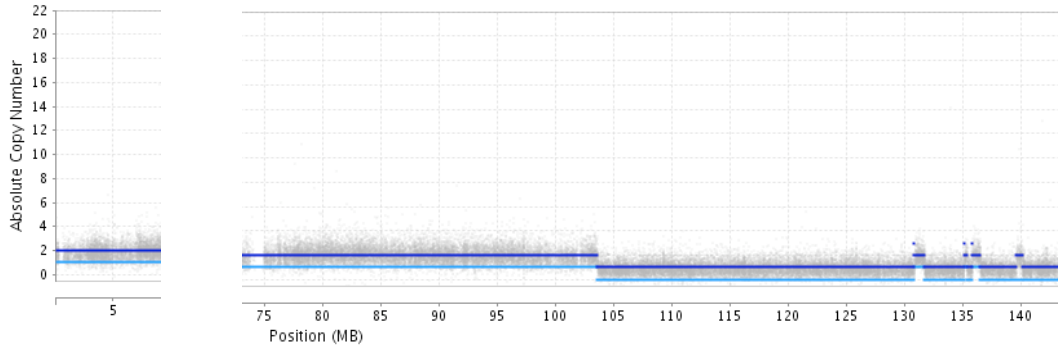


Cosmic: Genes

Gene Name	**BRAF** (HGNC Symbol) Synonyms: MGC126806, B-raf1, BRAF1, B-raf 1, RAFB1, MGC138284
Small Intragenic Mutation Summary	 **Histogram** - Click for a histogram of the full gene sequence and mutation details.
Rearrangement Mutation Summary	Fused With: AGTRAP (**1** mutations in 1 samples) FCHSD1 (**1** mutations in 2 samples) KIAA1549 (**130** mutations in 284 samples) SLC45A3 (**1** mutations in 6 samples)
Additional Info	**Fasta Files:** **cDNA:** NM_004333 **Protein:** BRAF **Transcript and Protein:** **Aligned:** NM_004333+BRAF **Internal Databases** **Array Copy Number Analysis:** BRAF **External Databases:** **OMIM:** 164757 **ENSEMBL:** P15056 **NCBI Entrez Gene:** 673 **CCDS:** CCDS5863.1 **Swiss-Prot:** P15056 **Atlas Genetics Oncology:** BRAFID828 **HGNC:** 1097 **DAS:** NCBI 36 **Ensembl Contig View** Click here to switch on the tracks if you have not previously used COSMIC DAS GRCh37 **Ensembl Contig View** Click here to sw **COSMIC GBrowse:** Please clic

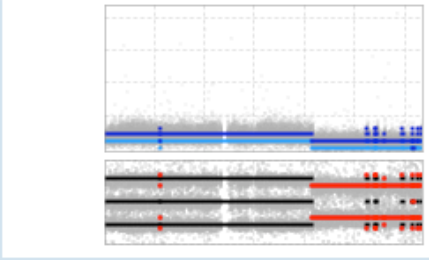


Copy Number Profiles



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

Future Development

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