

# New advanced solutions for Genomic *big data* Analysis and Visualization

Ignacio Medina (*Nacho*)  
[im411@cam.ac.uk](mailto:im411@cam.ac.uk)

<http://www.hpc.cam.ac.uk/compbio>  
Head of Computational Biology Lab  
HPC Service, University of Cambridge  
Cambridge, UK

Team lead for Bioinformatics Software Development  
Bioinformatics Department, Genomics England  
London, UK



UNIVERSITY OF  
CAMBRIDGE



# Overview

- Introduction
  - OpenCB initiative
- Advanced Computing Technologies
- OpenCB Projects
  - CellBase
  - Genome Maps, *Big data* visualization
  - High-Performance Genomics (HPG)
  - OpenCGA
- Q&A

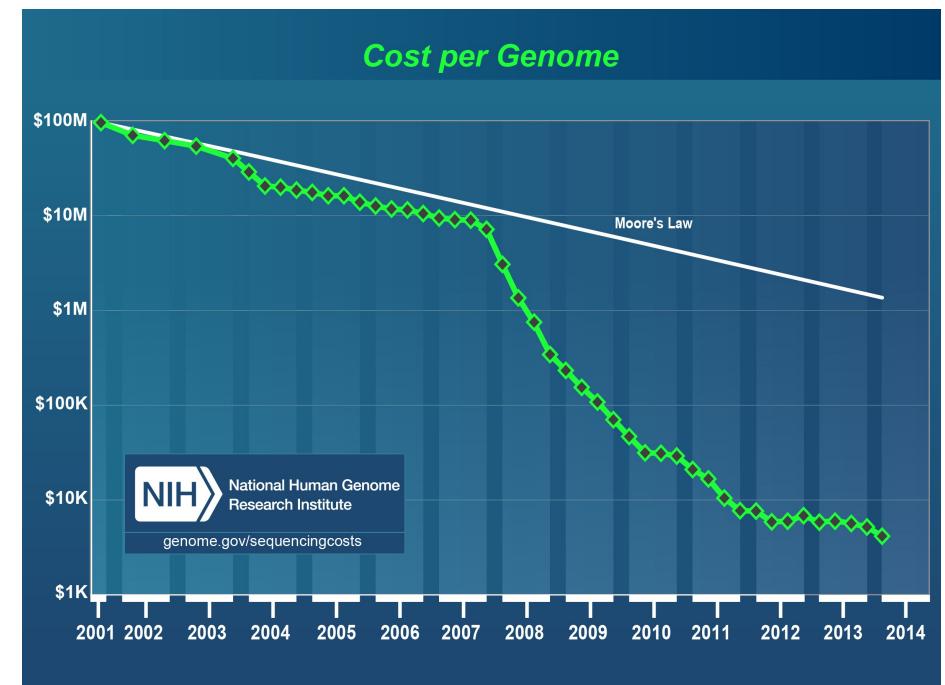
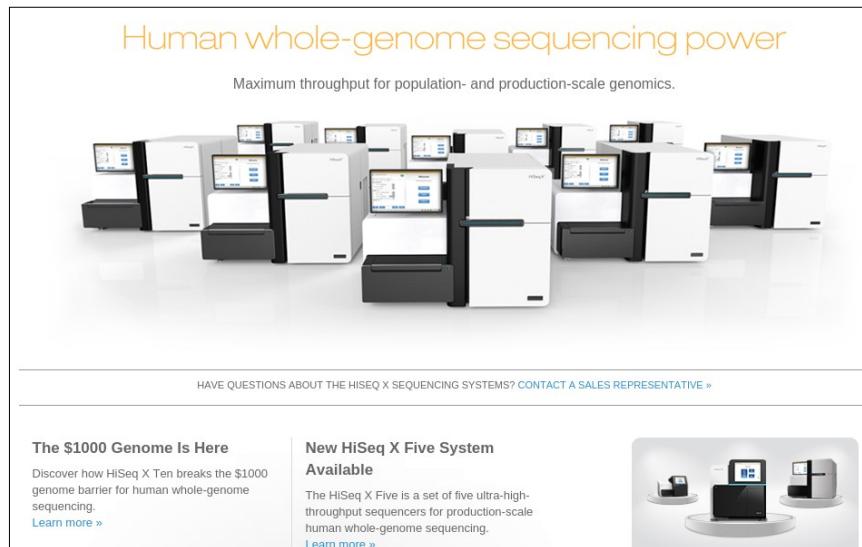
# Introduction

## *Big data in Genomics, a new scenario in biology*

**Next-Generation Sequencing (NGS)** is a high-throughput technology for DNA sequencing that is changing the way how researchers are performing genomic experiments. Most new experiments are being conducted by sequencing: *re-sequencing, RNA-seq, Meth-seq, ChIP-seq, ...* Experiments **have increased data size by more than 5000x when compared with microarrays. Surprisingly, most of the existing software solutions are not very different.**

Sequencing costs keep falling, today a whole genome can be sequenced by just **\$1000**, so much more data is expected. Data acquisition is **highly distributed** and involves heterogeneous formats.

A single HiSeq X Ten System can sequence ~20,000 human genomes a year



# Introduction

## Standard NGS experiment data analysis

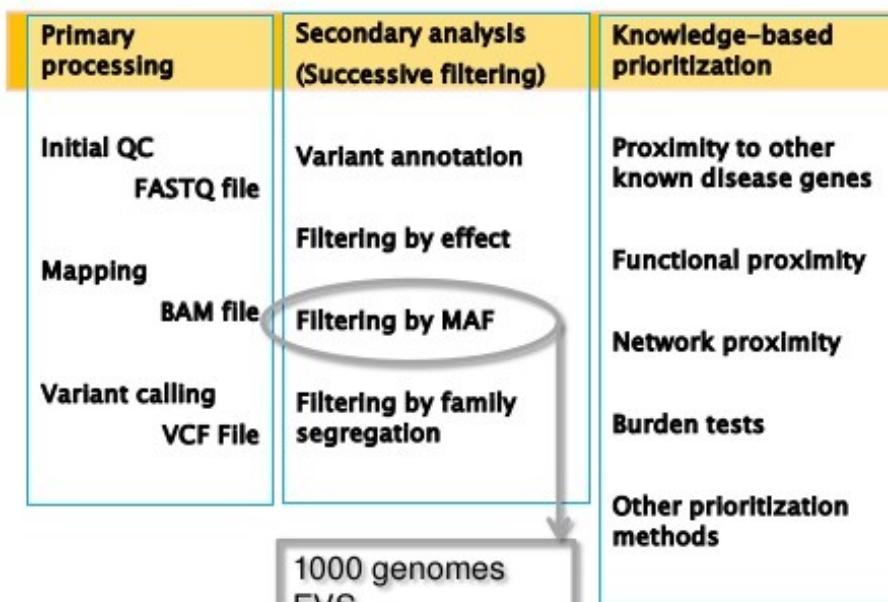
- Illumina NGS sequencer series:
  - **HiSeq 2500** provides high-quality 2x125bp: 50-1000Gb in 1-6 days, 90.2% bases above Q30. One human genome at ~60x coverage
  - **HiSeq 4000** provides high-quality 2x150bp: 125-1500Gb in 1-4 days, >75% bases above Q30. Up to 12 human genomes at ~40x coverage
- **Each sample** produces a **FASTQ** file ~1TB size containing ~1-2B reads
- New **Illumina X Ten**: Consists of 10 ultra-high-throughput HiSeq X sequencers. First \$1000 human genome sequencer, it can sequence up to 20,000 genomes per year



**Real flexibility.**  
**Real throughput.**  
**Real data quality.**

The HiSeq 2500 is ready for any application, any sample size—today.

## Pipeline of data analysis



**Variant Calling pipeline**

# Introduction

## *Big data in Genomics*

- Some current projects:
  - At EMBL-EBI and Sanger
    - **European Genome-phenome Archive (EGA)**: stores human datasets under controlled access  
<https://www.ebi.ac.uk/ega/home> Current size about 2PB, it is expected to increase 2x-3x over the next few years. Now new functionality is being implemented.
    - **European Variation Archive (EVA)**: open archive for all public genomic variation data for all species  
<http://www.ebi.ac.uk/eva/> A new project with only a few TBs of data so far
    - **1000G Phase 3**: about 2500 individuals from 26 populations, a few hundreds of TBs
  - Other ***big data*** projects
    - **NIHR BRIDGE**: 10,000 whole genomes from rare diseases, ~1-2PB of data expected
    - **Genomics England (GEL)**: is sequencing **100,000** whole genomes from UK, several rare diseases and cancers being studied, data estimation: ~20PB of BAM and **~400TB of VCF** data are expected! About 100 whole genomes/day, ~5-10TB/day
    - **International Cancer Genome Consortium (ICGC)**: store more than 10,000 sequenced cancers, few PB of data
- And of course **many medium-sized** projects. Data acquisition in Genomics is ***highly distributed, there is not a single huge project***
  - Square Kilometer Array (SKA) project will produce 750 terabytes/second by 2025
- **Data encryption and security** is a major concern in many of them
- Data stored in ***different*** data centers

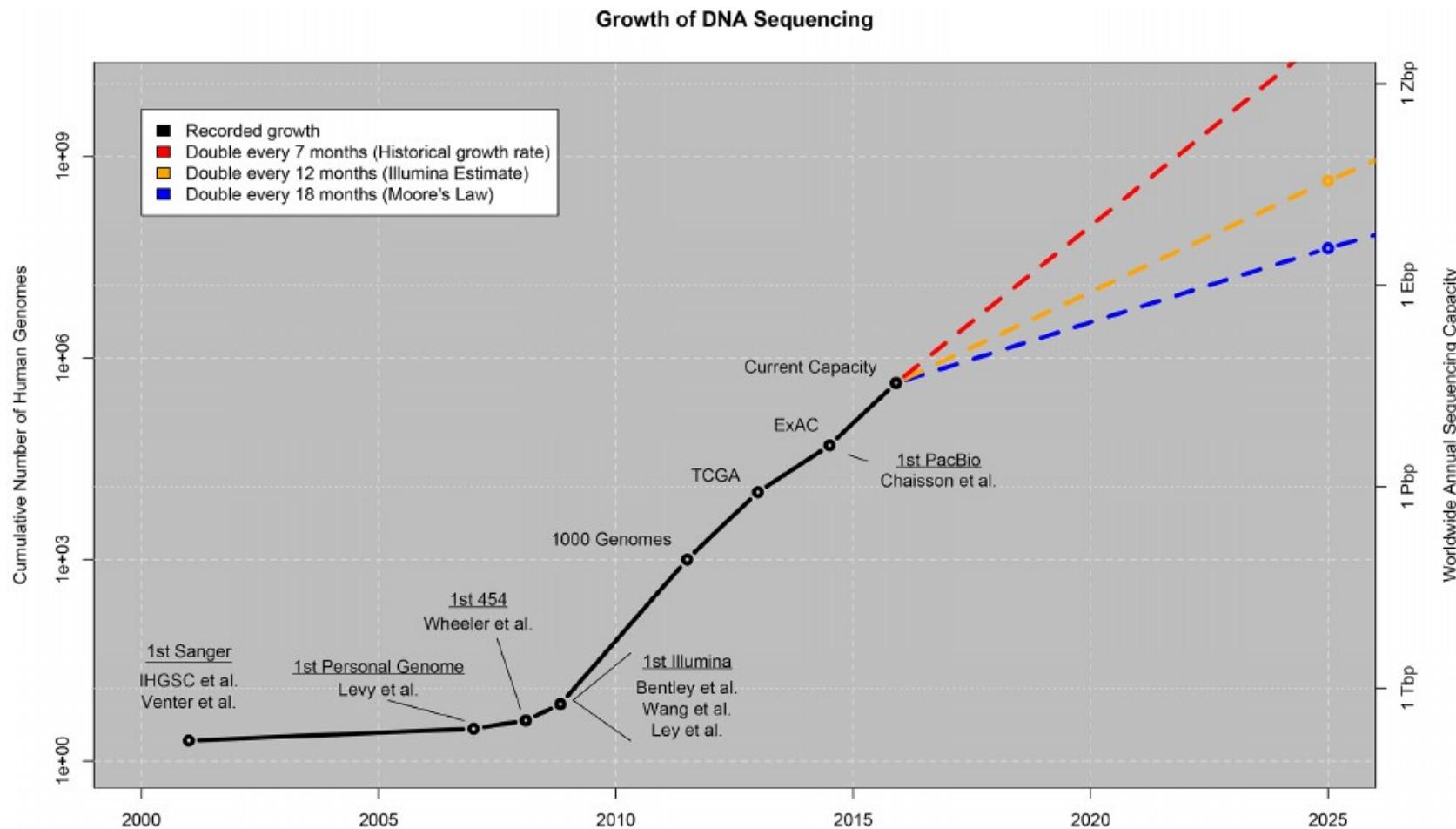
# Introduction

## Estimated sequencing data size projection for human

### **Big Data: Astronomical or Genomical?**

Stephens ZD. et al., PLOS Biology, July 2015

**ABSTRACT** - Genomics is a Big Data science and is going to get much bigger, very soon, but it is not known whether the needs of genomics will exceed other Big Data domains. **Projecting to the year 2025**, we compared genomics with three other major generators of Big Data: **astronomy, YouTube, and Twitter**. Our estimates show that **genomics is a “four-headed beast”—it is either on par with or the most demanding of the domains analyzed here in terms of data acquisition, storage, distribution, and analysis**



# Introduction

## Genomic Variant Dataset, big and complex

**Logical view** of genomic variation dataset,  
data come from **different VCF files**.

Hundreds of millions of mutations, some meta data needed: **Variant annotation**

- Clinical info
- Consequence types
- Conservation scores
- Population frequencies
- ...

**Genomics England** project:

- **200M variants x 100K samples**, about **20 trillion points**
- With different layers of data, about **80-100 trillion points**
- a lot of meta data for variants and samples
- about **400TB** to be indexed

		Samples						
		var_1	28	32	29	28	35	32
var_1	16,12	16,17	18,14	12,14	16,14	16,12		
var_2	C/C	C/G	C/C	C/G	C/C	G/G		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
..	..	..	..	..	..	..		
var_n	..	..	..	..	..	..		

**Different layers** of information:

- Genotype for samples
- Allele counts
- Quality scores
- ...

Meta data: **Sample annotation**

- Phenotype
- Family and population pedigree
- Clinical variables
- ...

**Heterogeneous data analysis and algorithms**, different technologies and solutions required:

- Search and filter using data and meta data
- Data mining, correlation
- Statistic tests
- Machine learning
- Interactive analysis
- Network-based analysis
- Visualization
- Encryption
- ...

**Applications:**

- Personalized medicine
- ...

# Introduction

## *Big data analysis challenges*

- **Data Analysis and visualization:** Real-time and **Interactive** graphical data analysis and visualization is needed.
- **Data mining:** Complex queries, aggregations, correlations, ...
- **Security:** sometimes data access require authentication, authorization, *encryption*, ...
- **Performance and scalability:** software must be high-performance and scalable
- **Data Integration:** different types of data such as variation, expression, ChIP, ...
- **Share and collaboration:** data models to ease the collaboration among different groups. Avoid moving data.
- **Knowledge base and sample annotations:** many of the visual analytic tools need genome and sample annotations



*Do current bioinformatic tools solve these problems?*

# Introduction

## Current status in bioinformatics

- Many bioinformatic tools are great but, in general, not designed and implemented for processing and analyzing big data.
- Tools usually don't exploit the parallelism of modern hardware and current high-performance and scalable technologies. Poor performance and scalability.
- We need to develop new generation of software and methodologies to:
  - Improve performance and scalability of analysis
  - Store data efficiently and secured to be queried and visualized
  - Easy to adapt to new technologies and uses cases to be useful to researchers
- So, are bioinformaticians doing things right? Are researchers happy with their current tools?
- Is needed a ***paradigm change*** in bioinformatics? Most tools are designed and implemented to run in a single ***workstation***, but PB scale data require more ***advanced computing technologies***



# OpenCB

## Open source initiative for Computational Biology

- Software in Biology is still usually developed in small teams, we must learn to collaborate in bigger projects to solve bigger problems. OpenCB tries to engage people to work in big problems: <http://www.opencb.org>
- Shared, collaborative and well designed platforms to build more **advanced solutions to solve current biology problems** are needed.
- OpenCB **aims** to design and develop high-performance and scalable solutions for genomic big data analysis using most modern computing technologies.
- No one computing programming language oriented: BioPerl, BioPython, Bioconductor, ... Good software solutions may use different languages and technologies to solve different problems and use cases
- So far, is where all the software we develop is being released. About 15 active committers. Available as open-source at GitHub <https://github.com/opencb>

OpenCB is a collaborative project with more than 15 actives developers and data analysts and more than 12 repositories

Many papers published during last two years, very good adoption

Many different technologies used: HPC, Hadoop, web applications, NoSQL databases, ...

The screenshot shows the GitHub profile for the 'opencb' organization. It features three repository cards:

- opencga**: An Open Computational Genomics Analysis platform for big data processing and analysis in genomics. Last updated 6 hours ago.
- bionetdb**: BioNetDB implements a storage engine to work with biological networks using a NoSQL Graph database. Last updated 13 hours ago.
- hpg-bigdata**: This repository implements converters and tools for working with NGS data in HPC or Hadoop cluster. Last updated 16 hours ago.

On the right side, there's a 'People' section showing a grid of developer profiles and a 'Invite someone' button. The top navigation bar includes links for 'Repositories', 'People 23', 'Teams 7', and 'Settings'.

# OpenCB

## Some relevant projects

- **biodata** (<https://github.com/opencb/biodata>) and **ga4gh** (<https://github.com/opencb/ga4gh>)
  - Contain all data models, parsers and converters (*avro, protobuf*) for all OpenCB projects
- **CellBase** (<https://github.com/opencb/cellbase>)
  - The biological knowledge-base for OpenCB project (*ensembl core, variation and regulatory, uniprot, clinvar, cosmic, expression, conservation, ...*) A **Variant Annotation** tool implemented.
  - A high-performance NoSQL implementation, a CLI and web services implemented
- **HPG BigData** (<https://github.com/opencb/hpg-bigdata>)
  - Hadoop-based implementation of data converters (*avro, parquet*) and bioinformatic tools (ie. samtools)
  - Simple indexing for HBase, Hive and Impala developed
  - C code embedded using JNI to speed-up processing
- **OpenCGA** (<https://github.com/opencb/opencga>)
  - Integrates most of the OpenCB projects to provide a scalable and high-performance platform
  - **OpenCGA Catalog** provides an authenticated environment, files and sample annotations, system audit, ...
  - **OpenCGA Storage** is a plugin-oriented framework that allows to index hundreds of millions of variants for thousands of samples in different storage engines. Stats and annotation implemented..
- **Genome Maps** (<https://github.com/opencb/genome-maps>)
  - A web-based NGS and genome browser: <http://genomemaps.org/>
- Many other related projects for big data analysis and visualization, check <https://github.com/opencb>

# OpenCB

## A *big data* friendly architecture

### Client

Rich Web applications and visualization. New HTML5 web technologies: SVG, IndexedDB, WebWorkers, WebGL, SIMD.js, ...

### Command Line Interfaces (CLI)



### Visualization and rich HTML5 web applications



#### Genome Maps



#### BierApp



### Server

Distributed and HPC technologies for real-time and interactive data analysis and visualization: NoSQL, Hadoop, HPC, ...

Search, filter and aggregate the **data needed** by researchers

### CellBase

Java APIs and *RESTful* web services

#### Knowledge Base

Genome  
Gene & transcripts  
Variation & Clinical  
Regulatory  
Systems Biology



### OpenCGA

Application Layer: Java APIs and *RESTful* web services  
*Query Data, launch jobs, sessions*

#### Analysis

Exploratory and Genomic Data Analysis  
Big Data indexing and analysis

#### Execution Framework

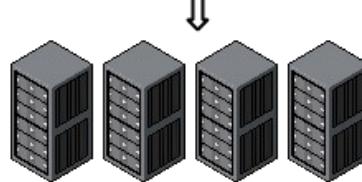
HPC and Hadoop clusters  
Slurm, MapReduce, Spark

#### Storage Engine

Searches, filters and complex queries  
MongoDB, HBase

#### Catalog and Security

Authentication & authorization  
Samples, files and jobs



# Advanced Computing Technologies

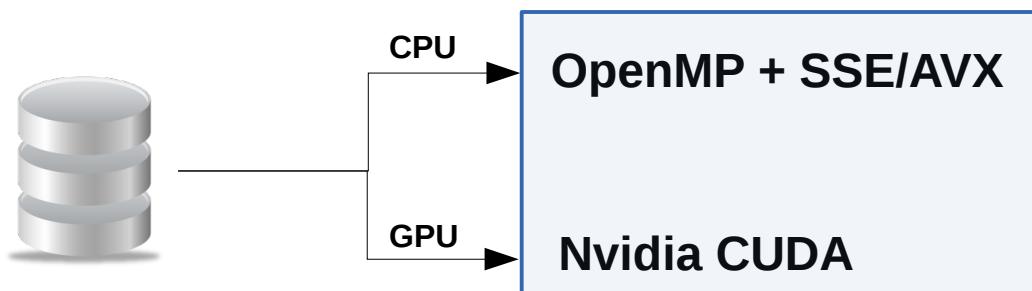
## Web technologies, HTML5 standard and RESTful WS

- **HTML5** brings many new standards and libraries to JavaScript allowing developing amazing web applications in a easy way
  - **Canvas/SVG inline**: browsers can render dynamic content and inline static SVG easily
  - **IndexedDB**: client-side indexed storage for high-performance query
  - **WebWorkers**: simple mean for creating OS threads
  - **WebGL**: hardware-accelerated 3D graphics to web, based on OpenGL ES
- Many new web tools, frameworks and libraries being developed to build great and bigger web applications:
  - Performance: asm.js, SIMD.js, ...
  - Web Components: Polymer
  - Build: Grunt, Bower, Yeoman, ...
  - Visualization: d3.js, Three.js, Highcharts, ...
- RESTful web services and JSON ease the development of light and fast RPC

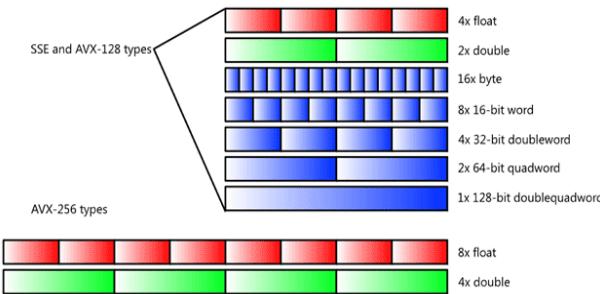
# Advanced Computing Technologies

## High-Performance Computing (HPC)

- HPC hardware:
  - **Intel MIC architecture:** Intel Xeon and Intel Xeon Phi coprocessor, 1.01Tflops DP and more than 50 cores
  - **Nvidia Tesla:** Tesla K20X almost 1.31Tflops DP and 2688 CUDA cores
- Some HPC frameworks available:
  - **Shared-memory parallel:** OpenMP, OpenCL
  - **GPGPU computing:** CUDA, OpenCL
  - **Message passing Interface (MPI)**
  - **SIMD:** SSE4 instructions extended to AVX2 with a 256-bit SIMD
- Heterogeneous HPC in a shared-memory
  - CPU (OpenMP+AVX2) + GPU (CUDA)
- Hybrid approach:



MPI  
Hadoop  
MapReduce



# Advanced Computing Technologies

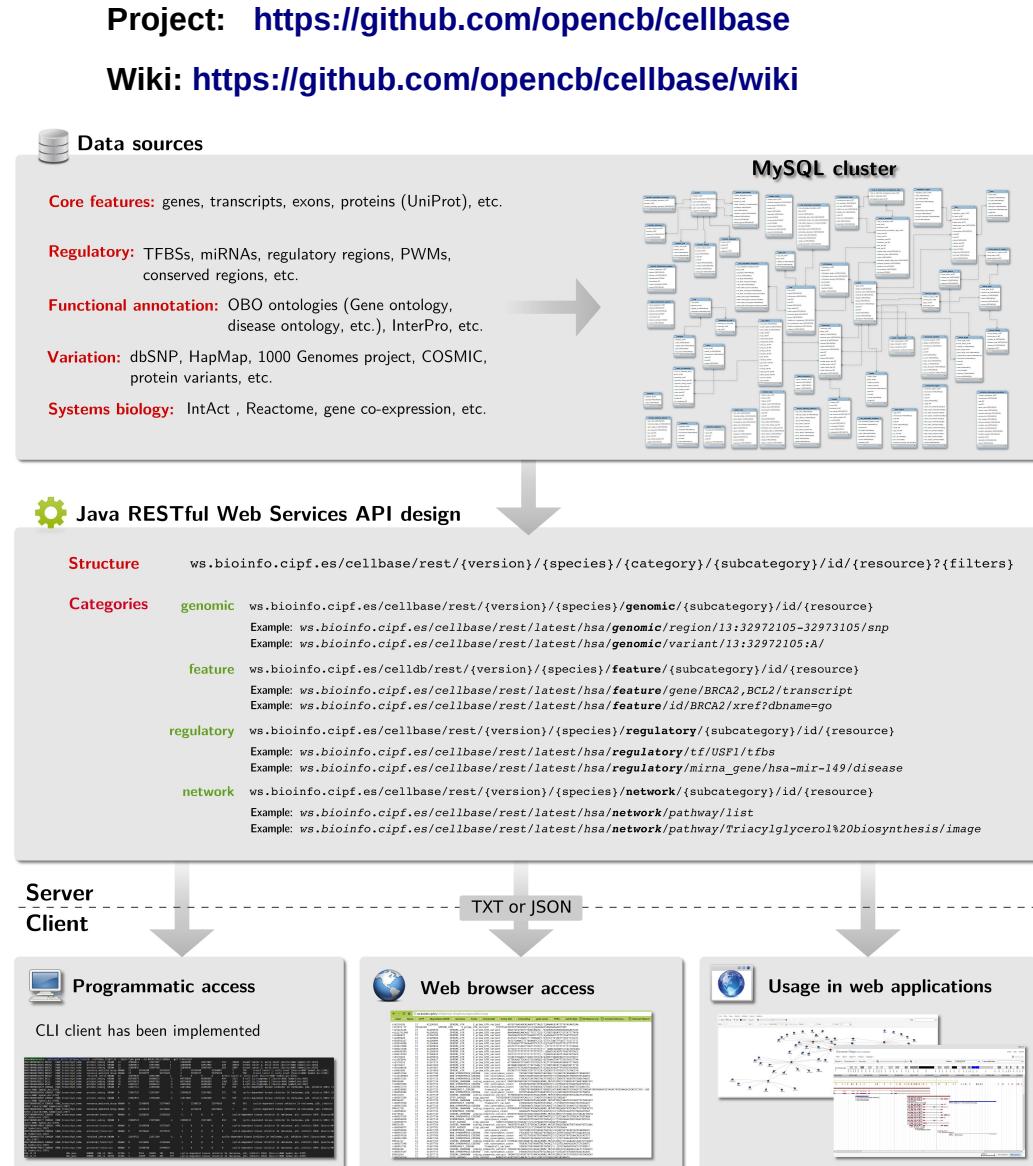
## *Big data* analysis and NoSQL databases

- Apache Hadoop (<http://hadoop.apache.org/>) is currently *de facto* standard for ***big data processing and analysis***:
  - **Core**: HDFS, MapReduce, HBase
  - **Spark**: SparkML, SparkR
- **NoSQL databases**, four main families of ***high-performance distributed and scalable*** databases:
  - *Column store*: Apache Hadoop HBase ...
  - *Document store*: MongoDB, Solr, ...
  - *Key-Value*: Redis, ...
  - *Graph*: Neo4J, ...
- New solutions for PB scale ***interactive analysis***:
  - *Google Dremel* (Google BigQuery) and similar implementations: new *Hive*, *Cloudera Impala*
  - Nested data, and comma and tab-separated data, **SQL queries allowed**

# CellBase

## An integrative database and RESTful Web Service API

- **CellBase** is a comprehensive integrative NoSQL database and a *RESTful Web Service API*, designed to provide a **high-performance a scalable** solution. Currently contains more than 1TB of data:
  - *Ensembl Core features*: genome sequence, genes, transcripts, exons, gene expression, conservation, ...
  - Protein: UniProt, Interpro
  - Variation: dbSNP and Ensembl Variation, Cosmic, ClinVar, ...
  - Functional: OBO ontologies(Gene Ontology), Interpro domains, drug interactions, ...
  - Regulatory: TFBS, miRNA targets, conserved regions, CTCF, histones, ...
  - Systems biology: Reactome, Interactome (IntAct)
- Published at NAR 2012:
  - <http://nar.oxfordjournals.org/content/40/W1/W609>
- Used by EMBL-EBI, ICGC, GEL, ... among others



# CellBase

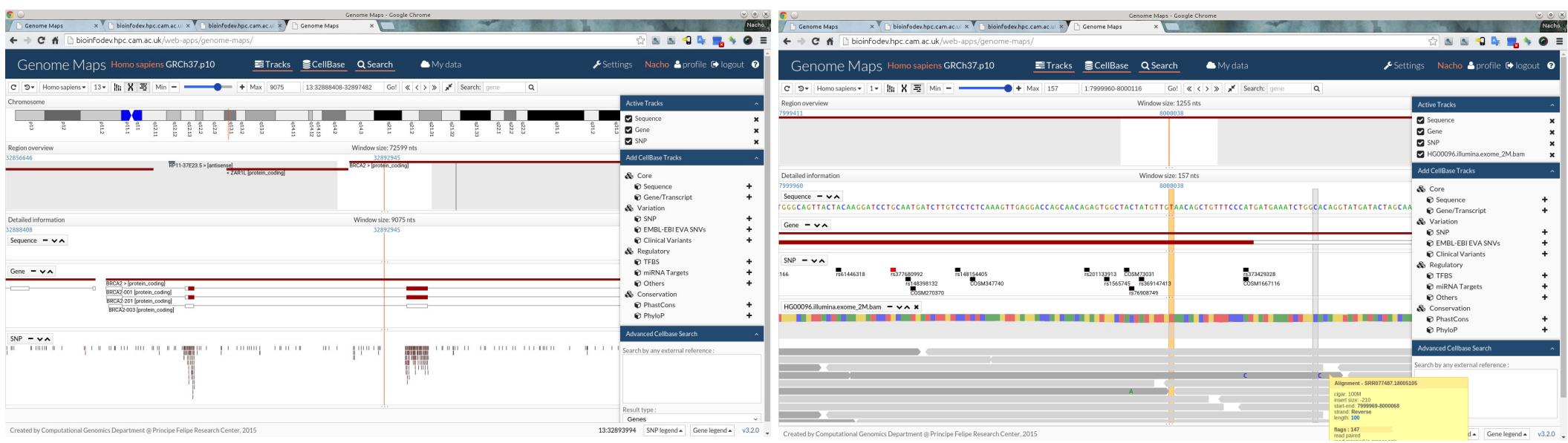
## New features in version 3.x and 4.x

- CellBase **v3.0** moved to **MongoDB** NoSQL database to provide a much higher performance and scalability, most queries run in <40ms. Installed at University of Cambridge and EMBL-EBI, some links:
  - GitHub <https://github.com/opencb/cellbase>
  - Wiki <https://github.com/opencb/cellbase/wiki>
  - Official domain and Swagger <http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/>
- **Variant annotation** integrated:
  - <http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/rest/v3/hsapiens/genomic/variant/19:45411941:T:C/annotation>
- Coming features in **v4.0**:
  - Focus on Clinical data
  - Many more species
  - Aggregation and stats
  - R and Python clients
  - Data customization
  - Richer and scalable API

# Genome Maps

## A *big data* HTML5+SVG genome browser

- Genome scale data **visualization** is an important part of the data analysis: *Do not move data!*
- Main features of **Genome Maps** ([www.genomemaps.org](http://www.genomemaps.org), published at NAR 2013)
  - 100% HTML5 web based: **HTML5+SVG, and other JavaScript libraries**. Always updated, **no browser plugins needed**
  - Genome data is mainly consumed from **CellBase and OpenCGA** database through **RESTful web services**. **JSON** data is parsed and SVG is rendered, making server lighter and improve network transfers
  - Other features: NGS data viewer, Multi species, Feature caches, API oriented, embeddable, key navigation, ...
  - Beta: <http://bioinfodev.hpc.cam.ac.uk/web-apps/genome-maps/>



# Genome Maps

## Software Architecture and Design

Genome Maps design goals are

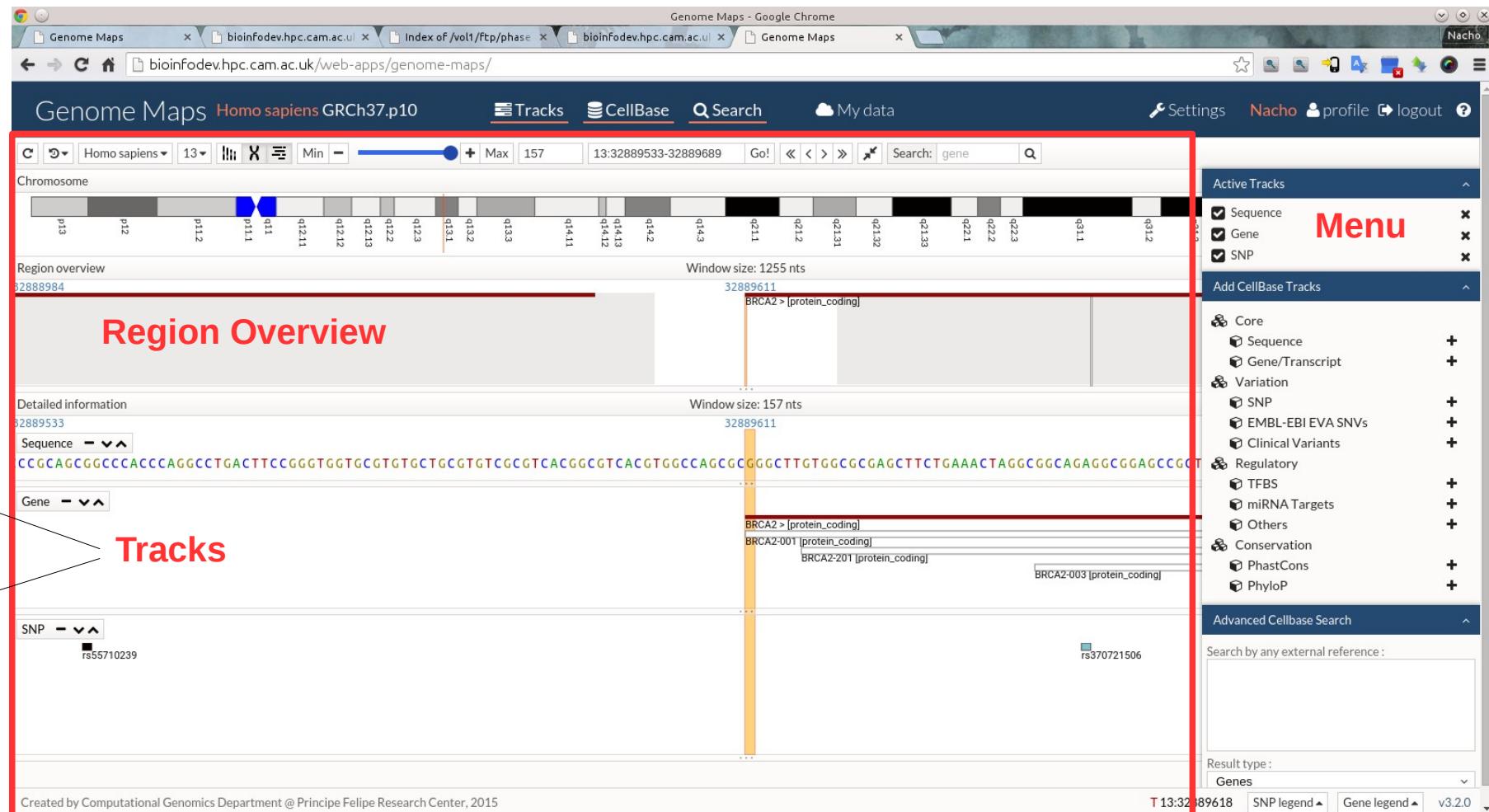
- provide a high-performance visualization → **JSON data** fetched remotely: CellBase, OpenCGA, EVA, ... **no images** sent from the server
- easy to integrate: **event system** and **JavaScript API** developed
- memory efficient: integrated cached (**IndexedDB**), reduces the number of calls
- OpenCB JSorolla JS library

Genome  
Viewer



DataAdapter  
(cache)

Renderer



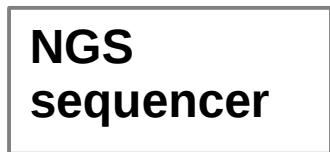
# Genome Maps

## New features and coming releases

- Version **v3.2** (summer 2015)
  - More species (~20) and a more efficient **IndexedDB** based **FeatureCache**, less memory footprint and less remote queries
  - More NGS data friendly, better rendering and features for BAM and VCF files
  - More secure, uses HTTPS and can read and **cache encrypted data** (ie. AES)  
<http://bioinfo.cipf.es/apps-beta/genome-maps/encryption/>
- Future version **v4.0**, tentative release date during 1Q16
  - *JSCircos* for structural variation and other visualizations ( <http://bioinfo.cipf.es/apps-beta/circular-genome-viewer/> )
  - *RNA-seq* and 3D *WebGL* visualization
  - Local data browsing using Docker
  - New Hadoop-based server features being added
- Used by some projects:
  - EMBL-EBI EVA: <http://www.ebi.ac.uk/eva/?Home>
  - ICGC data portal: <http://icgc.org/>
  - New genome browser at Peer Bork group <http://ct.bork.embl.de/ctbrowser/>
  - Lens Beta website: <http://patseq.dev.lens.org/lens/>

# HPG Suite

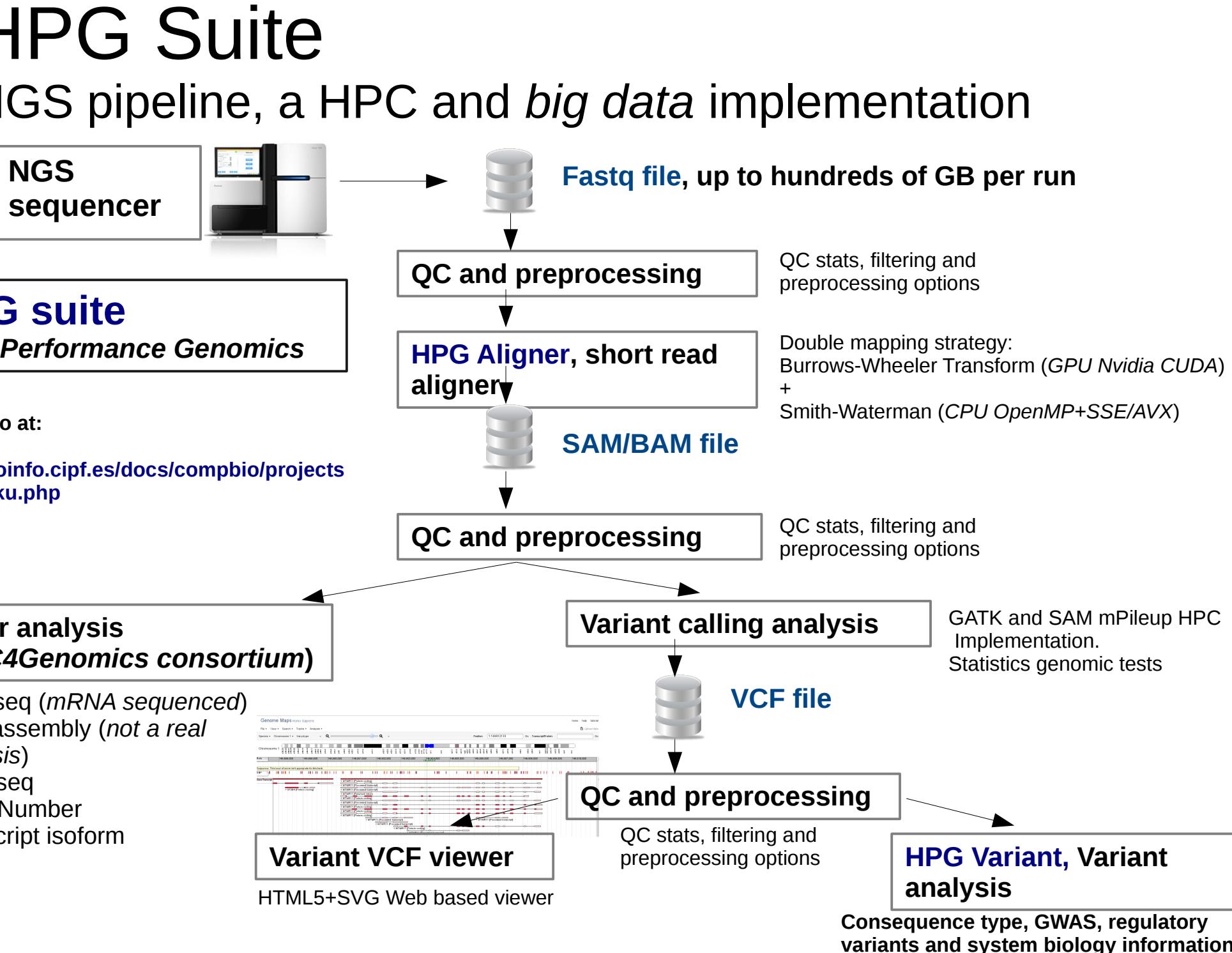
NGS pipeline, a HPC and *big data* implementation



**HPG suite**  
*High-Performance Genomics*

More info at:

<http://bioinfo.cipf.es/docs/compbio/projects/hpg/doku.php>



# HPG Aligner

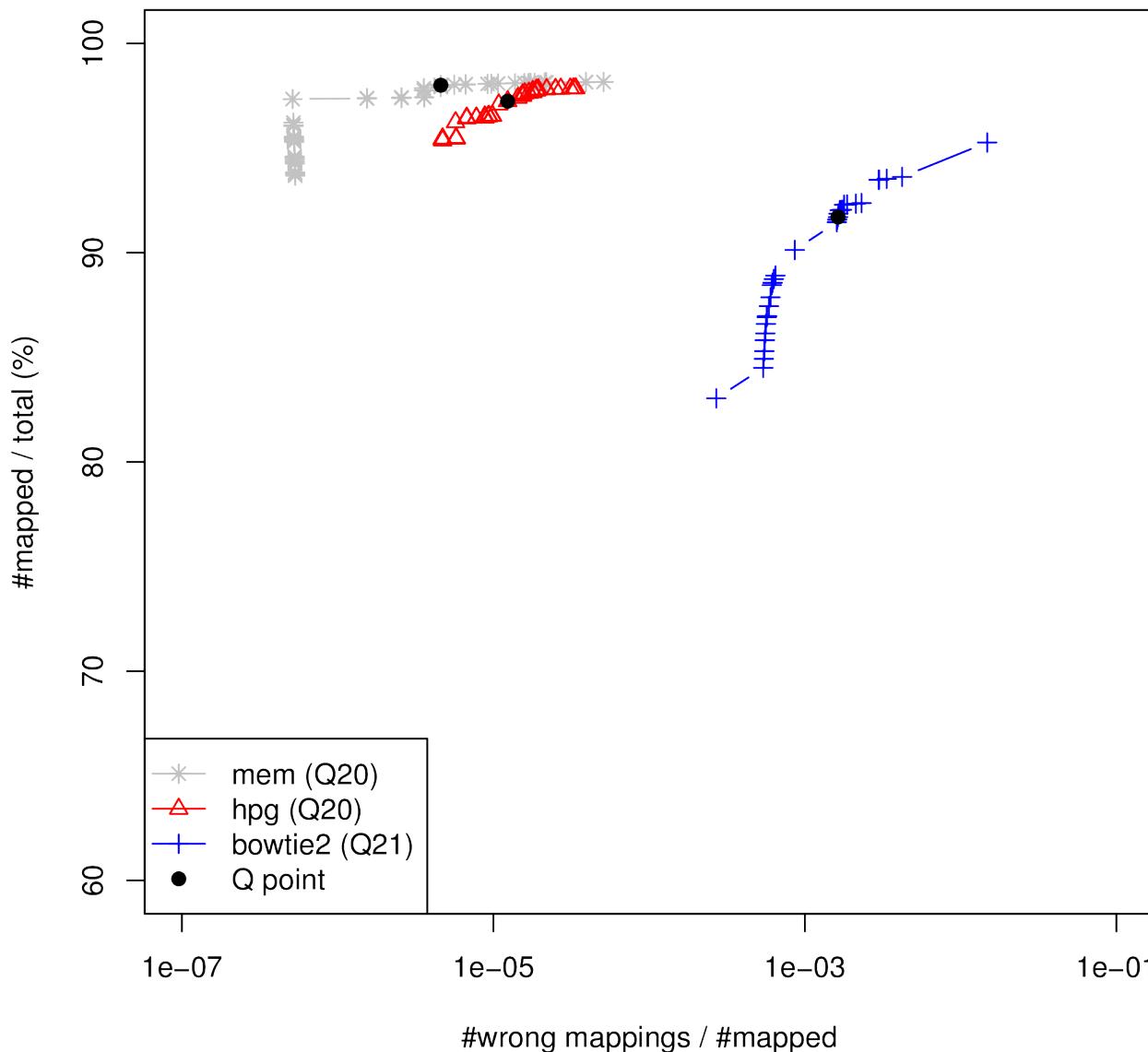
## Why another NGS read aligner

- There are more than 70 aligners
  - [http://wwwdev.ebi.ac.uk/fg/hts\\_mappers/](http://wwwdev.ebi.ac.uk/fg/hts_mappers/)
- This project began as a *experimental project*, could a well designed algorithm implemented using modern HPC technologies speed up NGS data mapping?
  - “*It's hardware that makes a machine fast. It's software that makes a fast machine slow*”  
Craig Bruce
- Focus on **performance** and **sensitivity**. Current software is designed for standard workstations, but **new projects are processed in clusters**. During these two years we have developed one of the fastest and one of the most sensitive NGS aligners for DNA and RNA by using different HPC technologies
- Some computing papers
  - <http://dl.acm.org/citation.cfm?id=2223945>
  - <http://arxiv.org/abs/1304.0681>
- Bioinformatics paper:  
<http://bioinformatics.oxfordjournals.org/content/early/2014/09/01/bioinformatics.btu553.long>

# HPG Aligner

DNA aligner results, quality similar to new BWA-MEM

Comparison: base error: 0.1%, mutation: 0.1% (125 bp length)



Similar sensitivity than BWA-MEM but **3-5x times faster**

Other features:

- Adaptor support
- INDEL realignment
- Base recalibration

# HPG Aligner

## Main and coming features

- Part of the HPG Aligner suite (<http://www.opencb.org/technologies/hpg>) with other tools: *hpg-fastq*, *hpg-bam* and *hpg-aligner*
- Usability: the two aligners under the same binary, only one execution is needed to generate the BAM output file, faster index creator, multi-core implementation
- Focused on providing the best sensitivity and the best performance by using HPC technologies: multicore, SSE4/AVX2, GPUs, MPI, ...
- Coming features
  - Smith-Waterman in AVX2 implementation and Xeon PHI
  - Initial support for **GRCh38** with ALTs
  - BS-seq released: for methylation analysis (being tested)
  - **Apache Hadoop implementation** will allow to run it in a distributed environment
  - **New SA index** (not BWT) in version 2.0 for performance improvements, more memory needed but first results show a speed-up of 2-4x. i.e. more than 10 billion exact reads of 100nt aligned in 1 hour in a 12-cores node.

# HPG Variant

## A suite of tools for variant analysis

- *HPG Variant*, a suite of tools for HPC-based genomic variant analysis
  - **VARIANT** = **VAR**iant **AN**alysis **T**ool
- Three tools are already implemented: **vcf**, **gwas** and **effect**. Implemented using *OpenMP*, SSE/AVX, Nvidia CUDA and *MPI* for large clusters. Hadoop version coming soon.
- **VCF**: C library and tool: allows to analyze large VCFs files with a low memory footprint: stats, filter, split, **merge**, ... (*paper in preparation*)
  - Example: *hpg-variant vcf –stats –vcf-file ceu.vcf*
- **GWAS**: suite of tools for gwas variant analysis (~*Plink*)
  - Genetic tests: association, TDT, Hardy-Weinberg, ...
  - **Epistasis**: HPC implementation with SSE4 and MPI, 2-way 420K SNPs epistasis in 9 days in a 12-core node.
  - Example: *hpg-variant gwas –tdt –vcf-file tumor.vcf*
- **EFFECT**: A CLI and web application, it's a cloud-based genomic variant **effect** predictor tool has been implemented (<http://variant.bioinfo.cipf.es>, published in NAR 2012)

# HPG BigData

## A suite of tools for working in genomics with Hadoop

- New project to provide big data analysis in genomics based on Hadoop and Spark
- Main features:
  - Hadoop-based implementation of data converters (avro, parquet) and some bioinformatic tools (ie. samtools)
  - Simple VCF indexing for HBase, Hive and Impala developed
  - C code embedded using JNI to speed-up processing
  - Java libraries and command line implemented
- Many new NGS data analysis being implemented with MapReduce and Spark
- Available at <https://github.com/opencb/hpg-bigdata>

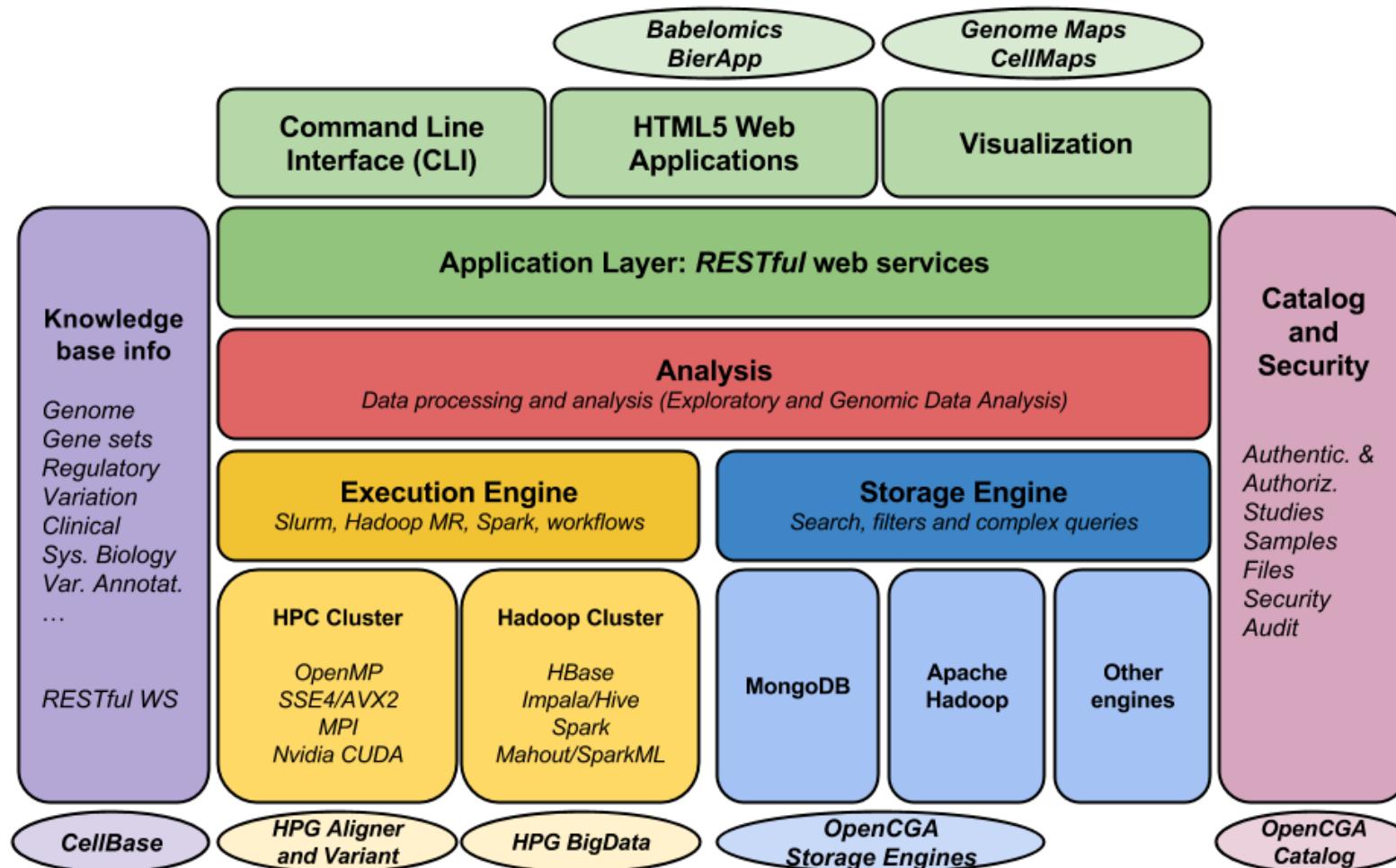
# OpenCGA

## Overview and goals

Open-source Computational Genomics Analysis (**OpenCGA**) aims to provide to researchers and clinicians a **high performance and scalable solution** for genomic big data processing and analysis

**OpenCGA** is built on OpenCB: CellBase, Genome Maps, Cell Maps, HPG Aligner, HPG BigData, Variant annotation

Project at GitHub: <https://github.com/opencb/opencga>

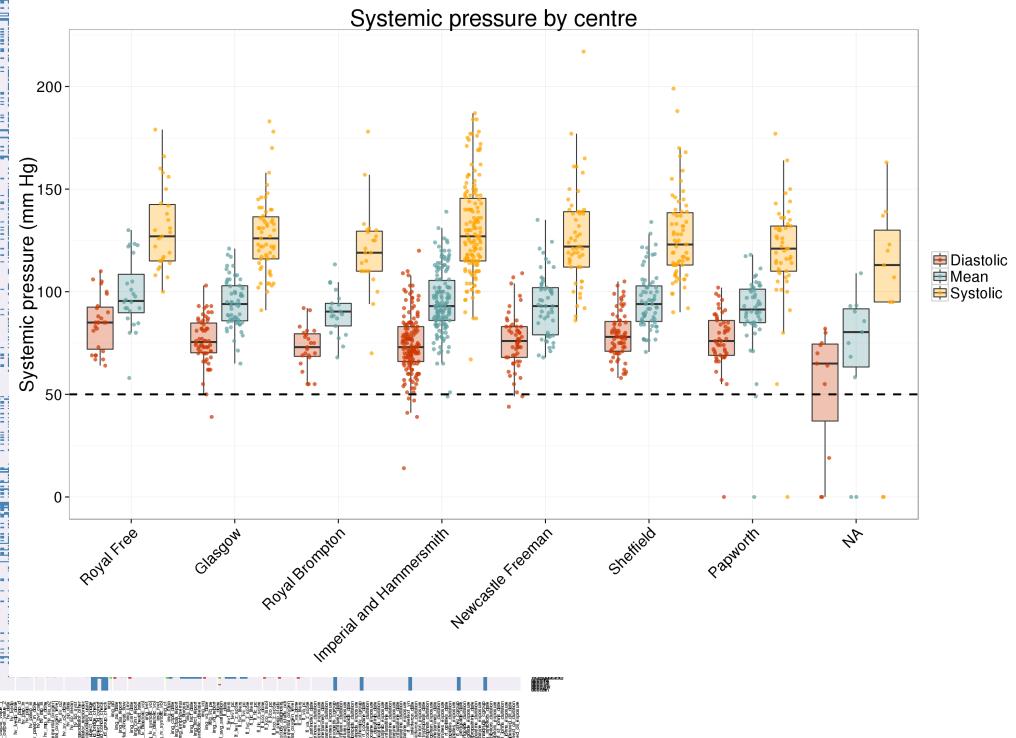
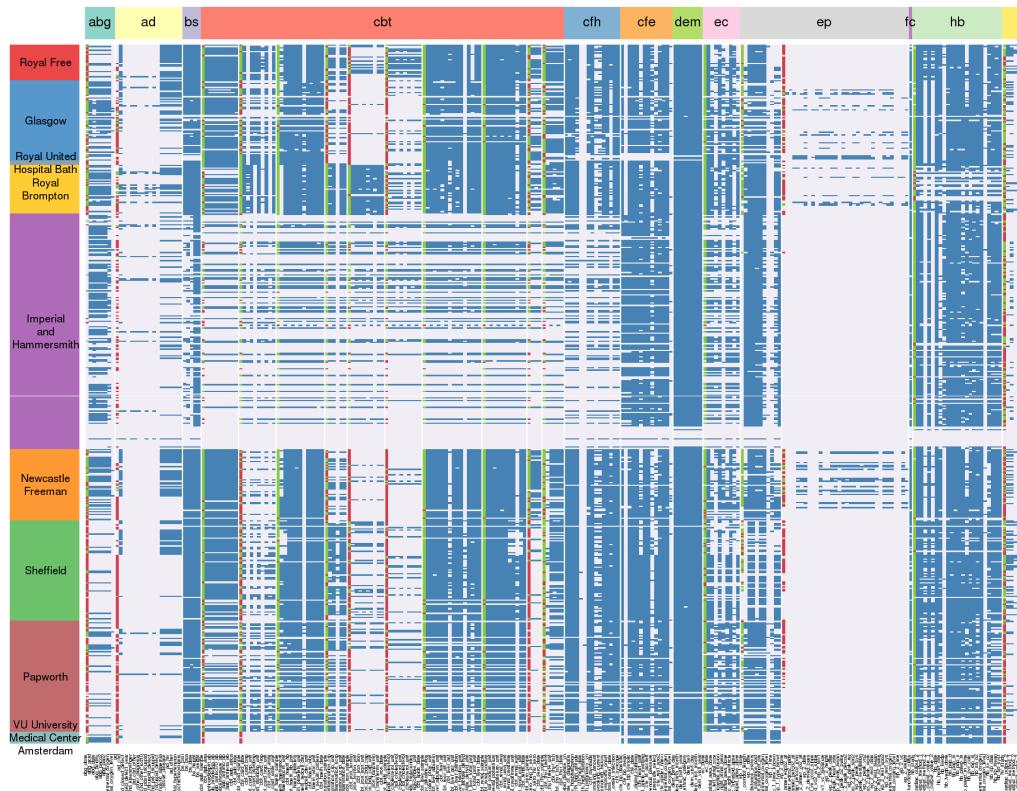


# OpenCGA

## Metadata Catalog

- **OpenCGA Catalog** provides user authentication, authorization, *sample annotation*, file and job tracking, audit, ...
- Allow to the different storage engines to perform optimizations
- **Sample annotations** is one of the main feature:
  - Allow complex queries and aggregations
  - Allow to detect bias and other problems with the data

R prototypes, interactive web-based plots being implemented

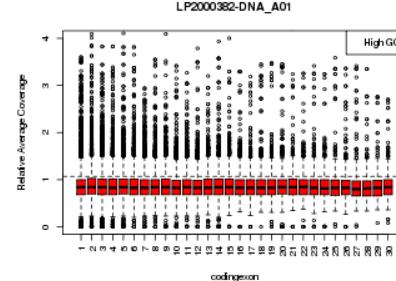
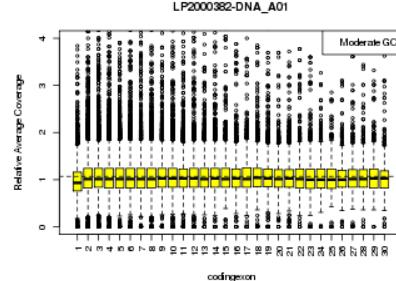
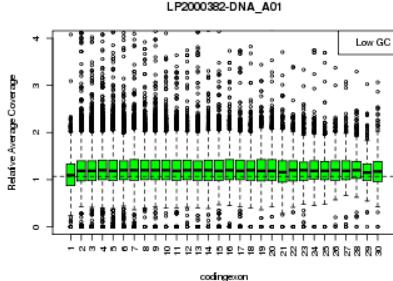
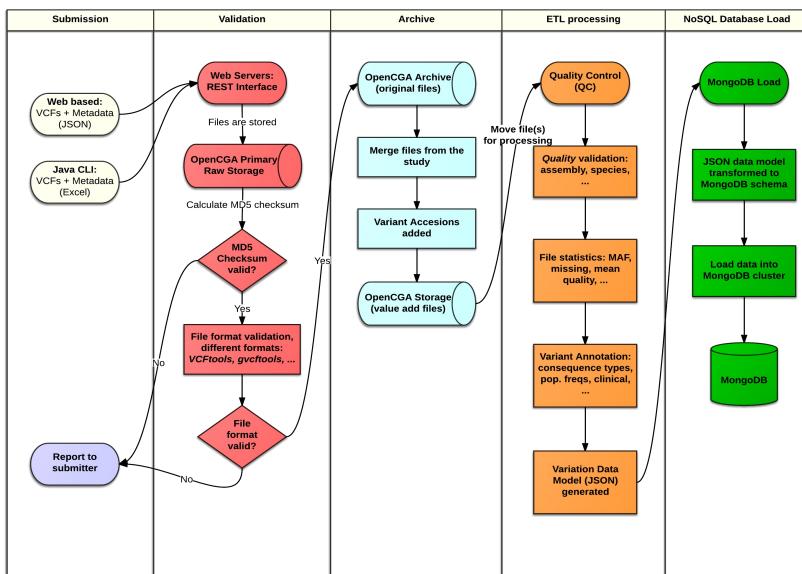


# OpenCGA

## Storage Engines

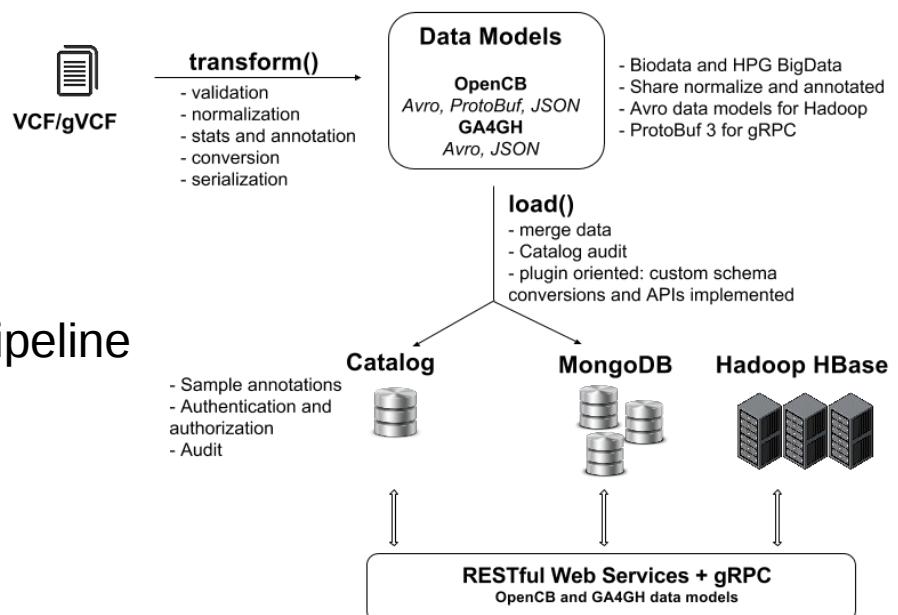
**OpenCGA Storage** provides a *pluggable* Java framework for storing and querying alignment and variant data

- Data is processed, normalized, **annotated and indexed**, also some **stats are precomputed and indexed**



Two default implementations: **MongoDB** and **Hadoop** for huge performance and scalability (**~500k inserts/second**)

Highly customizable and easy to extend (ie. only two Java classes are needed to be implemented)



Precomputed data allows interactive analysis

# HPCS-Dell Genomic Collaboration

New projects coming soon

- New collaboration between HPCS and Dell to port some big data processing and algorithms to new Dell Statistica software
- Dell Statistica provides an amazing platform for data analysis in many different areas, genomics can take advantage of this
- Proof of concept being developed for next year

# Acknowledgements

- Bosses
  - **Paul Calleja**, Director of HPC Service, University of Cambridge
  - **Augusto Rendon**, Director of Bioinformatics of Genomics England
- Team and core OpenCB developers:
  - Jacobo Coll (*Genomics England*)
  - Javier Lopez (*EMBL-EBI*)
  - Joaquin Tarraga (*CIPF*)
  - Cristina González (*EMBL-EBI*)
  - Jose M. Mut (*EMBL-EBI*)
  - Francisco Salavert (*CIPF*)
  - Matthias Haimel (*Addenbrooke's*)
  - Marta Bleda (*Addenbrooke's*)
- Collaborators
  - **Joaquin Dopazo** (*CIPF*)
  - Justin Paschall (*EMBL-EBI*)
  - Stefan Gräf (*Addenbrooke's*)
  - UJI: Enrique Quintana, Héctor Martinez, Sergio Barrachina, Maribel Castillo
- Dell
- Cloudera
  - Tom White
  - Uri Laserson