

Next Generation Sequencing

and

Bioinformatics Analysis Pipelines

Adam Ameur
National Genomics Infrastructure
SciLifeLab Uppsala
adam.ameur@igp.uu.se

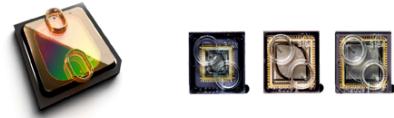
Today's lecture

- Management of NGS data at NGI/SciLifeLab
- Examples of analysis pipelines:
 - Human exome & whole genome sequencing
 - Assembly using long reads
 - Clinical routine sequencing

illumina®



ThermoFisher
SCIENTIFIC
life
ion torrent

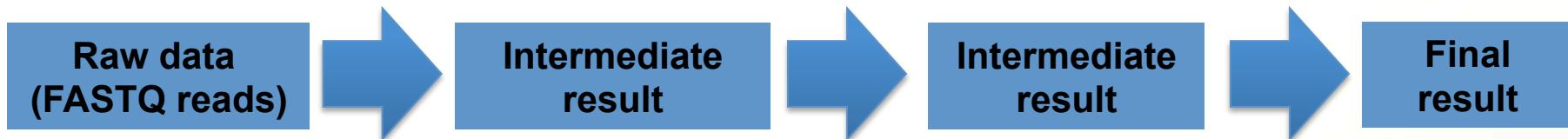


PB
PACIFIC
BIOSCIENCES®

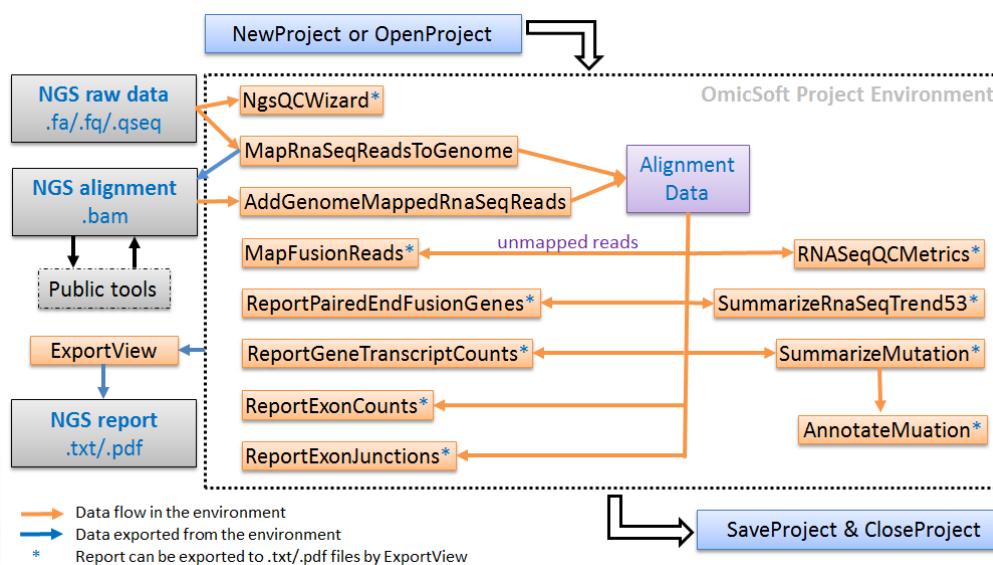


What is an analysis pipeline?

- Basically just a number of steps to analyze data



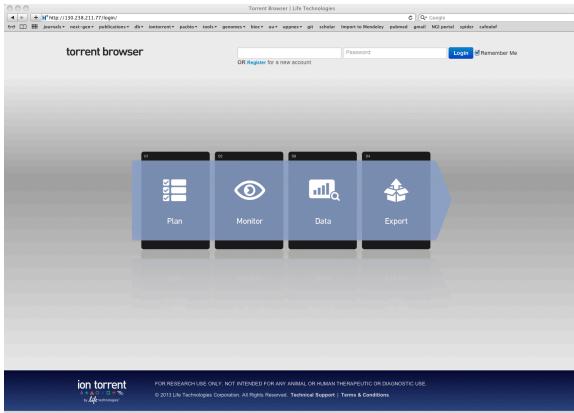
- Pipelines can be simple or very complex...



Some analysis pipelines for NGS data

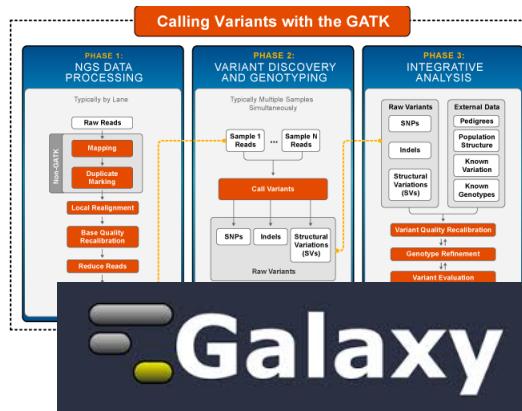
Ion Torrent

Torrent Suite Software



Illumina

GATK, Galaxy,...



PacBio

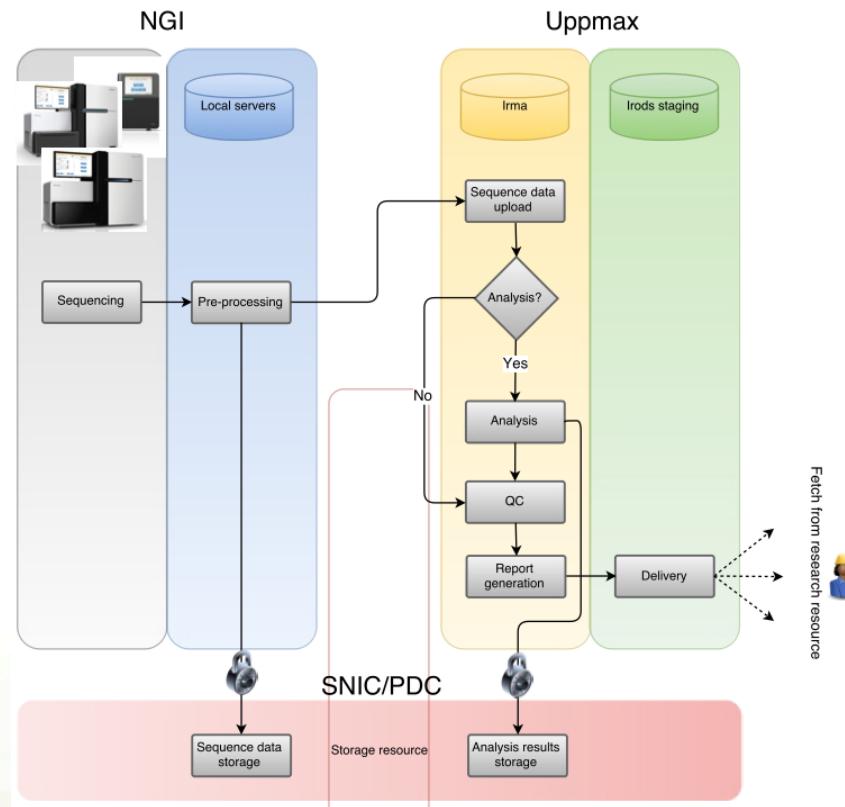
SMRT analysis portal



- Enables variant calling, de novo assembly, RNA expression analyses, ...
- Many other tools exists, also from commercial vendors

Data processing at NGI

- Raw data from is processed in automated pipelines
- Delivered to user accounts at UPPNEX



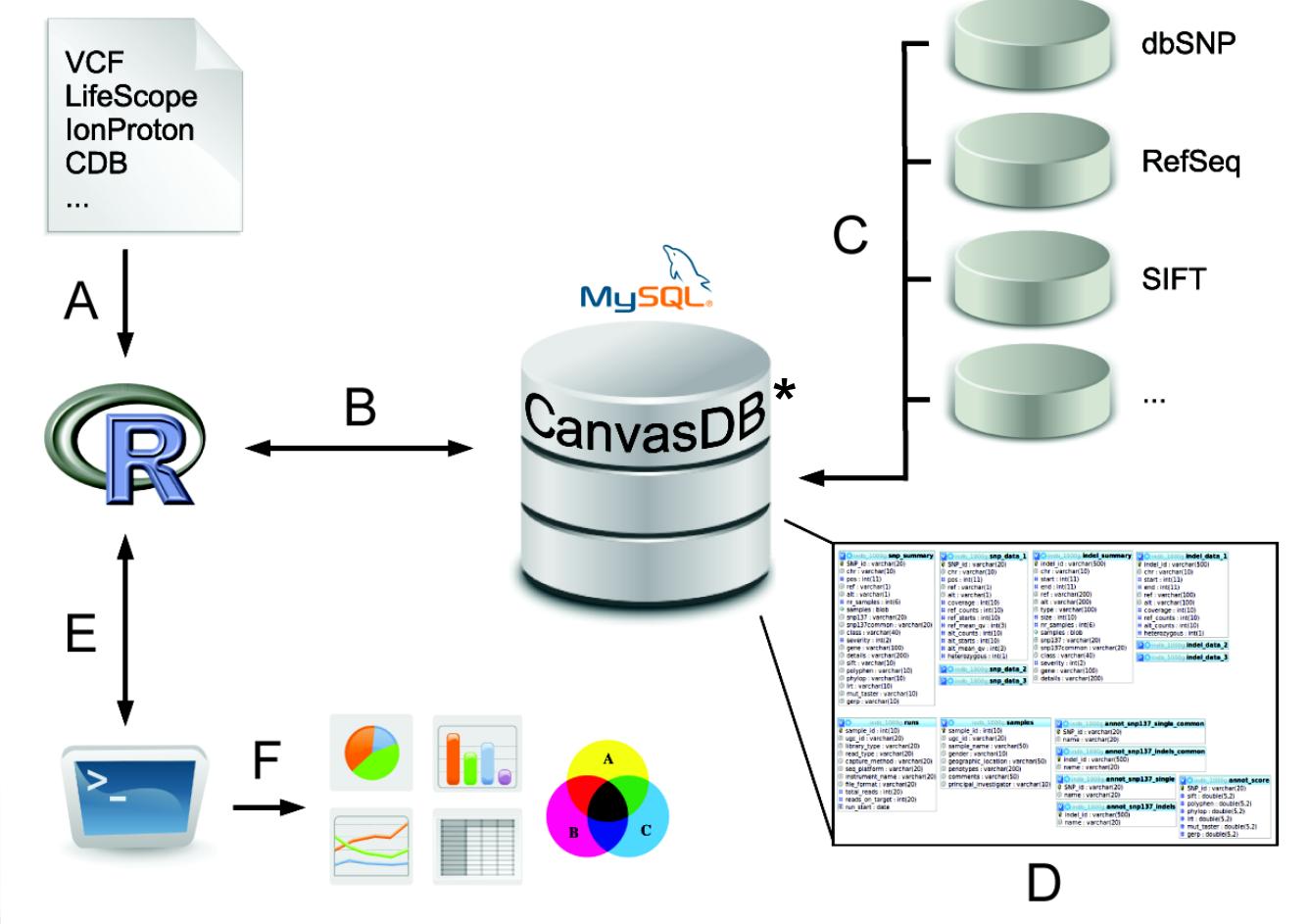
In-house development of pipelines

- In some cases NGI develops own pipelines
- But only when we see a need for a specific analysis

Some examples follows:

- I. Building a local variant databases (WES/WGS)
- II. Assembly of genomes using long reads
- III. Clinical sequencing – Leukemia Diagnostics

Example I: Computational infrastructure for exome-seq data

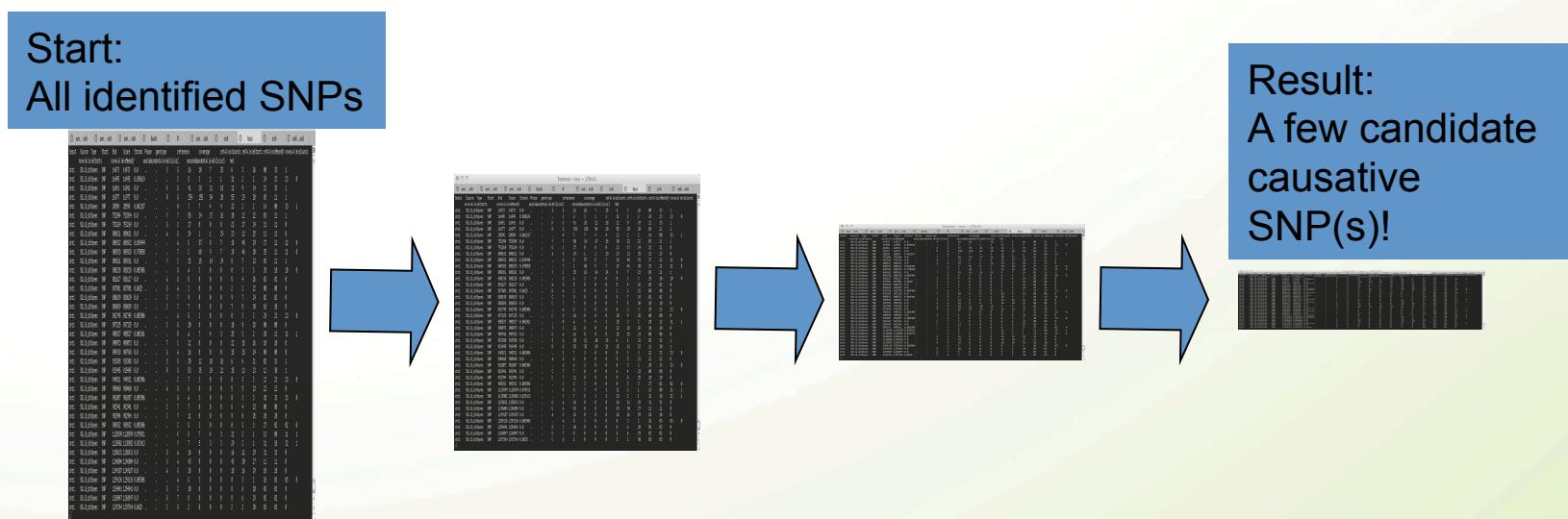


Background: exome-seq

- Main application of exome-seq
 - Find disease causing mutations in humans
- Advantages
 - Allows investigate all protein coding sequences
 - Possible to detect both SNPs and small indels
 - Low cost (compared to WGS)
 - Possible to multiplex several exomes in one run
 - Standardized work flow for data analysis
- Disadvantage
 - All genetic variants outside of exons are missed (~98%)

How to analyze this?

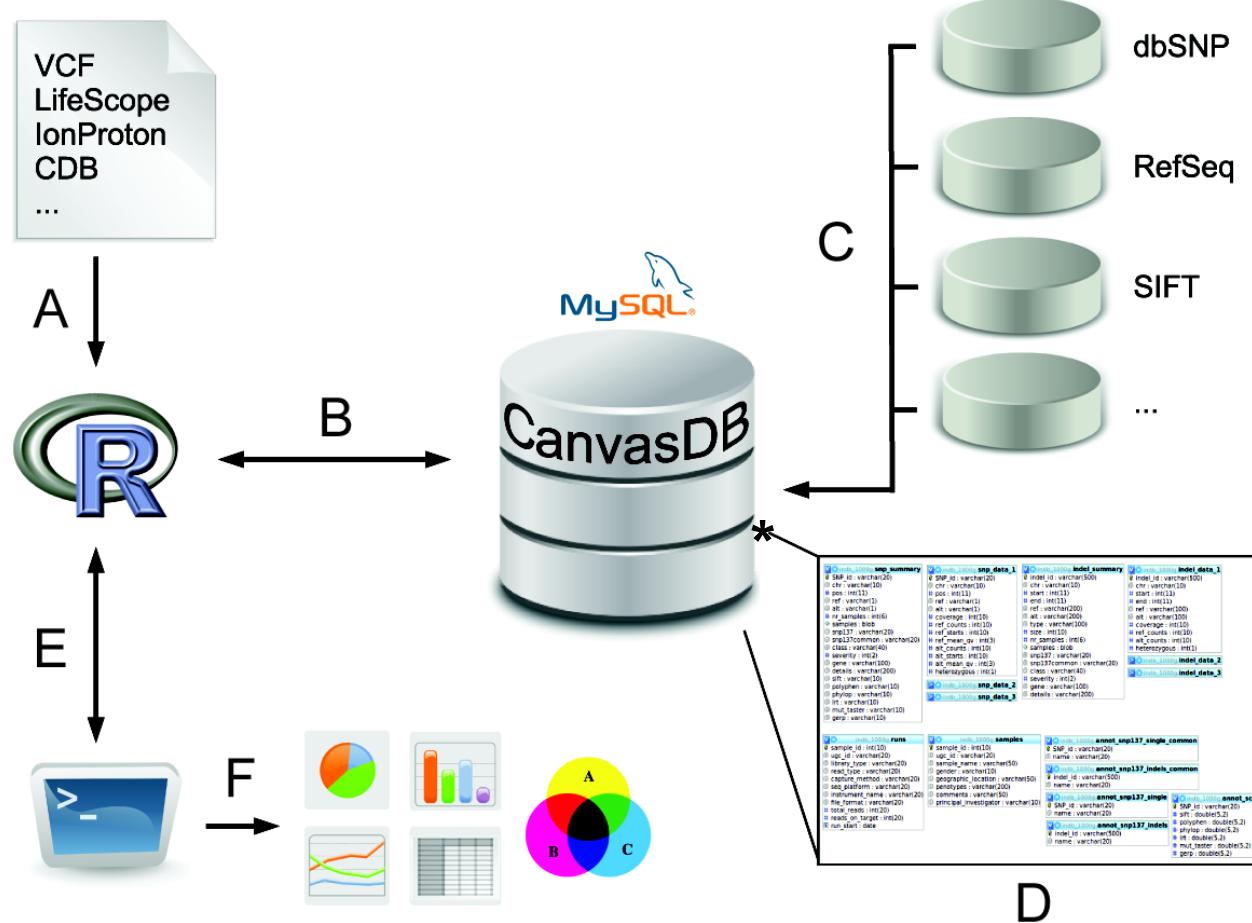
- Traditional analysis - A lot of filtering!
 - Typical filters
 - Focus on rare SNPs (not present in dbSNP)
 - Remove FPs (by filtering against other exomes)
 - Effect on protein: non-synonymous, stop-gain etc
 - Heterozygous/homozygous
 - This analysis can be automated (more or less)



Why is this not optimal?

- Drawbacks
 - Work on one sample at time
 - Difficult to compare between samples
 - Takes time to re-run analysis
 - When using different parameters
 - No standardized storage of detected SNPs/indels
 - Difficult to handle 100s of samples
- Better solution
 - A database oriented system
 - Both for data storage and filtering analyses

Analysis: In-house variant database

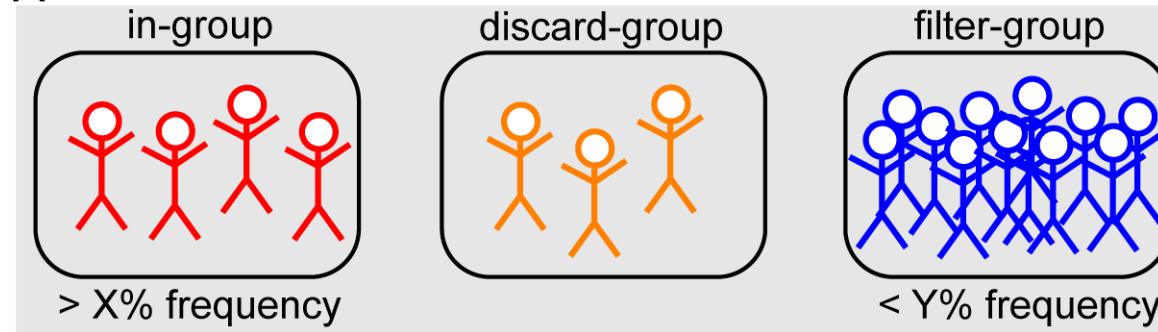


*CANdidate Variant Analysis System and Data Base

Ameur et al., Database Journal, 2014

CanvasDB - Filtering

A



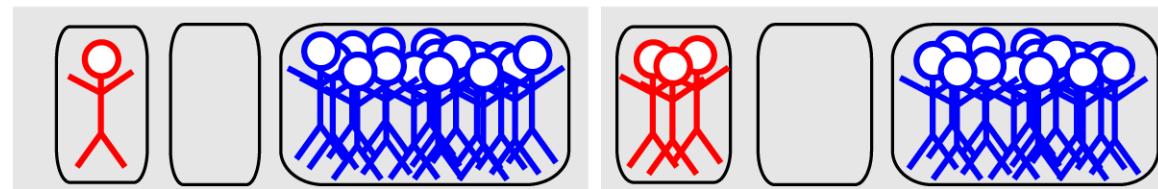
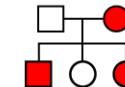
B

parent-offspring trio



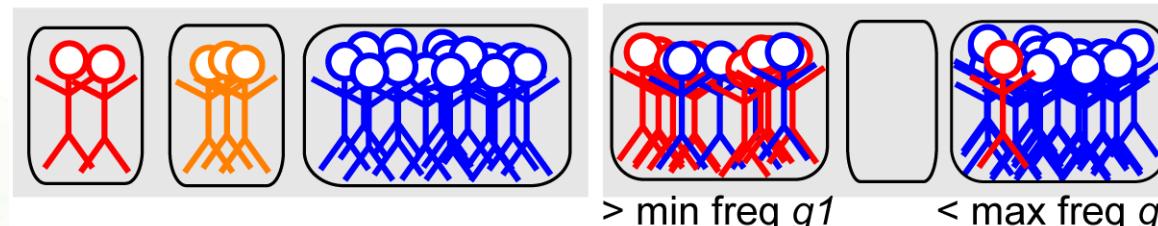
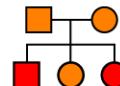
C

dominant variant



D

recessive variant



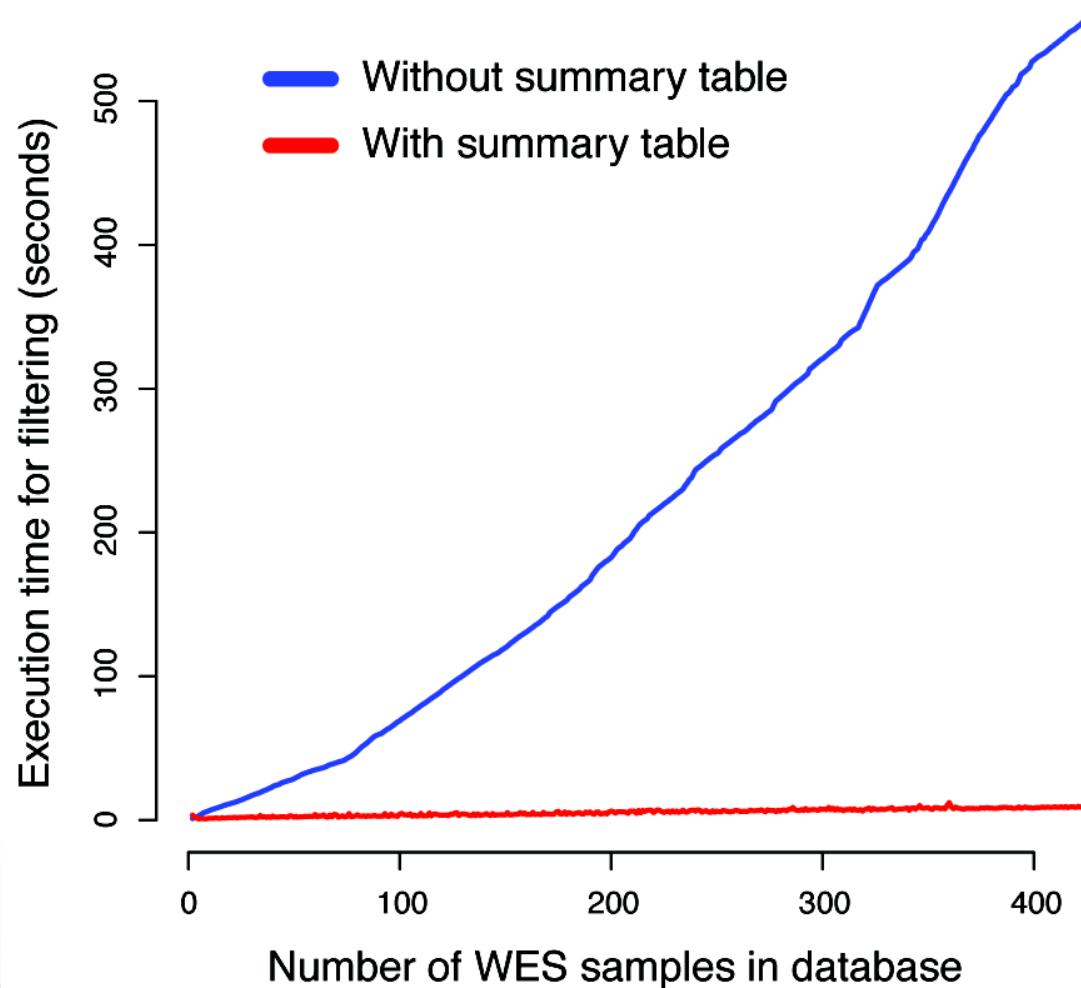
E

comparing groups



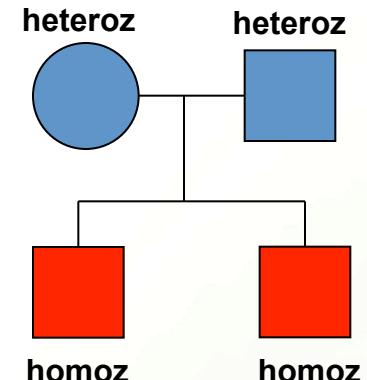
CanvasDB - Filtering speed

- Rapid variant filtering, also for large databases



An exome-seq project

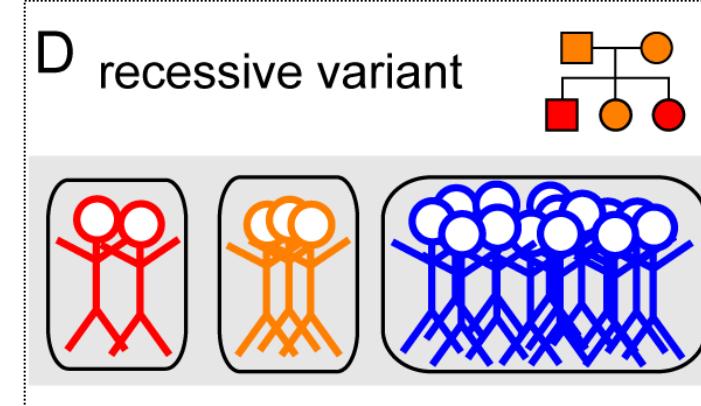
- Hearing loss: 2 affected brothers
 - Likely a rare, recessive disease
 - => Shared homozygous SNPs/indels
- Sequencing strategy
 - TargetSeq exome capture
 - One sample per PI chip



| | | |
|--------------|---------------|---------------|
| nr reads | (% mapped) | 76M-89M (97%) |
| mapped reads | (% on target) | 73M-88M (83%) |
| SNPs | (% in dbSNP) | 85k-93k (93%) |
| Indels | (% in dbSNP) | 5k-6k (48%) |

Filtering analysis

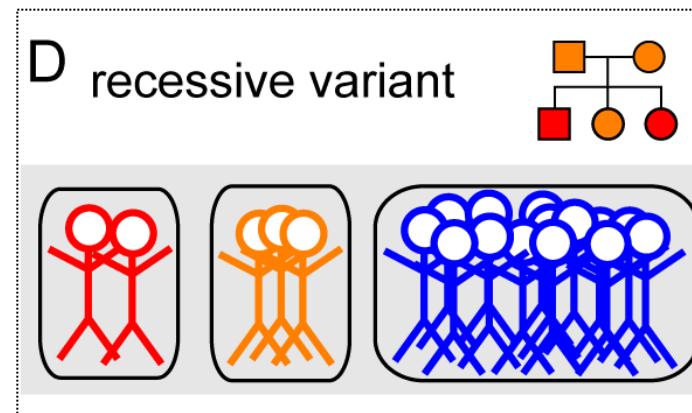
- *CanvasDB* filtering for a variant that is...
 - rare
 - at most in 1% of ~700 exomes
 - shared
 - found in both brothers
 - homozygous
 - in brothers, but in no other samples
 - deleterious
 - non-synonymous, frameshift, stop-gain, splicing, etc..



```
> cand <- filterRecessive(c("up_001_1", "up_001_2"), outfile="cand.txt")
Total time for filtering: 27.012s
```

Filtering results

- Homozygous candidates
 - 2 SNPs
 - stop-gain in *STRC*
 - non-synonymous in *PCNT*
 - 0 indels
- Compound heterozygous candidates (lower priority)
 - in 15 genes



| sample_name | class | chr | pos | ref | alt | snp137 | gene | ref_counts | alt_counts |
|-------------|---------------|-------|----------|-----|-----|-------------|------|------------|------------|
| up_001_1 | stopgain | chr15 | 43896948 | G | A | rs144948296 | STRC | 3 | 58 |
| up_001_2 | stopgain | chr15 | 43896948 | G | A | rs144948296 | STRC | 5 | 55 |
| up_001_1 | nonsynonymous | chr21 | 47808772 | G | A | rs35044802 | PCNT | 0 | 21 |
| up_001_2 | nonsynonymous | chr21 | 47808772 | G | A | rs35044802 | PCNT | 1 | 14 |

=> Filtering is fast and gives a short candidate list!

STRC - a candidate gene

STRC

From Wikipedia, the free encyclopedia

Stereocilin is a protein that in humans is encoded by the *STRC* gene.^{[1][2][3]}

This gene encodes a protein that is associated with the hair bundle of the sensory hair cells in the inner ear. The hair bundle is composed of stiff **microvilli** called **stereocilia** and is involved with **mechanoreception** of sound waves. This gene is part of a tandem duplication on chromosome 15; the second copy is a **pseudogene**. Mutations in this gene cause autosomal recessive **non-syndromic deafness**.^[3]

=> Stop-gain in STRC is likely to cause hearing loss!

IGV visualization: Stop gain in STRC

Unrelated sample

[0 - 336]

Brother #1

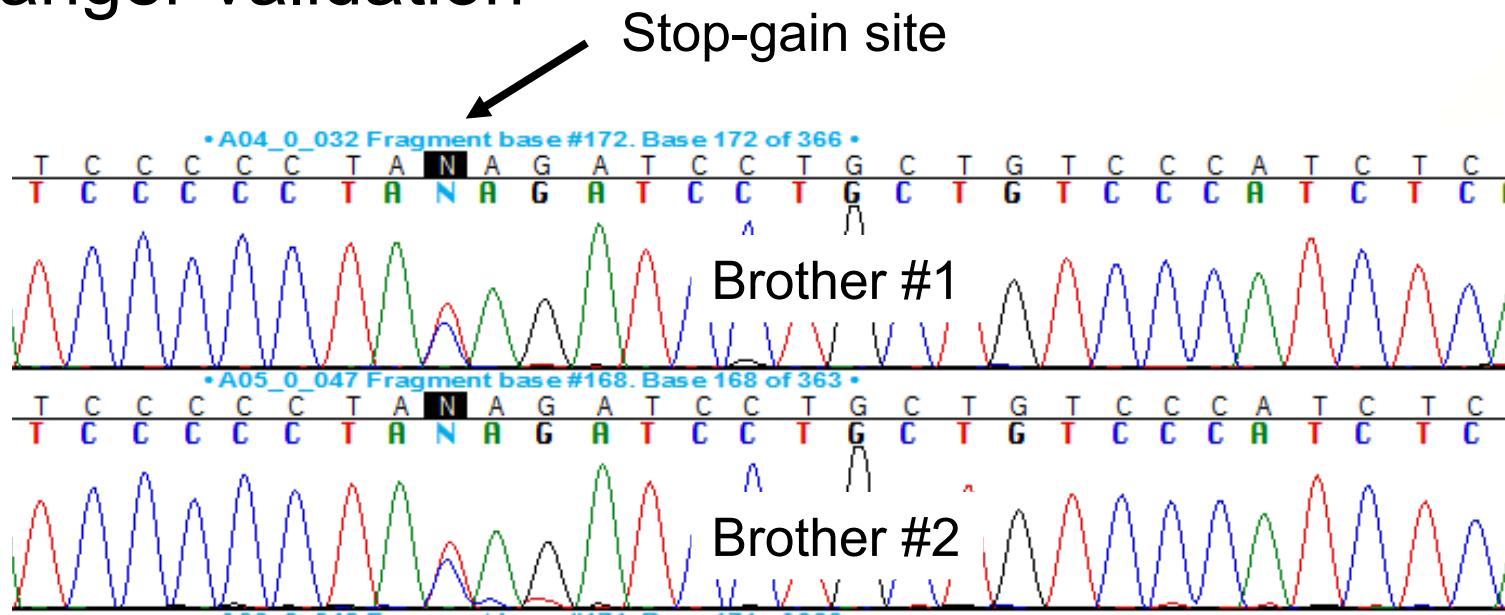
[0 - 331]

Brother #2

G A G A T G G G A C A G C A G G A T C T G T A G G G A T C T G T C G T C G T G T
L H S L I Q L P I Q R T

STRC, validation by Sanger

- Sanger validation



- Does not seem to be homozygous..
 - Explanation: difficult to sequence STRC by Sanger
 - Pseudo-gene with very high similarity
- New validation showed mutation is homozygous!!

CanvasDB – some success stories

Solved cases, exome-seq - Niklas Dahl/Joakim Klar

| | |
|-------------------------------|---------------|
| <i>Neuromuscular disorder</i> | <i>NMD11</i> |
| <i>Arthrogryposis</i> | <i>SKD36</i> |
| <i>Lipodystrophy</i> | <i>ACR1</i> |
| <i>Achondroplasia</i> | <i>ACD2</i> |
| <i>Ectodermal dysplasia</i> | <i>ED21</i> |
| <i>Achondroplasia</i> | <i>ACD9</i> |
| <i>Ectodermal dysplasia</i> | <i>ED1</i> |
| <i>Arythroderma</i> | <i>AV1</i> |
| <i>Ichthyosis</i> | <i>SD12</i> |
| <i>Muscular dystrophy</i> | <i>DMD7</i> |
| <i>Neuromuscular disorder</i> | <i>NMD8</i> |
| <i>Welanders myopathy (D)</i> | <i>W</i> |
| <i>Skeletal dysplasia</i> | <i>SKD21</i> |
| <i>Visceral myopathy (D)</i> | <i>D:5156</i> |
| <i>Ataxia telangiectasia</i> | <i>MR67</i> |
| <i>Exostosis</i> | <i>SKD13</i> |
| <i>Alopecia</i> | <i>AP43</i> |
| <i>Epidermolysis bullosa</i> | <i>SD14</i> |
| <i>Hearing loss</i> | <i>D:9652</i> |

Success rate >80% for recent Proton projects!

CanvasDB - Availability

- CanvasDB system freely available on GitHub!

Installation of the CanvasDB system

This section describes how to download and install CanvasDB on your local computer. Make sure that [MySQL](#), [R](#) and [ANNOVAR](#) are running on your computer before starting the installation.

Step 1. Download code from github

```
$ git clone https://github.com/UppsalaGenomeCenter/CanvasDB.git  
$ cd CanvasDB
```

Step 2. Set the current path to 'rootDir' in canvasDB.R

Next Step: Whole Genome Sequencing



Capacity of HiSeq X Ten: 320 whole human genomes/week!!!

⇒ More work on pipelines and databases needed!

Whole Genome Sequencing Projects



Business & Financial News, Breaking US & International News | Reuters.com

» Print

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to colleagues, clients or customers, use the Reprints tool at the top of any article or visit: www.reutersreprints.com.

UK to launch 100,000 genomes project as Obama backs DNA drive

Fri, Jan 30 2015

- * Genomes project starts recruitment in England on Feb. 2
- * U.S. to analyse genetic data from 1 million Americans

By Ben Hirschler

LONDON, Jan 30 (Reuters) - Gene research is getting a boost on both sides of the Atlantic, with scientists in England set to launch a project on Feb. 2 to analyse 100,000 entire human genomes and U.S. President Barack Obama backing a big new DNA data drive.

The twin projects show the accelerating work by researchers to understand the underlying basis of diseases and develop medicines targeted to the genetic profile of individual patients.

Obama will announce the U.S. plan to analyse genetic information from more than 1 million American volunteers on Friday as a central part of an initiative to promote so-called precision medicine, officials said.

The 100,000 genomes project in England, meanwhile, was first unveiled by the British government two years ago -- but the 11 centres charged with collecting samples will only begin full-scale recruitment from next week. The aim is to complete the programme by the end of 2017.

Such large-scale genomic research has become possible because the cost of genome sequencing has plummeted in recent years to around \$1,000 per genome. That is a far cry from 15 years ago when it cost some \$3 billion to get the first human genome.

In the case of the British project, all the sequencing will be carried out by U.S. biotech company Illumina, which has pioneered fast and cheap technology to read genetic code.

The 100,000 genomes project is focusing on patients with rare diseases, and their families, as well as people with common cancers. The idea is to tease out the common drivers of disease to help develop better drugs and diagnostic tests.

In addition to helping doctors understand more about disease, the government also hopes the scheme will make the state-run National Health Service a world leader in science and boost Britain's life sciences industry.

The project will actually recruit around 75,000 participants, rather than 100,000, since people with cancer will provide two genomes -- one derived from the healthy cells in their body and one from their tumour. By comparing the two, experts hope to find the exact genetic changes causing cancer. (Editing by Susan Thomas)

© Thomson Reuters 2015. All rights reserved. Users may download and print extracts of content from this website for their own personal and non-commercial use only. Reproduction or redistribution of Thomson Reuters content, including by framing or similar means, is expressly prohibited without the prior written consent of Thomson Reuters. Thomson Reuters and its logo are registered trademarks or trademarks of the Thomson Reuters group of companies around the world.

Thomson Reuters journalists are subject to an Editorial Handbook which requires fair presentation and disclosure of relevant interests.

This copy is for your personal, non-commercial use only. To order presentation-ready copies for distribution to colleagues, clients or customers, use the Reprints tool at the top of any article or visit: www.reutersreprints.com.

PERSPECTIVE

Big Data: Astronomical or Genomical?

Table 1. Four domains of Big Data in 2025. In each of the four domains, the projected annual storage and computing needs are presented across the data lifecycle.

| Data Phase | Astronomy | Twitter | YouTube | Genomics |
|--------------|--|-----------------------------|--|--|
| Acquisition | 25 zetta-bytes/year | 0.5–15 billion tweets/year | 500–900 million hours/year | 1 zetta-bases/year |
| Storage | 1 EB/year | 1–17 PB/year | 1–2 EB/year | 2–40 EB/year |
| Analysis | In situ data reduction | Topic and sentiment mining | Limited requirements | Heterogeneous data and analysis |
| | Real-time processing | Metadata analysis | | Variant calling, ~2 trillion central processing unit (CPU) hours |
| | Massive volumes | | | All-pairs genome alignments, ~10,000 trillion CPU hours |
| Distribution | Dedicated lines from antennae to server (600 TB/s) | Small units of distribution | Major component of modern user's bandwidth (10 MB/s) | Many small (10 MB/s) and fewer massive (10 TB/s) data movement |

doi:10.1371/journal.pbio.1002195.t001

Analysis of WGS data @ SciLifeLab

We have a working group for WGS at SciLifeLab!

wgs-toolbox@scilifelab.se

Contacts with Genomics England initiated for analyses

Genomics
england



The SciLifeLab Human WGS Initiative

- WGS of patient cohorts (n=10,000 ind/year)
- Genetic Variant Database for the Swedish Population (n=1000)

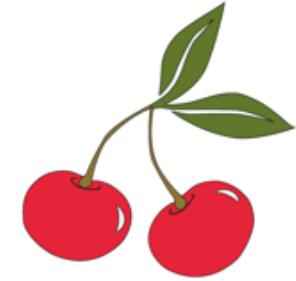


The Swedish Genetic Variant Project

- A. Identify a cohort that reflects the genetic structure of the Swedish population
- B. Generate WGS data using short- and long-read MPS technologies
- C. Establish a user-friendly database to make information available to the research community (association analyses) and clinical genetics laboratories.

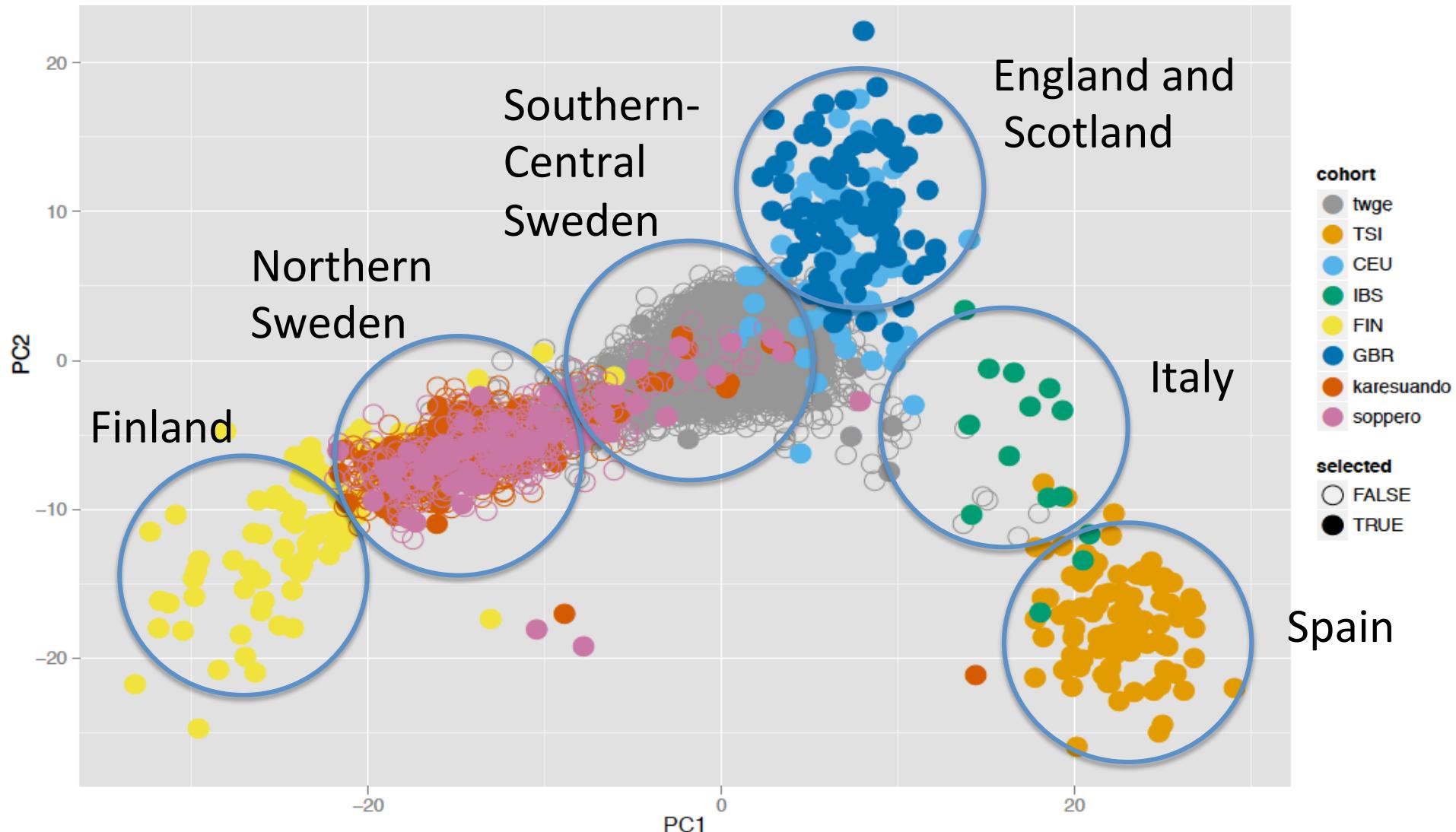
Twin Registry samples used as control cohort

- Inclusion based on twinning
- Distribution like population density
- General population-prevalence of disease
- 10,000 individuals have been analysed with SNP arrays



**Identify 1,000 individuals based on genetic structure
and diversity across Sweden**

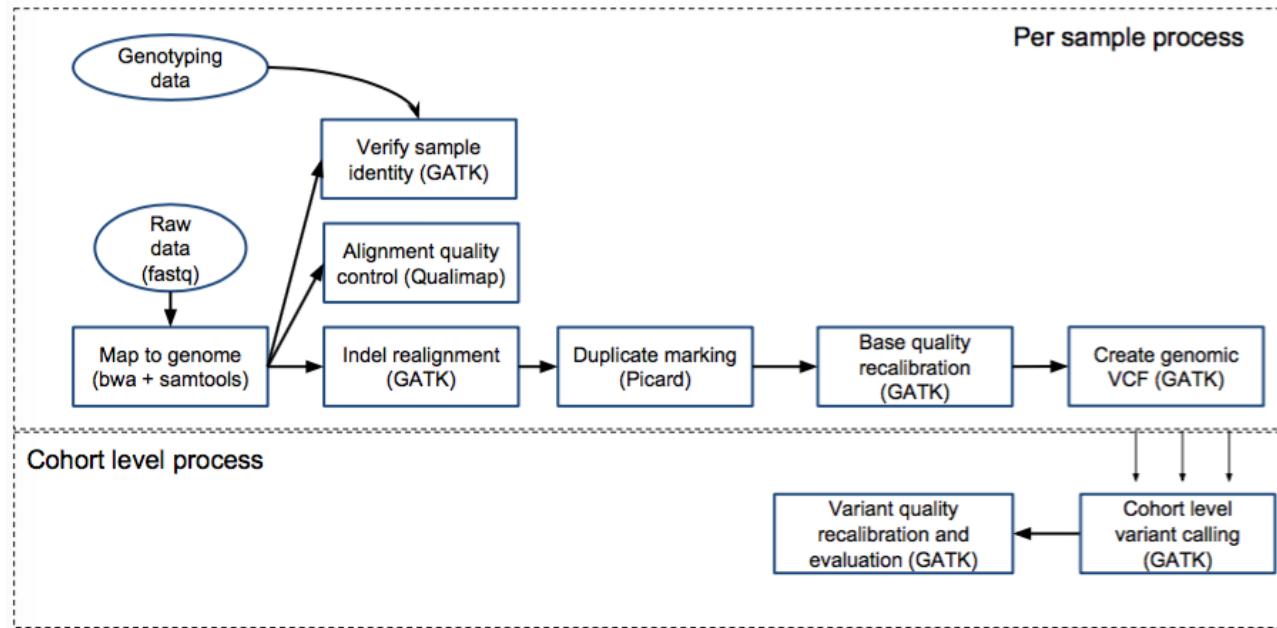
Principal components of European samples from 1,000 genomes project and 10,000 Swedish samples



Illumina WGS of Swedish control cohort

Step 1: 30X Illumina data of the 1,000 individuals

Step 2: Mapping and variant calling



Step 3: Making genotype frequencies available for download

Making frequency data available

SweFreq

Data Beacon

ExAC Browser

Login

SweGen Variant Frequency Database

This server hosts whole-genome variant frequencies for 1000 Swedish individuals generated within the SweGen project. The frequency data is intended to be used as a resource for the research community and clinical genetics laboratories. Individual positions in the genome can be viewed using the Data Beacon or ExAC Browser by clicking the links above. To access the variant frequency file you need to register.

Please note that the 1000 individuals included in the SweGen project represent a cross-section of the Swedish population and that no disease information has been used for the selection. The frequency data may therefore include genetic variants that are associated with, or causative of, disease.

We request that any use of data from the SweGen project cite this preprint on bioRxiv.



SciLifeLab

NATIONAL
EXAC
GENOMICS
INFRASTRUCTURE

NBIS

elixir
SWEDEN

Aggregated frequencies now available from ***swefreq.nbis.se!***

Example II:

Assembly of genomes using Pacific Biosciences

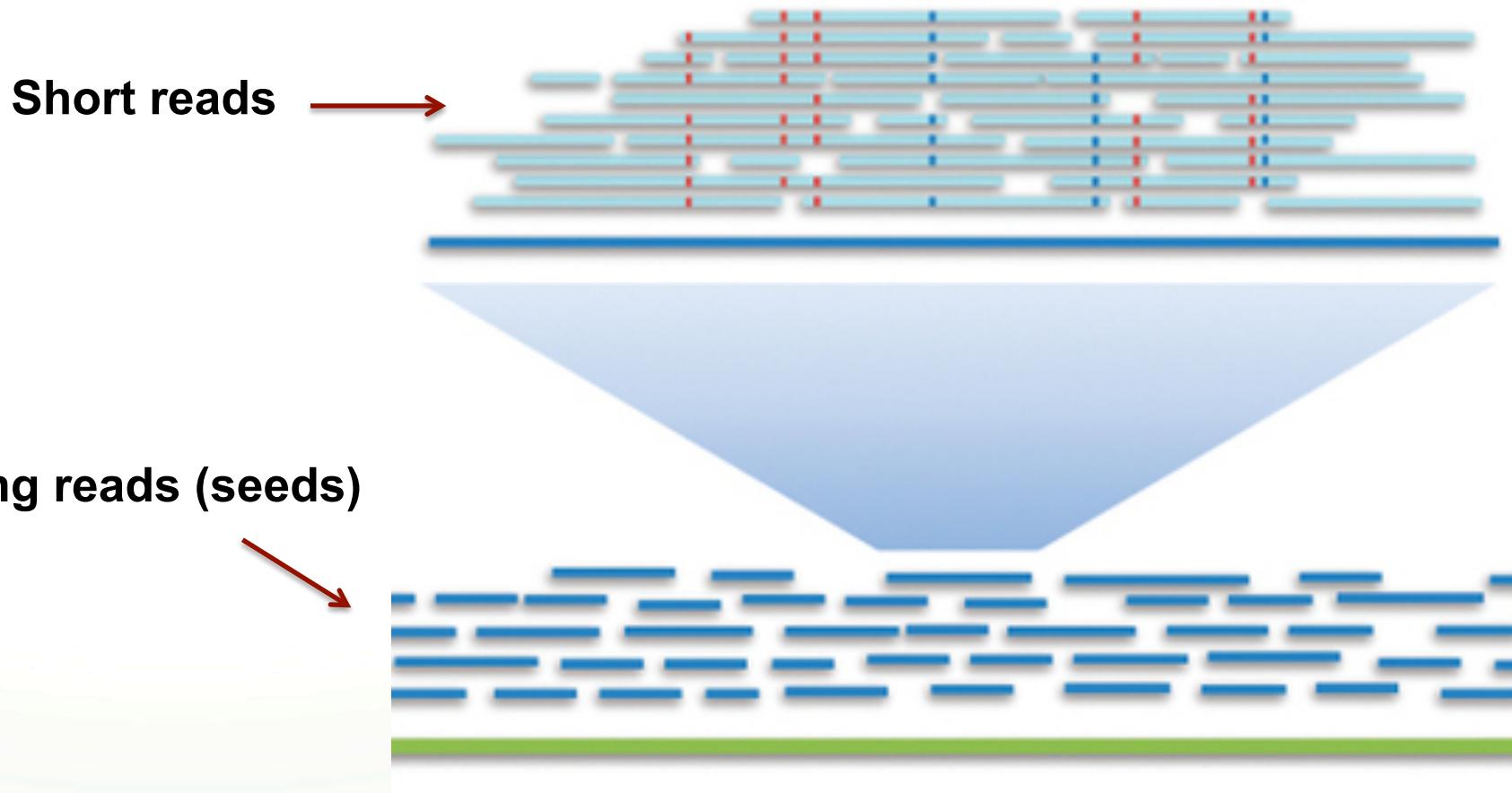


Genome assembly using NGS

- Short-read *de novo* assembly by NGS
 - Requires mate-pair sequences
 - Ideally with different insert sizes
 - Complicated analysis
 - Assembly, scaffolding, finishing
 - Maybe even some manual steps
- => Rather expensive and time consuming
- Long reads really makes a difference!!
 - We can assemble genomes using PacBio data only!

HGAP *de novo* assembly

- HGAP uses both long and shorter reads



PacBio assembly analysis

- Simple -- just click a button!!

The screenshot shows a web browser window with the URL 127.0.0.1:8080/smrtportal/#/Design-Job/Details-of-Job/16497. The page is titled "Details of Job assembly". The top navigation bar includes links for "Home", "Admin", "Tech Support Files", "Help", and "About". On the right, it shows "Welcome, ugc_admin!" and "Log Out".

The main content area has three tabs: "DESIGN JOB", "MONITOR JOBS", and "VIEW DATA". The "DESIGN JOB" tab is active. It contains fields for "Job Name" (assembly), "Comments", "Protocols" (set to RS_HGAP_Assembly.3), and "Reference" (None selected). Below these are two tables:

- SMRT Cells Available (Viewing 1 - 31 of 31)**

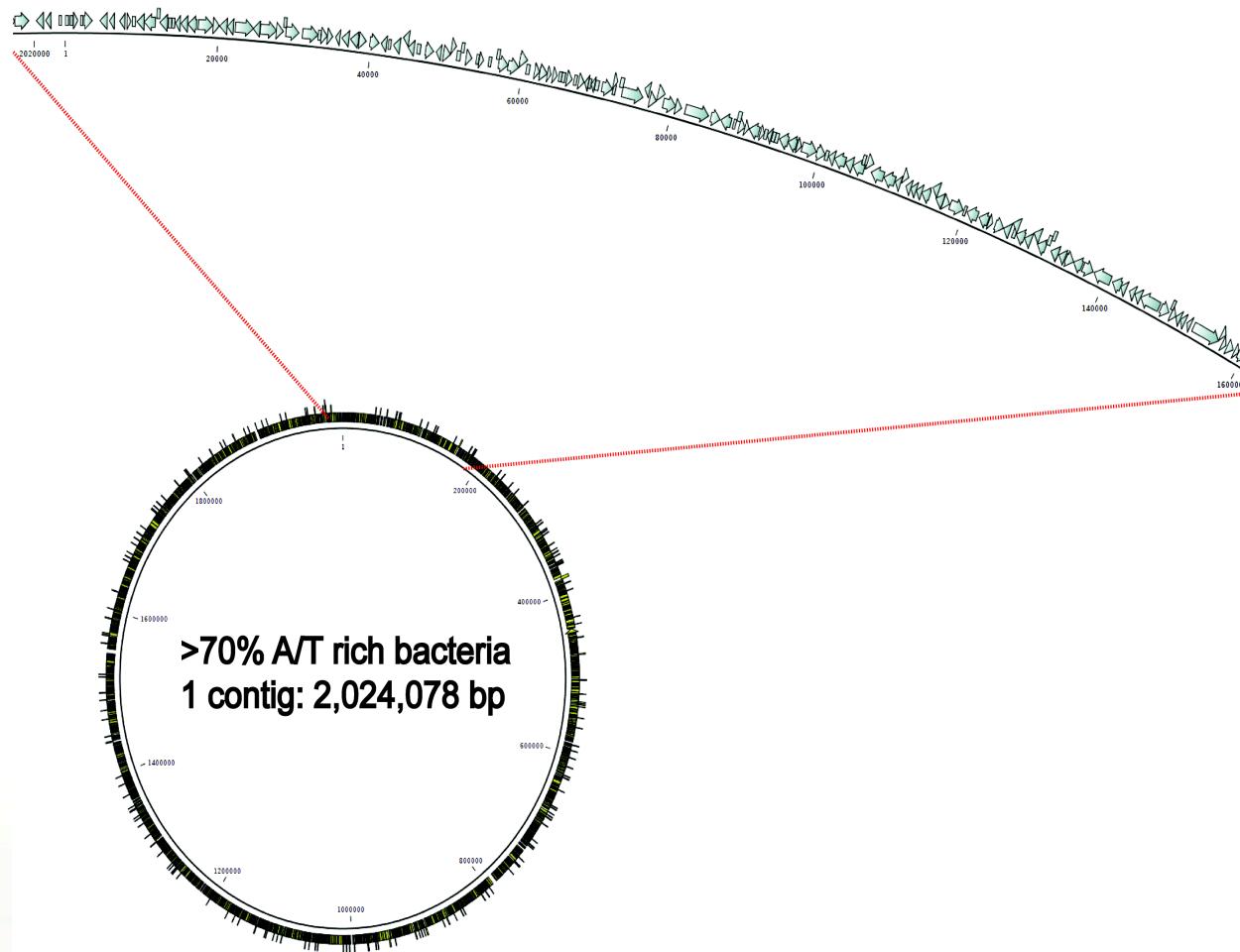
| Sample | Version | User | Groups | Started | Uri |
|----------------|---------|------|--------|--------------------------|---------------|
| Pb9_frax 21 | 2.0.2 | | all | 2014-02-20T19:28:20+0000 | /home/pacbio/ |
| Pb9_frax 44 | 2.0.2 | | all | 2014-02-20T19:28:20+0000 | /home/pacbio/ |
| Pb9_frax 63 | 2.0.2 | | all | 2014-02-20T19:28:20+0000 | /home/pacbio/ |
| Pb33_1 | 2.0.2 | | all | 2014-02-20T19:28:20+0000 | /home/pacbio/ |
| Pb33_2 | 2.0.2 | | all | 2014-02-20T19:28:20+0000 | /home/pacbio/ |
| Pb 33-5 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-7 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-6 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-3 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-9 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-8 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-4 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb 33-10 | 2.0.2 | | all | 2014-02-24T13:48:09+0000 | /home/pacbio/ |
| Pb55_f2 rpt | 2.1.0 | | all | 2014-05-09T10:48:14+0000 | /home/pacbio/ |
| Pb_46_3_repeat | 2.1.0 | | all | 2014-05-09T10:48:14+0000 | /home/pacbio/ |
| Pb55_f2 rpt | 2.1.0 | | all | 2014-05-09T10:48:14+0000 | /home/pacbio/ |
| Pb_46_9 | 2.1.0 | | all | 2014-05-09T10:48:14+0000 | /home/pacbio/ |
| Pb_46_10 | 2.1.0 | | all | 2014-05-09T10:48:14+0000 | /home/pacbio/ |
| Pb46_3 | 2.1.0 | | all | 2014-05-08T11:08:49+0000 | /home/pacbio/ |
| Pb46_5 | 2.1.0 | | all | 2014-05-08T11:08:49+0000 | /home/pacbio/ |
- SMRT Cells in Job (Viewing 1 - 1 of 1)**

| Sample | Version | User | Groups | Uri |
|--------|---------|------|--------|--------------------------------|
| Pb33_1 | 2.0.2 | | all | /home/pacbio/DATA/adam/Pb_33_F |

At the bottom, there are buttons for "Start", "Save", "Copy", and "Cancel".

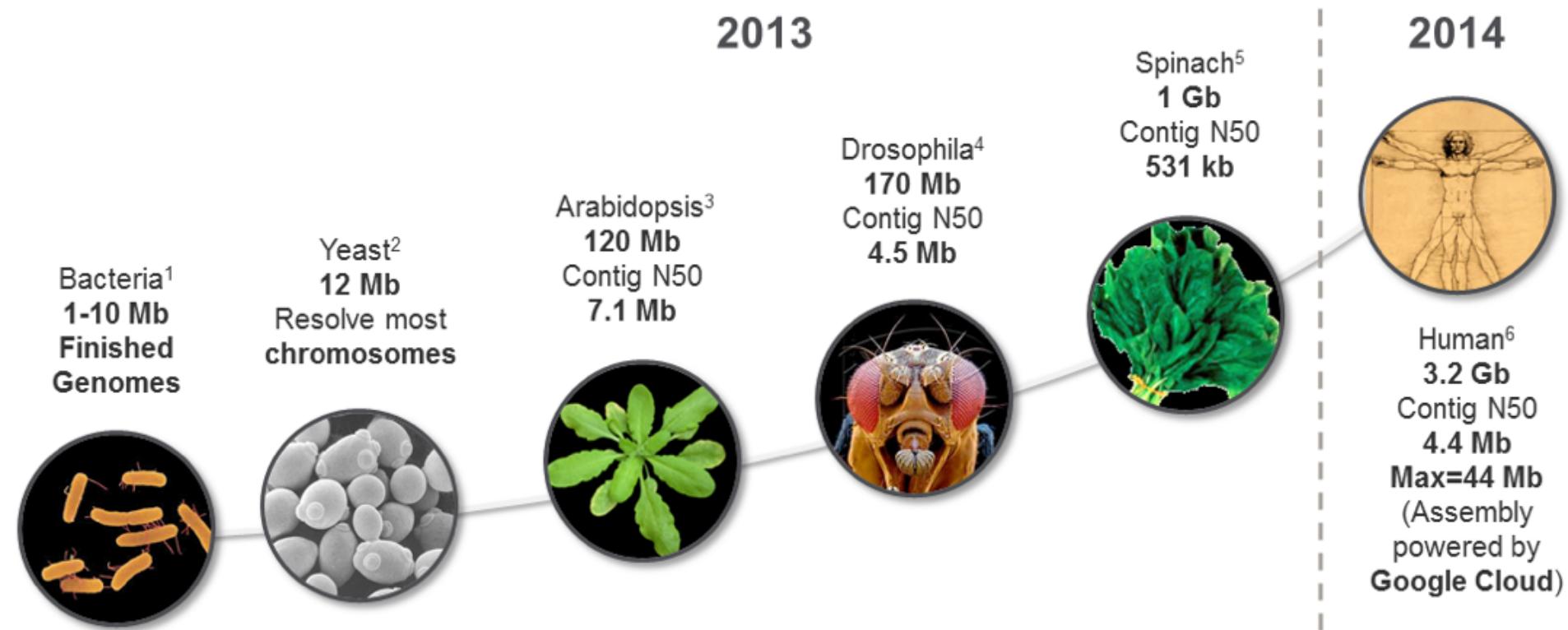
PacBio assembly, example result

- Example: Complete assembly of a bacterial genome



PacBio assembly – recent developments

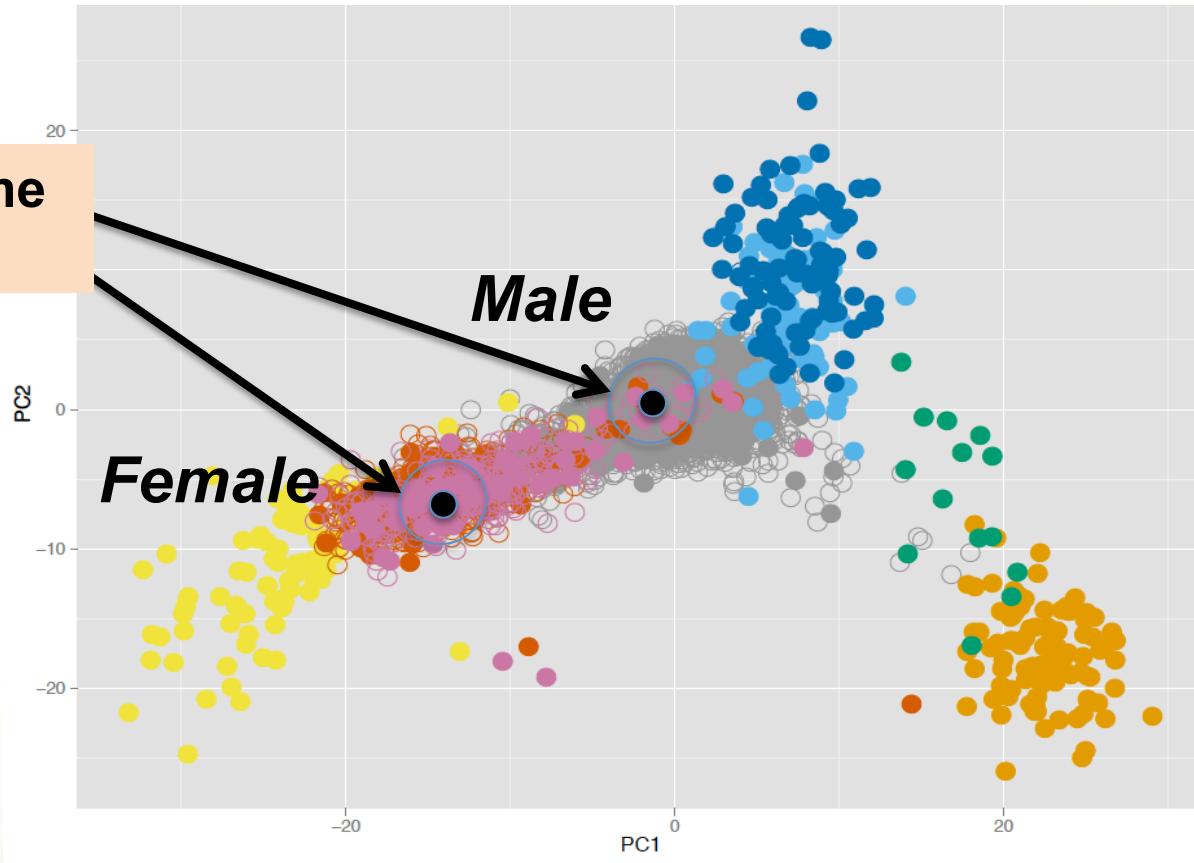
- Also larger genomes can be assembled by PacBio..



De novo WGS of Swedish cohort

Establish Swedish reference genome sequences by *de novo* assembly of long-reads: **PacBio+BioNano+10X Genomics**

Reference genome
individuals

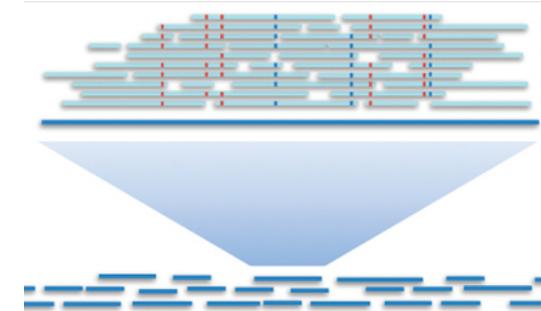


De novo assembly of 75X PacBio data

Assembly (FALCON)



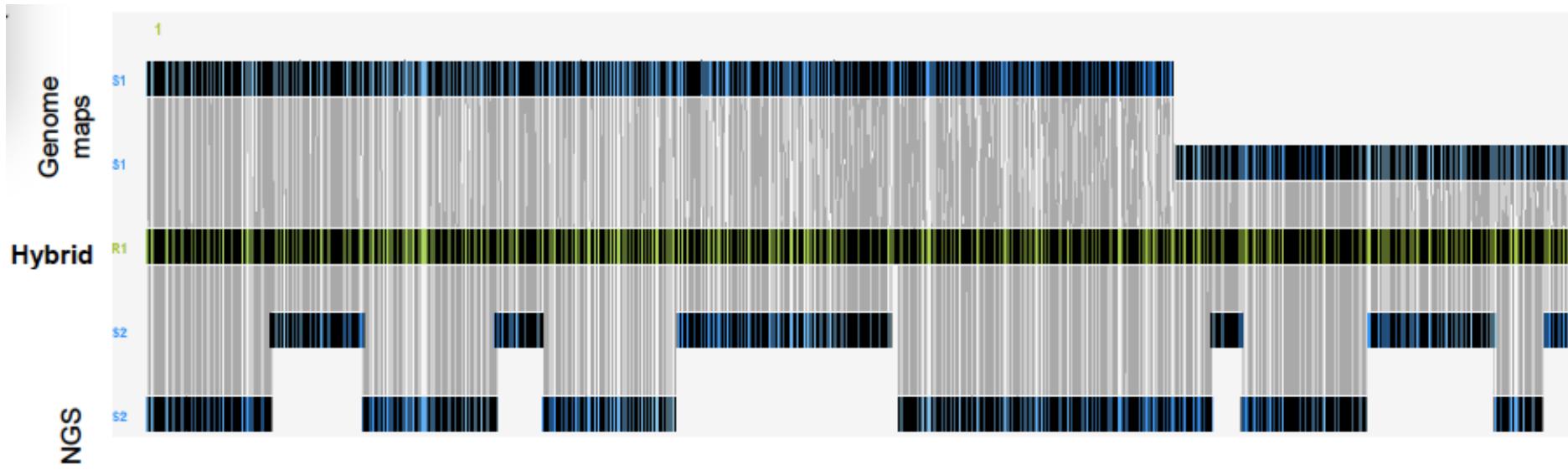
Error correction (2 x Quiver)



Analysis time: 1 month/genome

| | Individual 1 | Individual 2 |
|-----------------------|---------------|---------------|
| Assembly size | 3,039,619,582 | 3,024,752,299 |
| Nr contigs | 11,249 | 11,601 |
| Longest contig | 36,8 Mb | 54,1 Mb |
| N50 | 8,9 Mb | 8,3 Mb |

Hybrid scaffolding, PacBio + BioNano



Hybrid scaffolding with two labellings resulted in

- **3,1 Gb assembly, 51 Mb N50** (for individual #1)
- **3,1 Gb assembly, 46 Mb N50** (for individual #2)

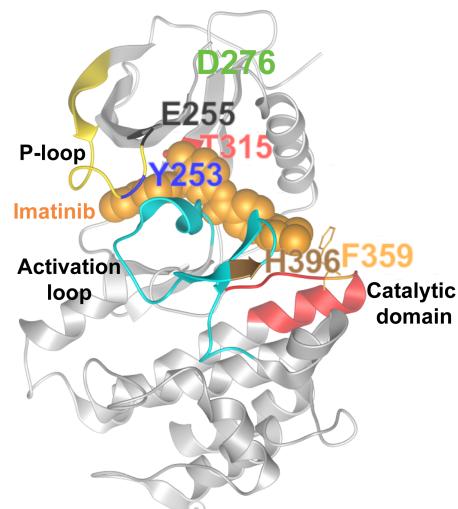
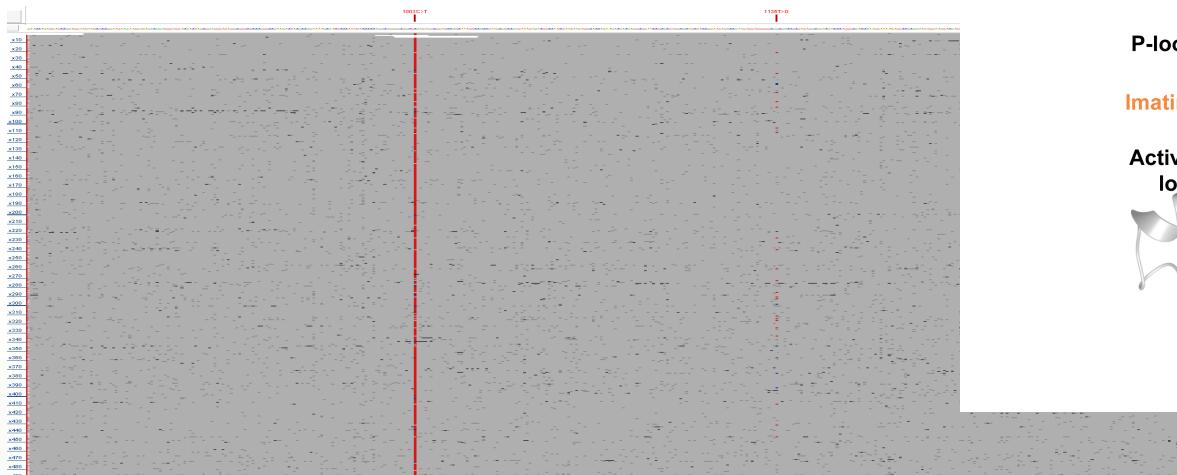
Aligning contigs to human reference

> 99% of bases can be aligned to human reference (hg38)



Example III:

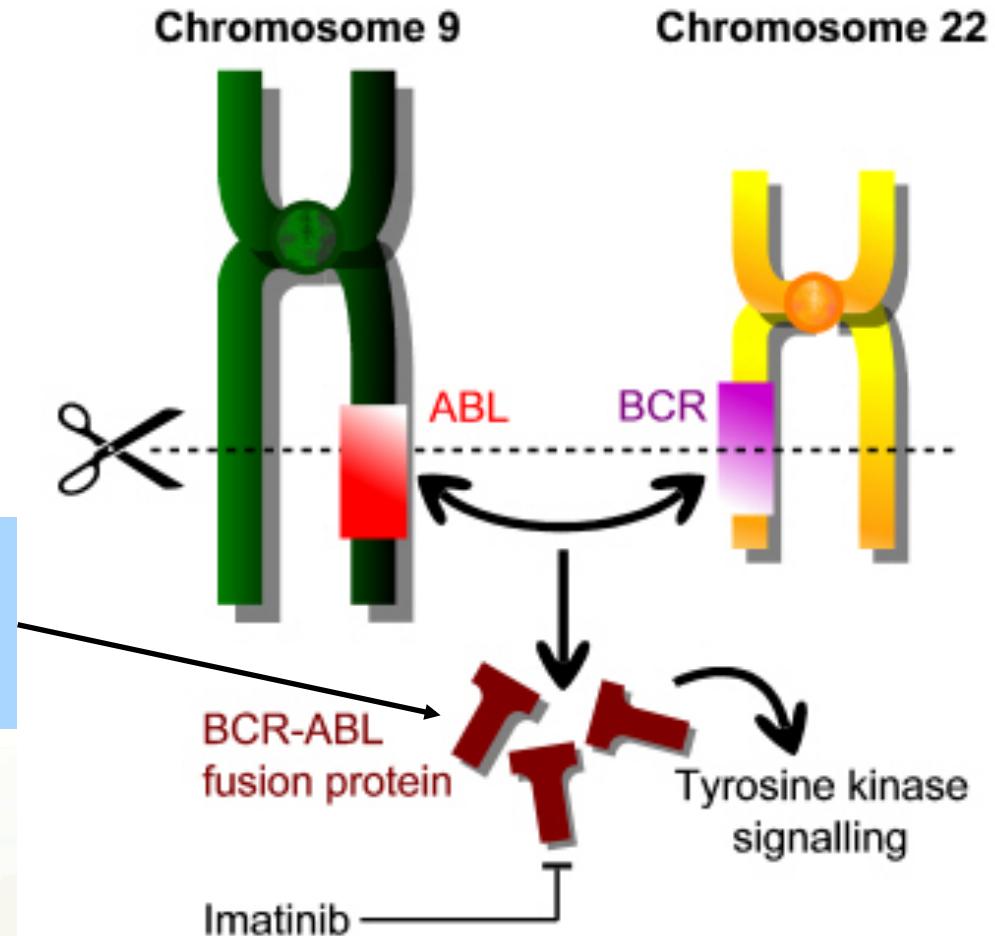
Clinical sequencing for Leukemia Treatment



Chronic Myeloid Leukemia

- BCR-ABL1 fusion protein – a CML drug target

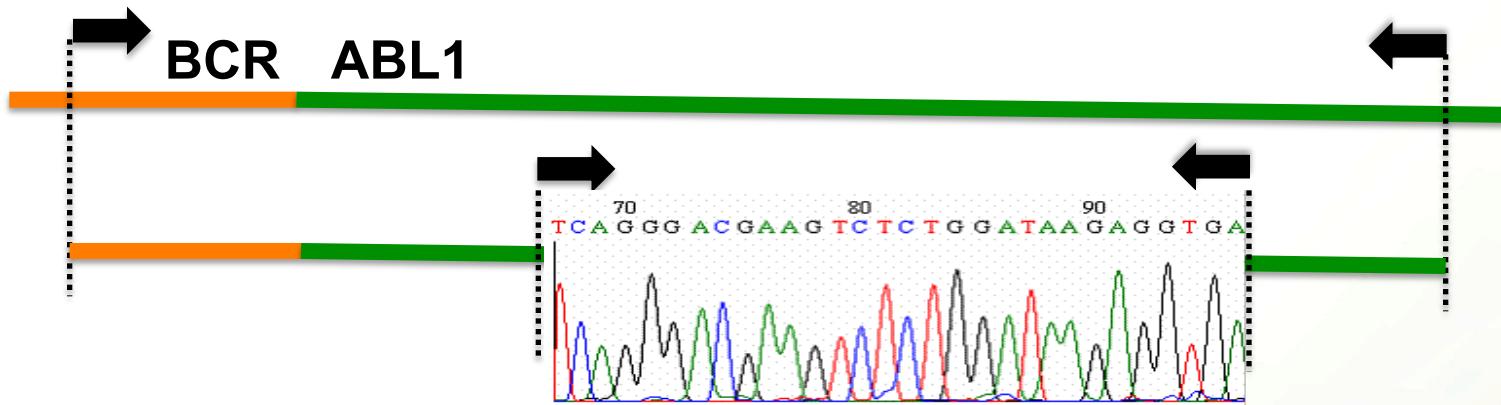
The BCR-ABL1 fusion protein can acquire resistance mutations following drug treatment



www.cambridgemedicine.org/article/doi/10.7244/cmj-1355057881

Traditional mutation screening in BCR-ABL1

Nested PCR and Sanger sequencing:



Limitations:

- Mutations at frequencies below 10-20% not seen
- Biases may be introduced by nested PCR
- Whole BCR-ABL1 fusion transcript not sequenced
- Clonal composition of mutations not determined

Our clinical diagnostics pipeline for BCR-ABL1

Cavelier et al. BMC Cancer (2015) 15:45
DOI 10.1186/s12885-015-1046-y



RESEARCH ARTICLE

Open Access

Clonal distribution of *BCR-ABL1* mutations and splice isoforms by single-molecule long-read RNA sequencing

Lucia Cavelier^{1*†}, Adam Ameur^{1†}, Susana Häggqvist¹, Ida Höijer¹, Nicola Cahill¹, Ulla Olsson-Strömberg² and Monica Hermanson¹

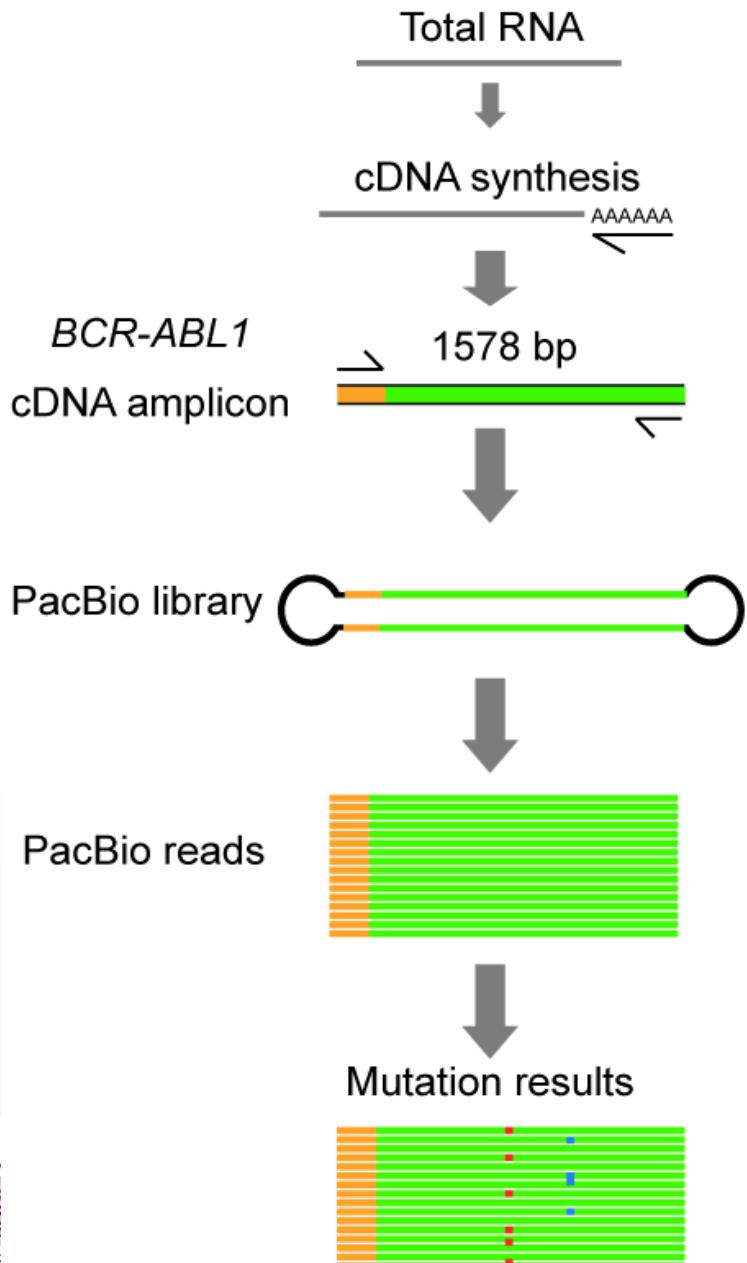
Abstract

Background: The evolution of mutations in the *BCR-ABL1* fusion gene transcript renders CML patients resistant to tyrosine kinase inhibitor (TKI) based therapy. Thus screening for *BCR-ABL1* mutations is recommended particularly in patients experiencing poor response to treatment. Herein we describe a novel approach for the detection and surveillance of *BCR-ABL1* mutations in CML patients.

Methods: To detect mutations in the *BCR-ABL1* transcript we developed an assay based on the Pacific Biosciences (PacBio) sequencing technology, which allows for single-molecule long-read sequencing of *BCR-ABL1* fusion transcript molecules. Samples from six patients with poor response to therapy were analyzed both at diagnosis and follow-up. cDNA was generated from total RNA and a 1,6 kb fragment encompassing the *BCR-ABL1* transcript was amplified using long range PCR. To estimate the sensitivity of the assay, a serial dilution experiment was performed.

Results: Over 10,000 full-length *BCR-ABL1* sequences were obtained for all samples studied. Through the serial dilution analysis, mutations in CML patient samples could be detected down to a level of at least 1%. Notably, the assay was determined to be sufficiently sensitive even in patients harboring a low abundance of *BCR-ABL1* levels. The PacBio sequencing successfully identified all mutations seen by standard methods. Importantly, we identified several mutations that escaped detection by the clinical routine analysis. Resistance mutations were found in all but one of the patients. Due to the long reads afforded by PacBio sequencing, compound mutations present in the same molecule were readily distinguished from independent alterations arising in different molecules. Moreover, several transcript isoforms of the *BCR-ABL1* transcript were identified in two of the CML patients. Finally, our assay allowed for a quick turn around time allowing samples to be reported upon within 2 days.

Conclusions: In summary the PacBio sequencing assay can be applied to detect *BCR-ABL1* resistance mutations in both diagnostic and follow-up CML patient samples using a simple protocol applicable to routine diagnosis. The method besides its sensitivity, gives a complete view of the clonal distribution of mutations, which is of importance when making therapy decisions.

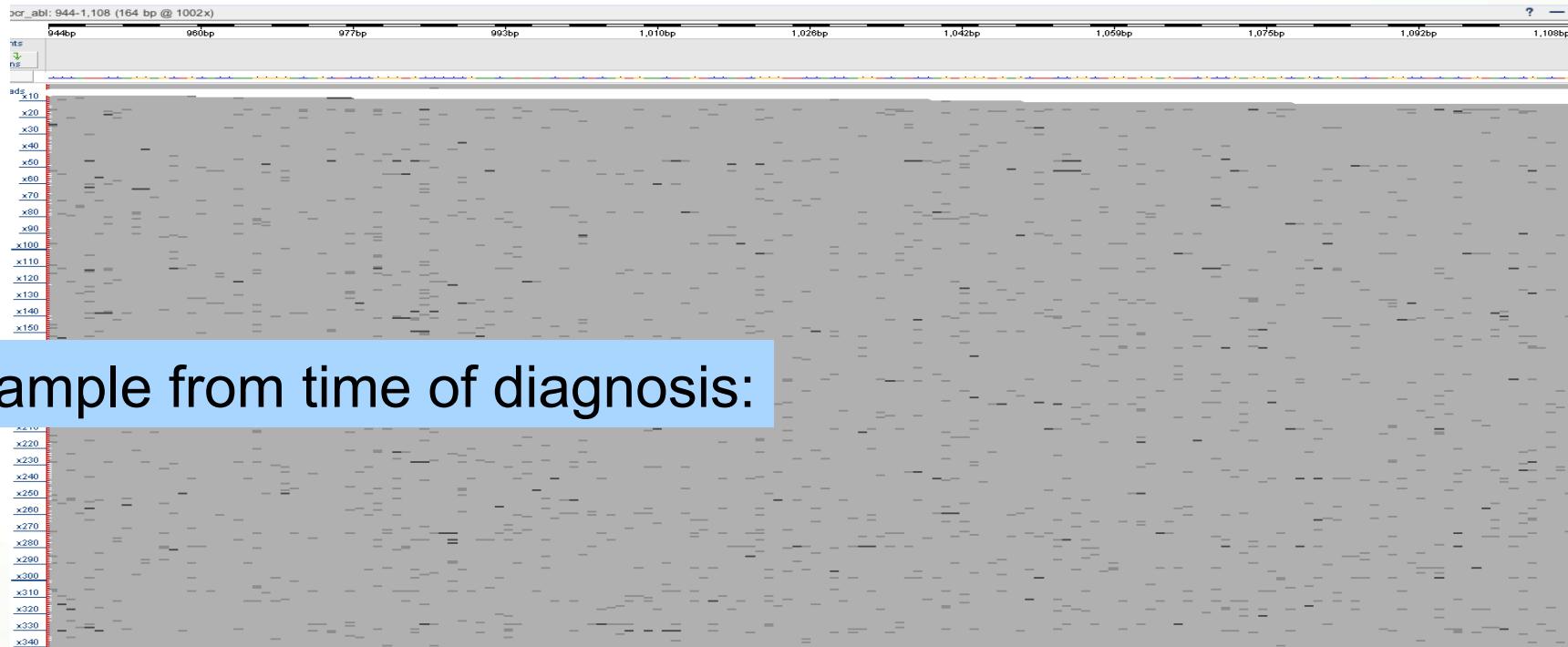


BCR-ABL1 mutations at diagnosis

PacBio sequencing generates ~10 000X coverage!

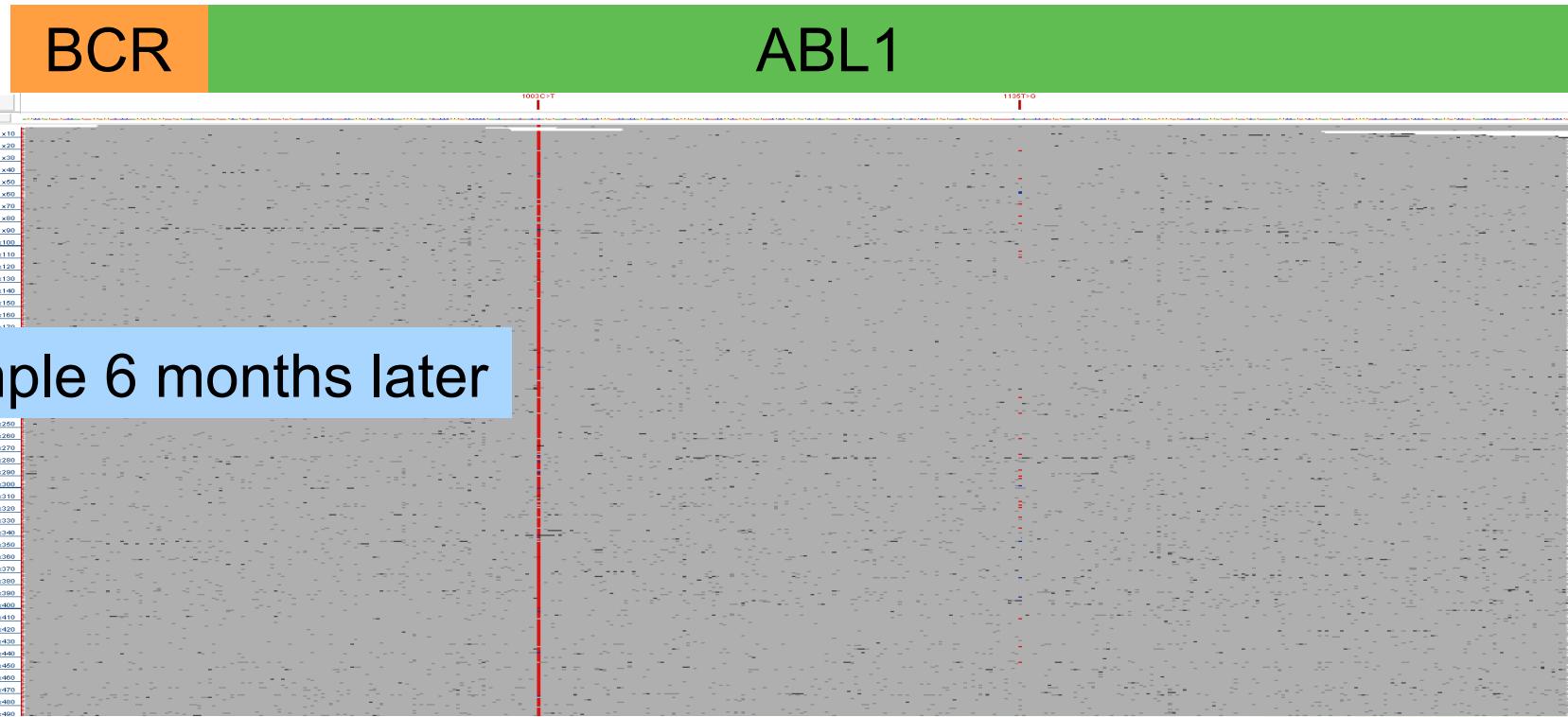
BCR

ABL1



Sample from time of diagnosis:

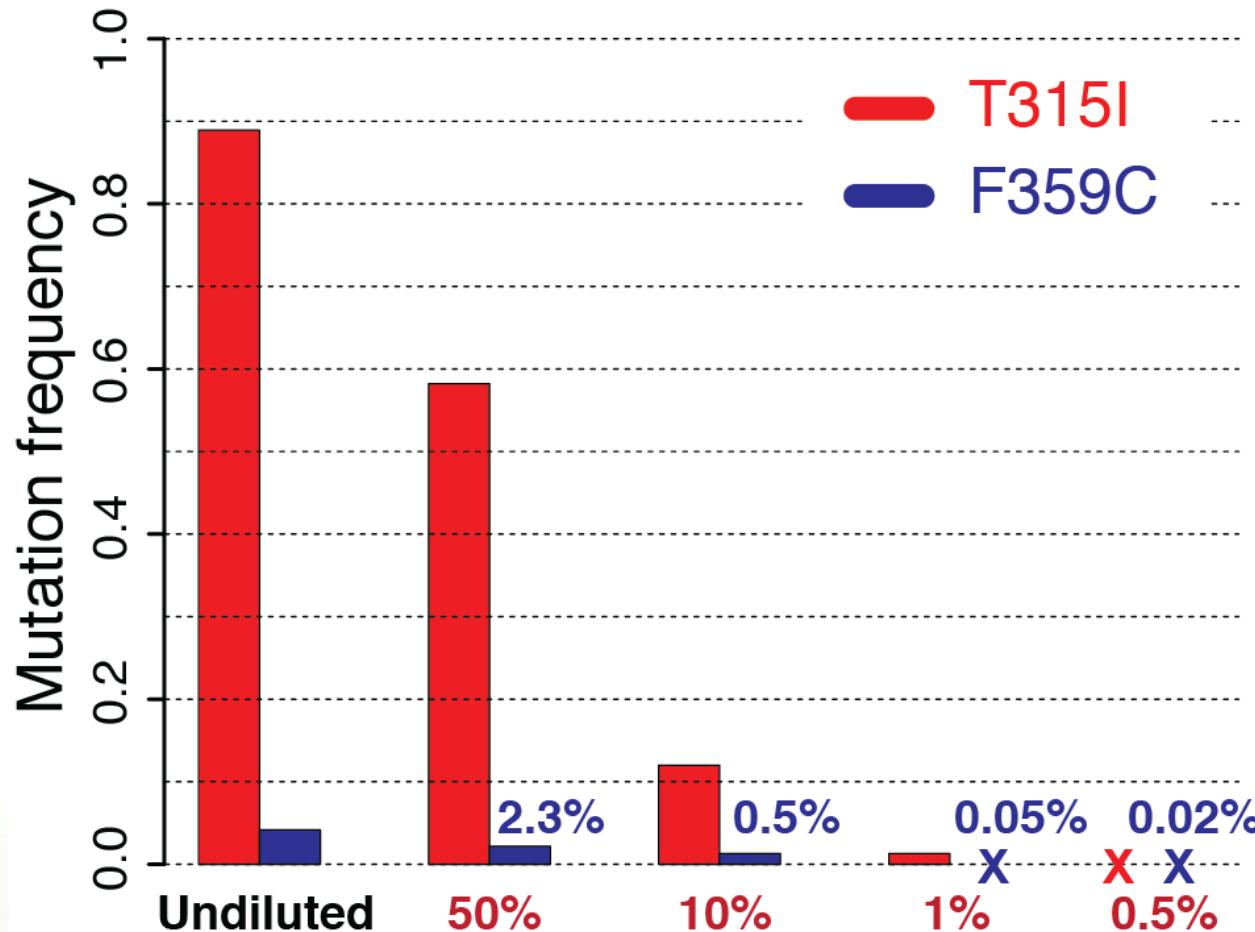
BCR-ABL1 mutations in follow-up sample



Mutations acquired in fusion transcript.
Might require treatment with alternative drug.

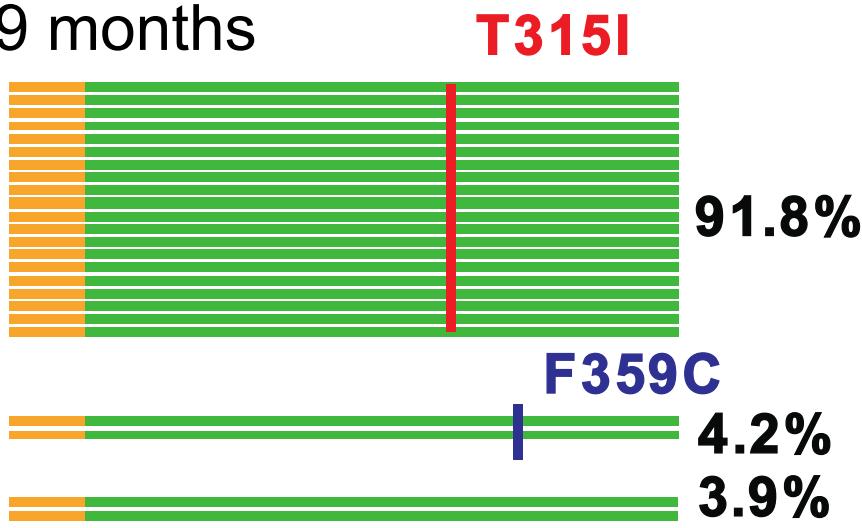
BCR-ABL1 dilution series results

- Mutations down to 1% detected!

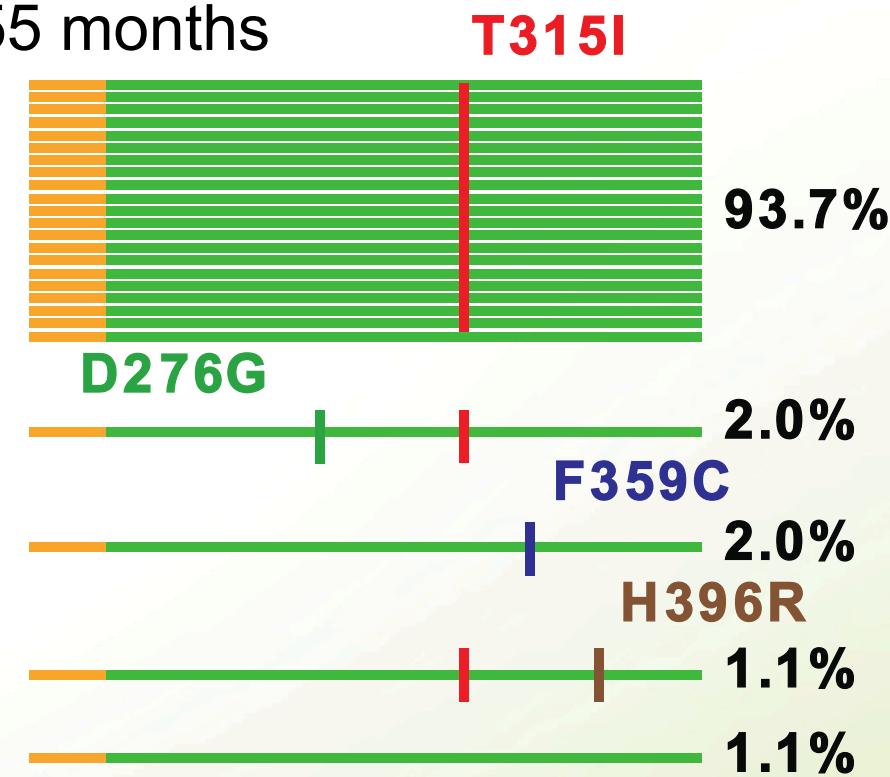


BCR-ABL1 - Compound mutations

49 months



55 months



Analysis method for BCR-ABL1 mutations

- Create CCS reads and screen all known resistance mutations
- CAVA analysis - count number of WT and MUT sequences

WT sequence: **TATATCATCACTGAGTTCATG**

MUT sequence: **TATATCATCAT**T**TGAGTTCATG**



BCR-ABL1 resistance mutation

- Classify each mutation
 - Less than 500X coverage => **Unresolved**
 - At least 0.5% mutation frequency => **Positive**
 - Otherwise => **Negative**

Web system for sharing results

| Details | Sample ID | Run ID | Unresolved (count) | Unknown (count) | M244V | Q252H | Y253H | E255K | E255V | K262N | D276G | T277A | L298V | T315I | T315A | M351T | F359V | L387M | E450G | E453G | E459G | M472I | E499E | Date |
|---------|-----------|-----------|--------------------|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| (91) | R12021 | cba_011_2 | | | | | | | | | | | | | | | | | | | | | | |
| (92) | R12023 | cba_011_3 | | | | | | | | | | | | | | | | | | | | | | |
| (93) | R12026 | cba_011_4 | | | | | | | | | | | | | | | | | | | | | | |
| (94) | R12091 | cba_012_1 | | | | | | | | | | | | | | | | | | | | | | |
| (95) | R12092 | cba_012_2 | | | | | | | | | | | | | | | | | | | | | | |
| (96) | R12093 | cba_012_3 | | | | | | | | | | | | | | | | | | | | | | |
| (97) | R12095 | cba_012_4 | | | | | | | | | | | | | | | | | | | | | | |
| (98) | R12124 | cba_013_1 | | | | | | | | | | | | | | | | | | | | | | |
| (99) | R12125 | cba_013_2 | | | | | | | | | | | | | | | | | | | | | | |
| (100) | R12126 | cba_013_3 | | | | | | | | | | | | | | | | | | | | | | |
| (101) | R12126 | cba_014_1 | | | | | | | | | | | | | | | | | | | | | | |
| (102) | R12149 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (103) | R12165 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (104) | R12143 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (105) | R12281 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (106) | R12282 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (107) | R12222 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (108) | R12291 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (109) | R12355 | cba_ | | | | | | | | | | | | | | | | | | | | | | |
| (110) | R12200 | cba_ | | | | | | | | | | | | | | | | | | | | | | |

101 Sample 102 103 [New Search](#)

| | | |
|----------------------------|----------------------------|---------------------------|
| Sample ID R12095 | Run ID cba_012_4 | Date 2015-09-17 |
|----------------------------|----------------------------|---------------------------|

[Results](#)
[Sequence](#)
Downloads:
[Coverage](#)
[Clonal txt](#)
[Clonal pdf](#)
[Log](#)

| mutation | sequence | wt_reads | mut_reads | other_reads | freq | detection |
|----------|--|----------|-----------|-------------|-------|-----------|
| M351T | CACTCAGATCTCGTCAGCCA[T/C]GGAGTACCTGGAGAAAGAAAA | 16134 | 19065 | 3 | 0.542 | positive |
| Q252H | CACAAGCTGGGGGGGCCAG[C/C]TACGGGGAGGTGTACCAAGGG | 12052 | 9920 | 8 | 0.451 | positive |
| K262N | GTGTACGAGGGCGTGAGAA[G/T]AAATACAGCCTGACGGTGCG | 25597 | 6996 | 16 | 0.215 | positive |
| M244V | TGGAACCCACGGACATCACC[A/G]TGAAGCACAAGCTGGCGGG | 32779 | 32 | 2 | 0.001 | negative |
| K247R | GGACATCACCATGAAGCACA[A/G]GCTGGCGGGGCCAGTACG | 27076 | 32 | 9 | 0.001 | negative |

Coverage of bcr-14-abl1, based on 500 reads

Frequency Reads

| | |
|--------|------|
| 49.9 % | 9268 |
| 23.8 % | 4418 |
| 17.4 % | 3245 |
| 8.69 % | 1613 |

Position in bcr-14-abl1

Stockholms universitet
UPPSALA UNIVERSITET
KAROLINSKA INSTITUTET

Clinical Diagnosis of BCR-ABL1 mutations

Clinical Genetics



- Collection of samples
- Seq library preparation

Sequencing Facility



- SMRT sequencing
- CAVA analysis

IT developers

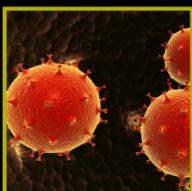
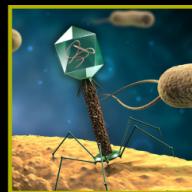
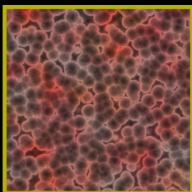


- Web server for results

- Ongoing routine service, 0-4 samples/week.
- Over 150 patient samples run so far
- 100% consistency with Sanger results

What we sequence at NGI /

SciLifeLab



- Diabetes
 - Alzheimer's disease
 - Whole-genome sequencing**
 - Gene therapy
 - Infection screen
 - Whole-transcriptome sequencing**
 - Target sequencing
 - Cancer prognosis
 - Gene regulation
 - Crohn's disease
 - Genomics of ageing
 - Exome sequencing**
 - Schizophrenia
 - Cancer diagnostics
 - Organ donor matching
 - Gut microflora
 - Gene fusions**
 - RNA editing
 - HIV
 - HPV**
 - HCV
 - Scoliosis
 - Immune response
 - Monogenic disorders
 - Sudden infant death
 - Cervical cancer**
 - Lynch syndrom
 - Leukemia
 - Scoliosis
 - HLA typing**
 - Dyslexia
 - MRSA / BRSA screen
 - Sudden cardiac arrest
 - Transcriptional regulation
 - Prenatal diagnostics**
 - Muscle dystrophy
 - Individualised cancer therapy
 - and much more...
- 