High-troughput technologies and databases

Irene Pérez Díez

Bioinformatics and Biostatistics Unit

Wednesday 16th October

Unidad de Bioinformática y Bioestadística

PRINCIPE FELIPE CENTRO DE INVESTIGACION

CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

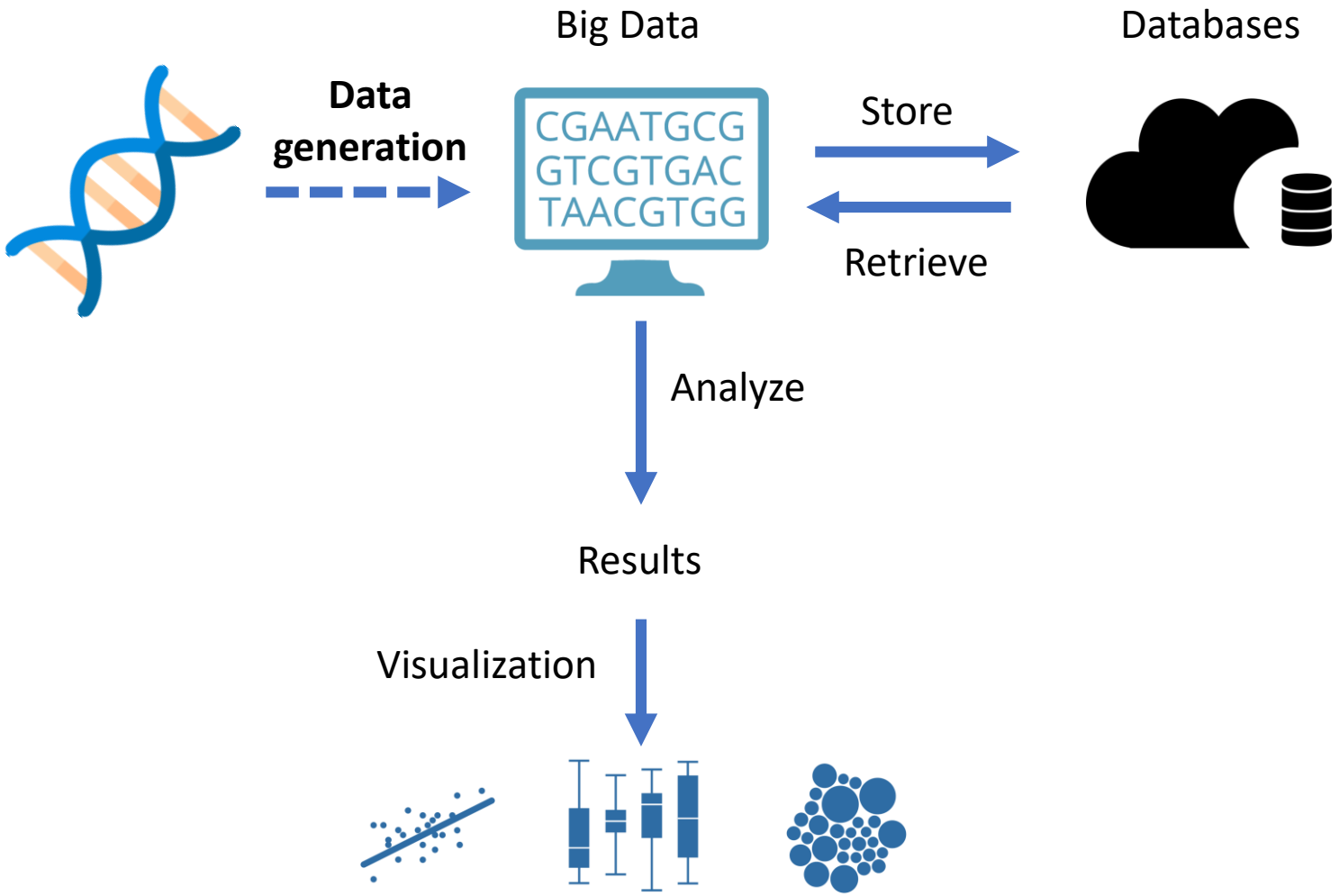
WODA

WEB-BASED OMICS DATA ANALYSIS

Outline

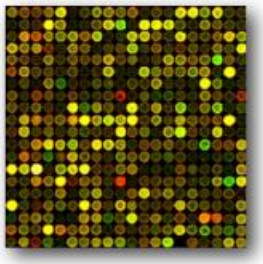
- High-throughput technologies
- Workflow and tools
- Databases

NGS technologies



NGS technologies

NGS / High-throughput



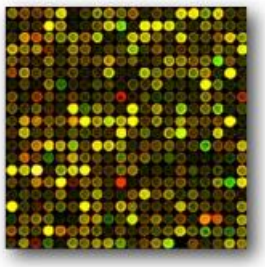
Sanger DNA Sequencing	Microarrays	2nd generation sequencing	3rd generation & single-molecule sequencing
Since 1977	Since mid-1990s	Since 2007	Since 2010

NGS technologies

Sanger

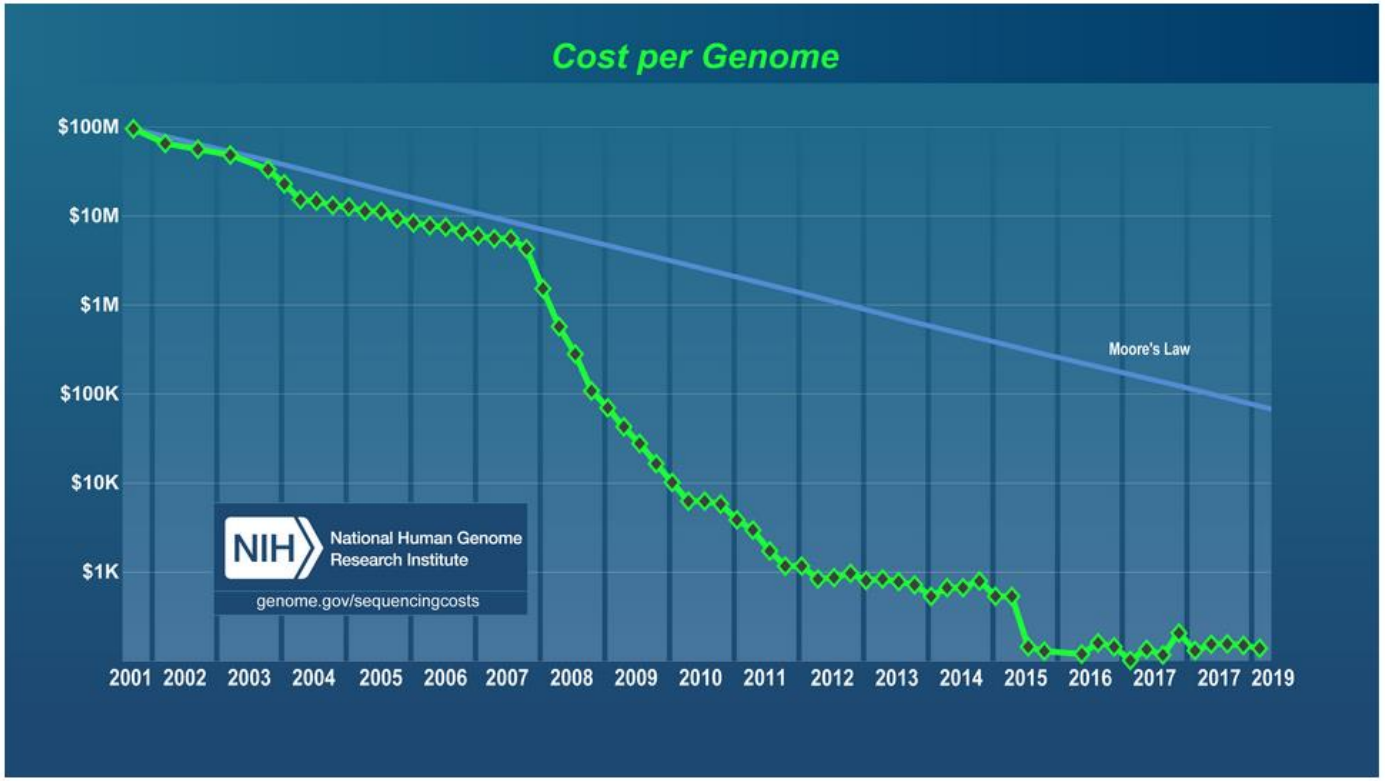
- *De novo* sequencing
- Targeted DNA sequencing
- NGS sequencing validation
- Microbial sequencing
- Mitochondiral sequencing

Microarray



- Comparative genomics
- Gene expresión profiling
- Clinical diagnostics
- Metylation analysis

NGS technologies



NGS technologies

Platform (Company)	Chemistry	Read length (bp)	No. reads	Raw error rate (%)	Applications
454 (Roche)	Pyro-sequencing	700	$1x10^6$	1	Bacterial and viral genomes, multiplex PCR, validation of point mutations, targeted somatic-mutation detection
HiSeq (Illumina)	Synthesis	150x2	$5x10^9$	0,8	Complex genomes (human, mouse and plants) and genome-wide NGS, RNA-seq, hybrid capture or multiplex-PCR, somatic-mutation detection, forensics, noninvasive prenatal testing
MiSeq (Illumina)	Synthesis	300x2	$3x10^8$	0,8	
SOLiD (Thermofisher)	Ligation	50	$1x10^9$	0,01	Complex genomes and genome-wide NGS, RNA-seq, hybrid capture or multiplex-PCR, somatic-mutation detection
Ion Torrent (Thermofisher)	Synthesis	200-400	$6x10^7$	1,7	Multiplex-PCR, microbiology and infectious diseases, somatic-mutation detection, validation of point mutations
3rd generation					
SMRT (Pac Bio)	Real-time SMS	> 10,000	$1x10^6$	12,9	Complex genomes, microbiology and infectious-disease genomes, transcript-fusion detection, methylation detection
MinION PromethION (Oxford Nanopore)	Real-time SMS	> 5000	$6x10^4$	34	Pathogen surveillance, targeted mutation detection, metagenomics, bacterial and viral genomes

Jerzy K. Kulski (January 14th 2016). Next-Generation Sequencing — An Overview of the History, Tools, and "Omic" Applications, Next Generation Sequencing - Advances, Applications and Challenges, Jerzy K Kulski, IntechOpen, DOI: 10.5772/61964

NGS chemistry overview

A. Library Preparation

B. Cluster Amplification

- Bridge PCR
- Emulsion PCR

C. Sequencing

- Pyrosequencing
- Sequencing by synthesis
- Ion semiconductor sequencing
- Sequencing by ligation
- Real-time SMS ← No PCR!!

D. Alignment/mapping and Data Analysis

FASTA - FASTQ

- Fasta
  - .fasta
  - .fa

- Fastq
  - .fastq
  - .fq

Header

>NP_414878.1 beta-galactosidase
MTMITDSLAVVLQRRDWENP...

@SRR010930.8436795/1
ACCCCAGGATCAACACTTCACATGCATTAGCAG
+
=>=??A?<@B@A:?B?D;AC@@CAAAD<AAA:9

Sequence

Quality

Computing requirements

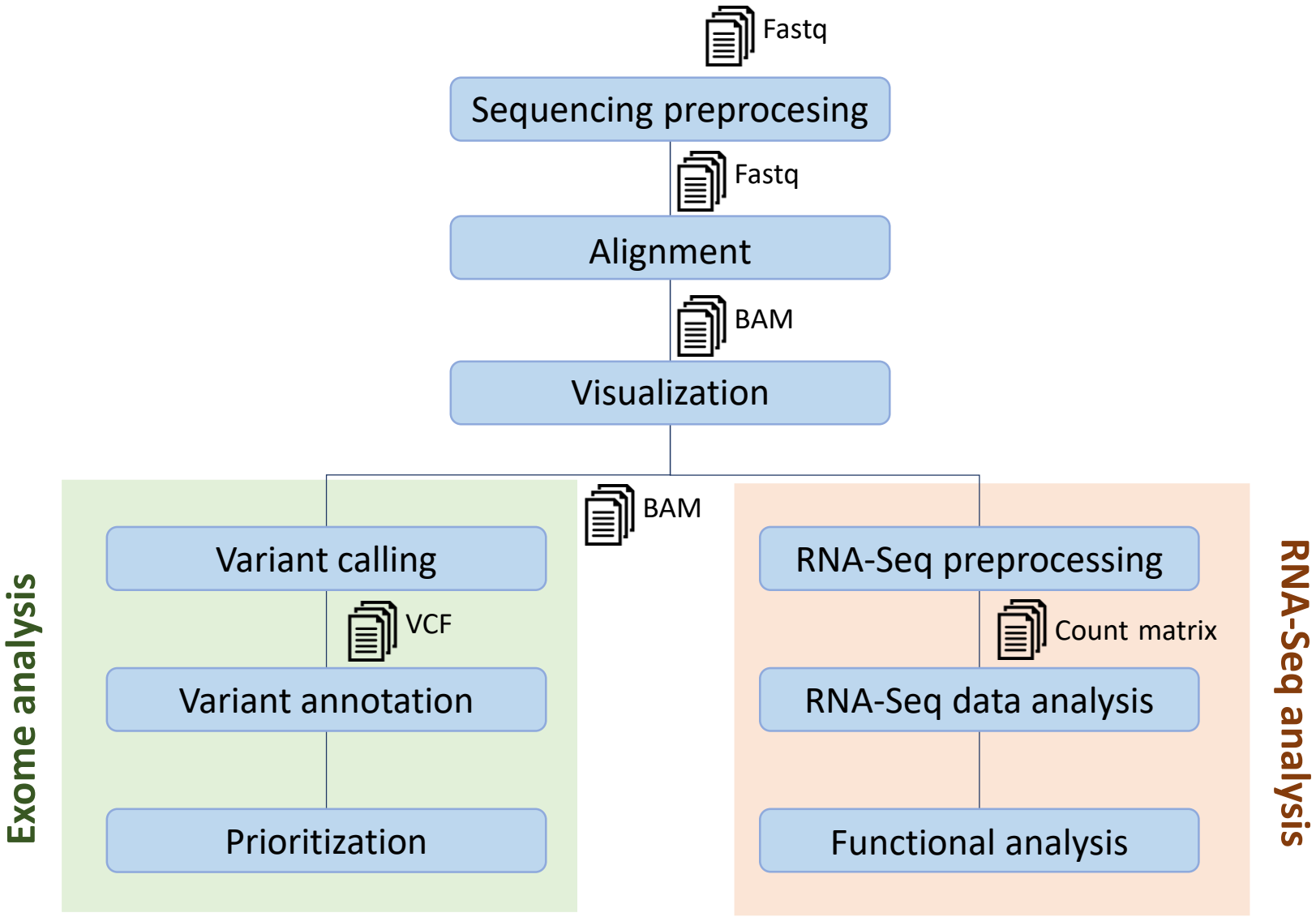
Conditioned data center (server roms)
Computing cluster
Many computer nodes (servers)
High performance and storage capacity
Fast networks
Sysadmins and developers

Cloud computing

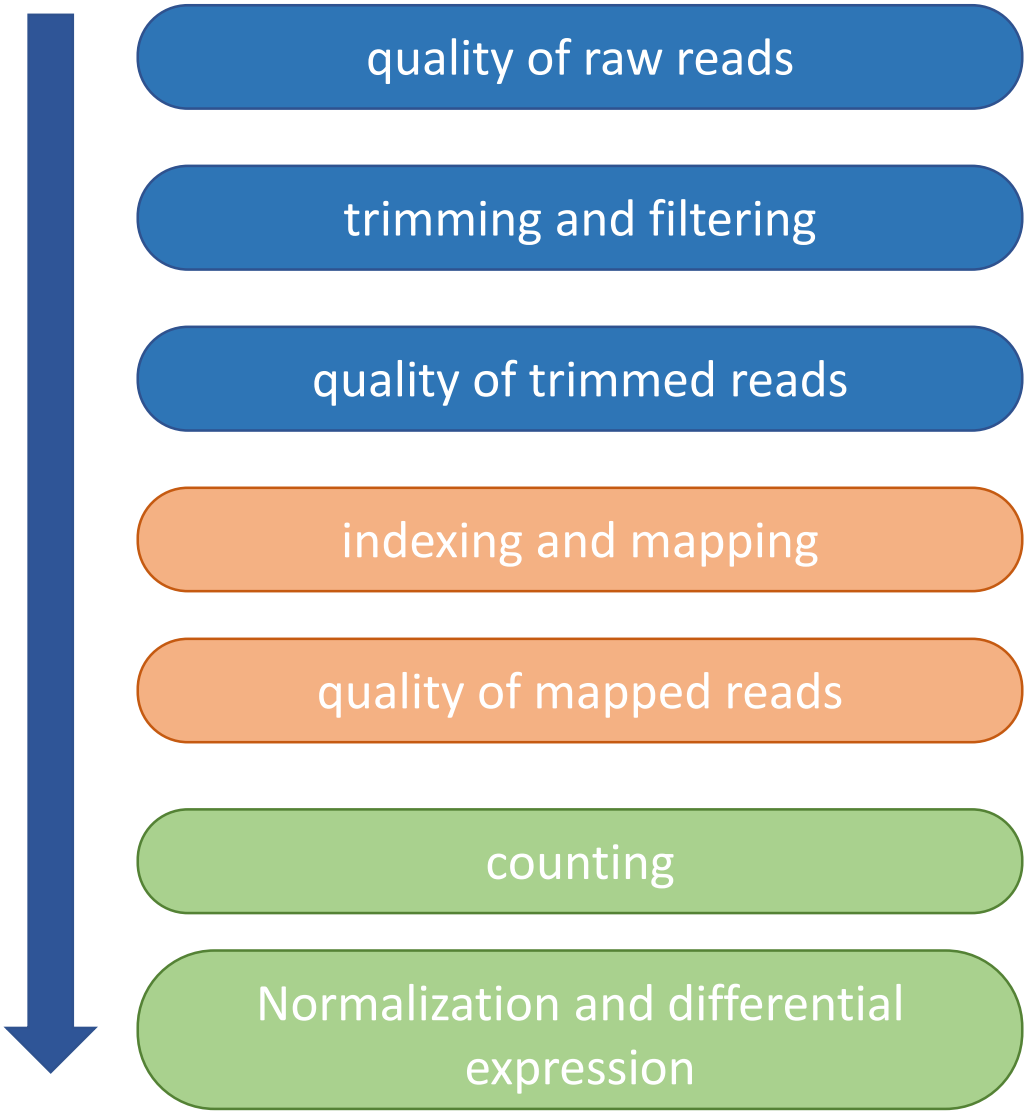
- Flexibility
- You pay what you use
- Don’t need to mantain a data center

- Transfer datasets through the internet is slow
- Lower performance
- Privacy and security concerns
- More expensive for big and long term projects

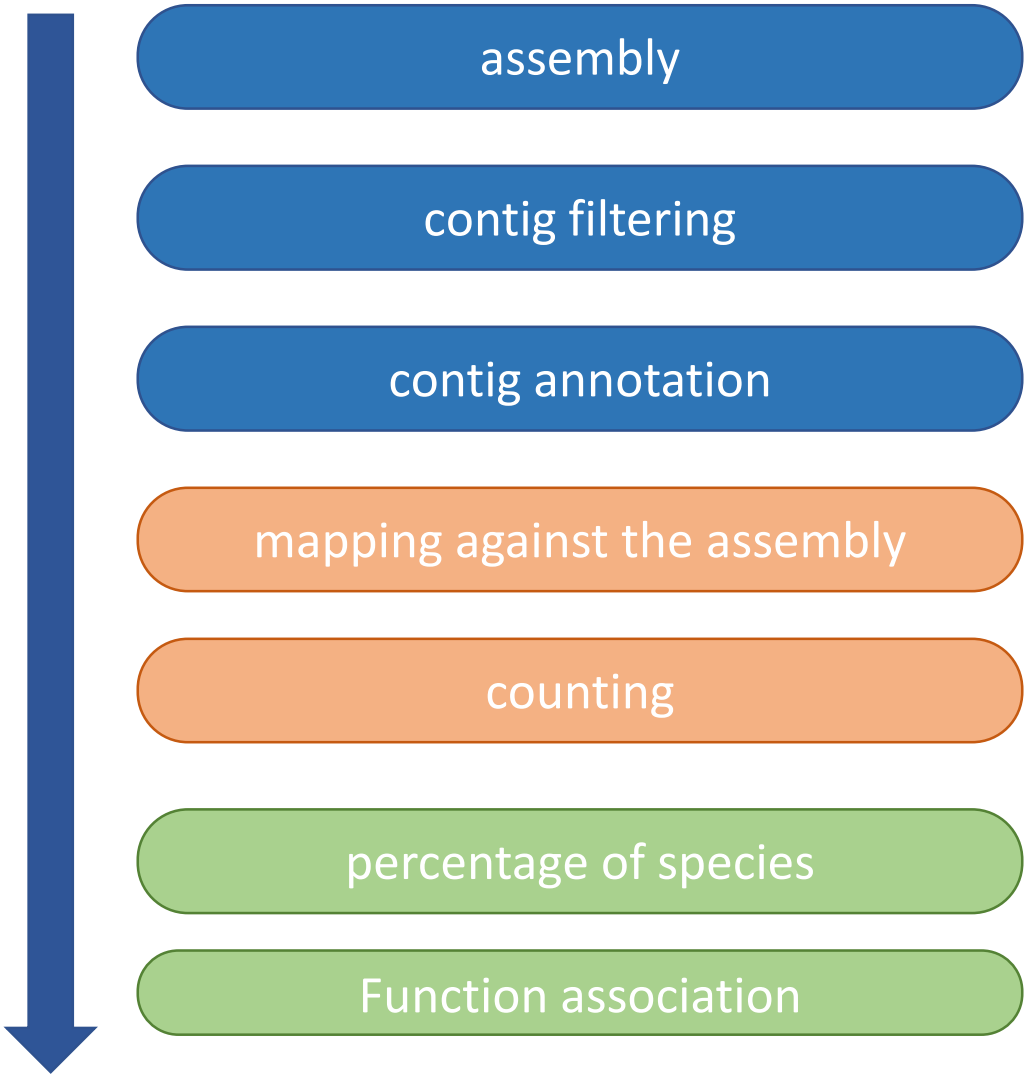
Basic workflow



RNAseq pipeline



Metagenomics pipeline



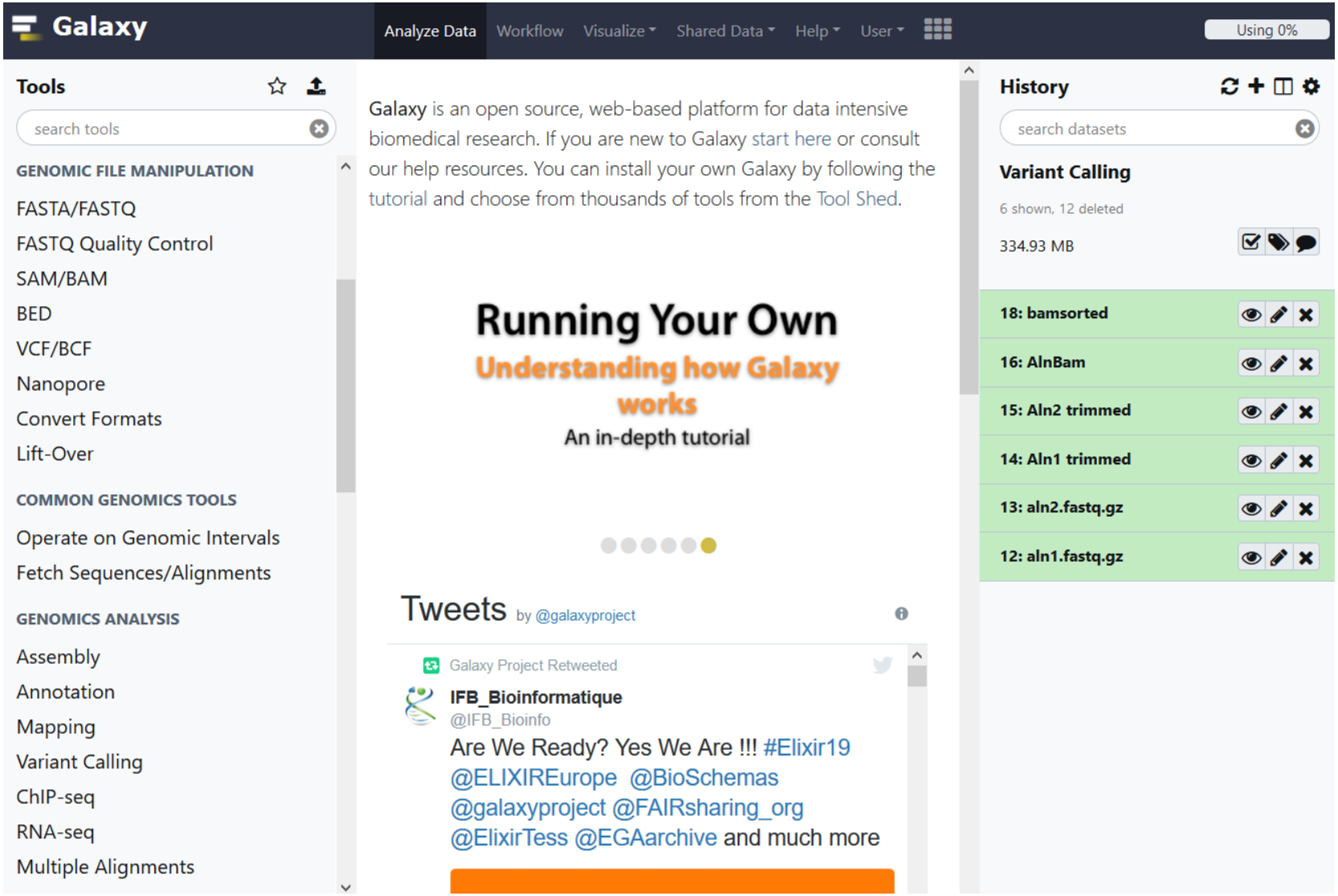
NGS tools

FastQC https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ Quality control	**Blast2GO** https://www.blast2go.com/ Functional annotation and analysis
Cutadapt https://cutadapt.readthedocs.io/en/stable/ Trimming: remove adaptors and other sequences	**Samtools** http://www.htslib.org/ Work with SAM/BAM/CRAM files
Bowtie2 http://bowtie-bio.sourceforge.net/bowtie2/index.shtml Alignment	**Vcftools** https://vcftools.github.io/index.html Work with VCF files
Bwa http://bio-bwa.sourceforge.net/ Alignment	**GATK** https://software.broadinstitute.org/gatk/ From variant discovery to metagenomics
STAR https://github.com/alexdobin/STAR RNA-seq aligner	**HISAT2** http://ccb.jhu.edu/software/hisat2/index.shtml RNA / DNA aligner

NGS tools

Cufflinks http://cole-trapnell-lab.github.io/cufflinks/ Transcriptome assembly and differential expression	**Mothur** https://www.mothur.org/ Microbial ecology toolbox
ABySS https://github.com/bcgsc/abyss *de novo* sequence assembler (large genomes)	**Bismark** https://www.bioinformatics.babraham.ac.uk/projects/bismark/ Bisulfite converted sequence reads – cytosine methylation
SPAdes http://cab.spbu.ru/software/spades/ Genome assembler (small genomes)	**BLAST** https://blast.ncbi.nlm.nih.gov/Blast.cgi Alignment
GLIMMER https://ccb.jhu.edu/software/glimmer/ Gene predictor - microbial DNA	**Augustus** http://bioinf.uni-greifswald.de/augustus/ Gene predictor – eukaryotic DNA
IGV https://software.broadinstitute.org/software/igv/ Genome visualization	**Quiime2** https://qiime2.org/ Microbiome bioinformatics platform

NGS tools


Galaxy
Analyze Data
Workflow
Visualize
Shared Data
Help
User
Using 0%
Tools
search tools
GENOMIC FILE MANIPULATION
FASTA/FASTQ
FASTQ Quality Control
SAM/BAM
BED
VCF/BCF
Nanopore
Convert Formats
Lift-Over
COMMON GENOMICS TOOLS
Operate on Genomic Intervals
Fetch Sequences/Alignments
GENOMICS ANALYSIS
Assembly
Annotation
Mapping
Variant Calling
ChIP-seq
RNA-seq
Multiple Alignments
Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy start here or consult our help resources. You can install your own Galaxy by following the tutorial and choose from thousands of tools from the Tool Shed.
Running Your Own
Understanding how Galaxy works
An in-depth tutorial
Tweets by @galaxyproject
Galaxy Project Retweeted
IFB_Bioinformatique
@IFB_Bioinfo
Are We Ready? Yes We Are !!! #Elixir19 @ELIXIREurope @BioSchemas @galaxyproject @FAIRsharing_org @ElixirTess @EGAarchive and much more
History
search datasets
Variant Calling
6 shown, 12 deleted
334.93 MB
18: bamsorted
16: AlnBam
15: Aln2 trimmed
14: Aln1 trimmed
13: aln2.fastq.gz
12: aln1.fastq.gz

NGS tools

Bowtie2 - map reads against reference genome (Galaxy Version 2.3.4.2)

Favorite | Versions | Options

Is this single or paired library

Paired-end

FASTA/Q file #1

12: aln1.fastq.gz

Must be of datatype "fastqsanger"or "fasta"

FASTA/Q file #2

13: aln2.fastq.gz

Must be of datatype "fastqsanger"or "fasta"

Write unaligned reads (in fastq format) to separate file(s)

Yes | No

--un/--un-conc (possibly with -gz or -bz2); This triggers --un parameter for single reads and --un-conc for paired reads

Write aligned reads (in fastq format) to separate file(s)

Yes | No

--al/--al-conc (possibly with -gz or -bz2); This triggers --al parameter for single reads and --al-conc for paired reads

Do you want to set paired-end options?

No

See "Alignment Options" section of Help below for information

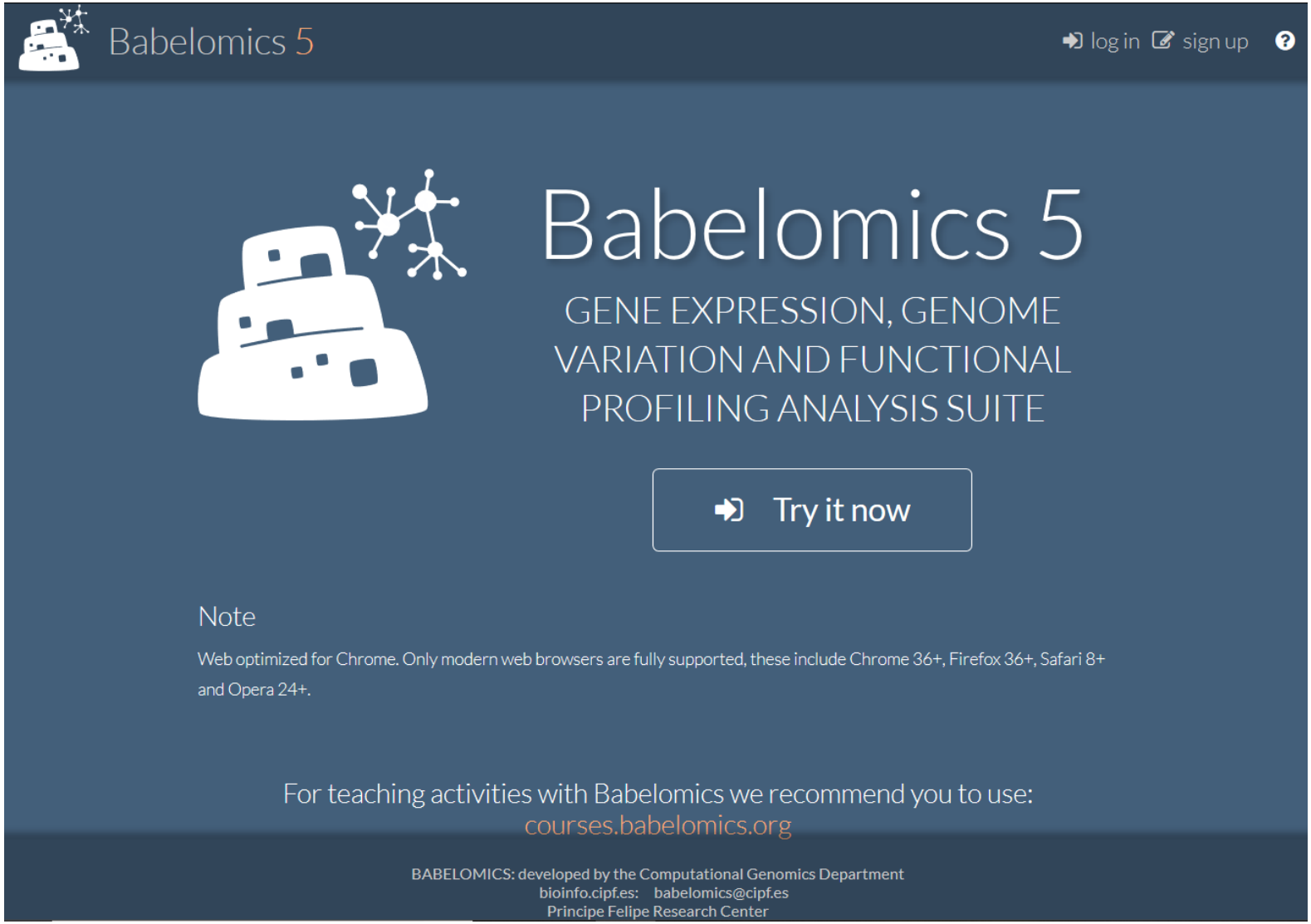
Will you select a reference genome from your history or use a built-in index?

19: Bowtie2 on data 13 and data 12: aligned reads (BAM)

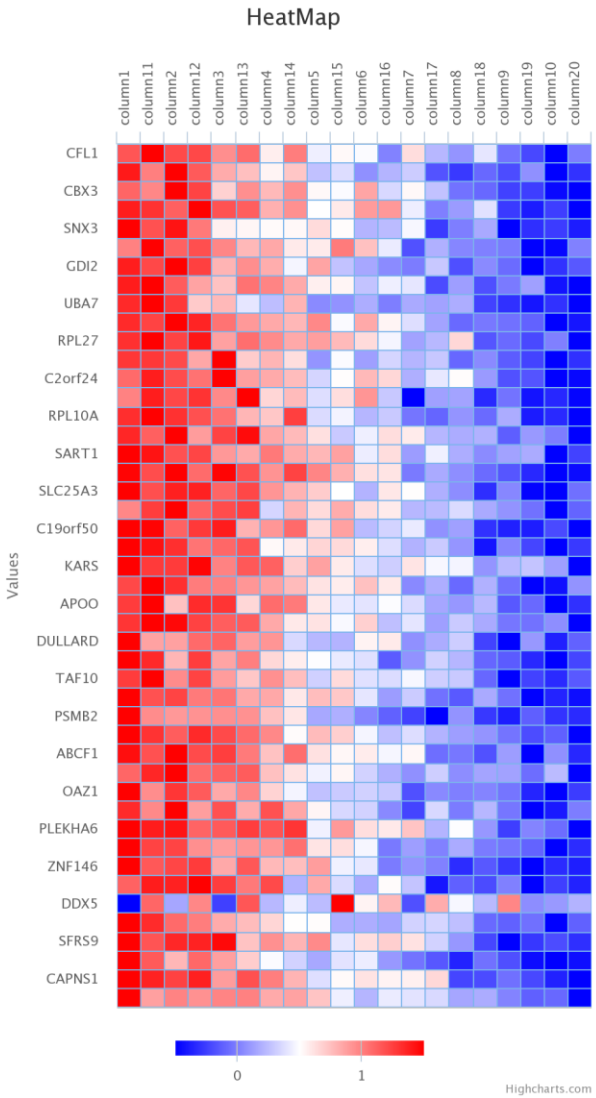
13: aln2.fastq.gz

12: aln1.fastq.gz

NGS tools


Babelomics 5
log in
sign up
Babelomics 5
GENE EXPRESSION, GENOME VARIATION AND FUNCTIONAL PROFILING ANALYSIS SUITE
Try it now
Note
Web optimized for Chrome. Only modern web browsers are fully supported, these include Chrome 36+, Firefox 36+, Safari 8+ and Opera 24+.
For teaching activities with Babelomics we recommend you to use:
courses.babelomics.org
BABELOMICS: developed by the Computational Genomics Department
bioinfo.cipf.es: babelomics@cipf.es
Principe Felipe Research Center

NGS tools



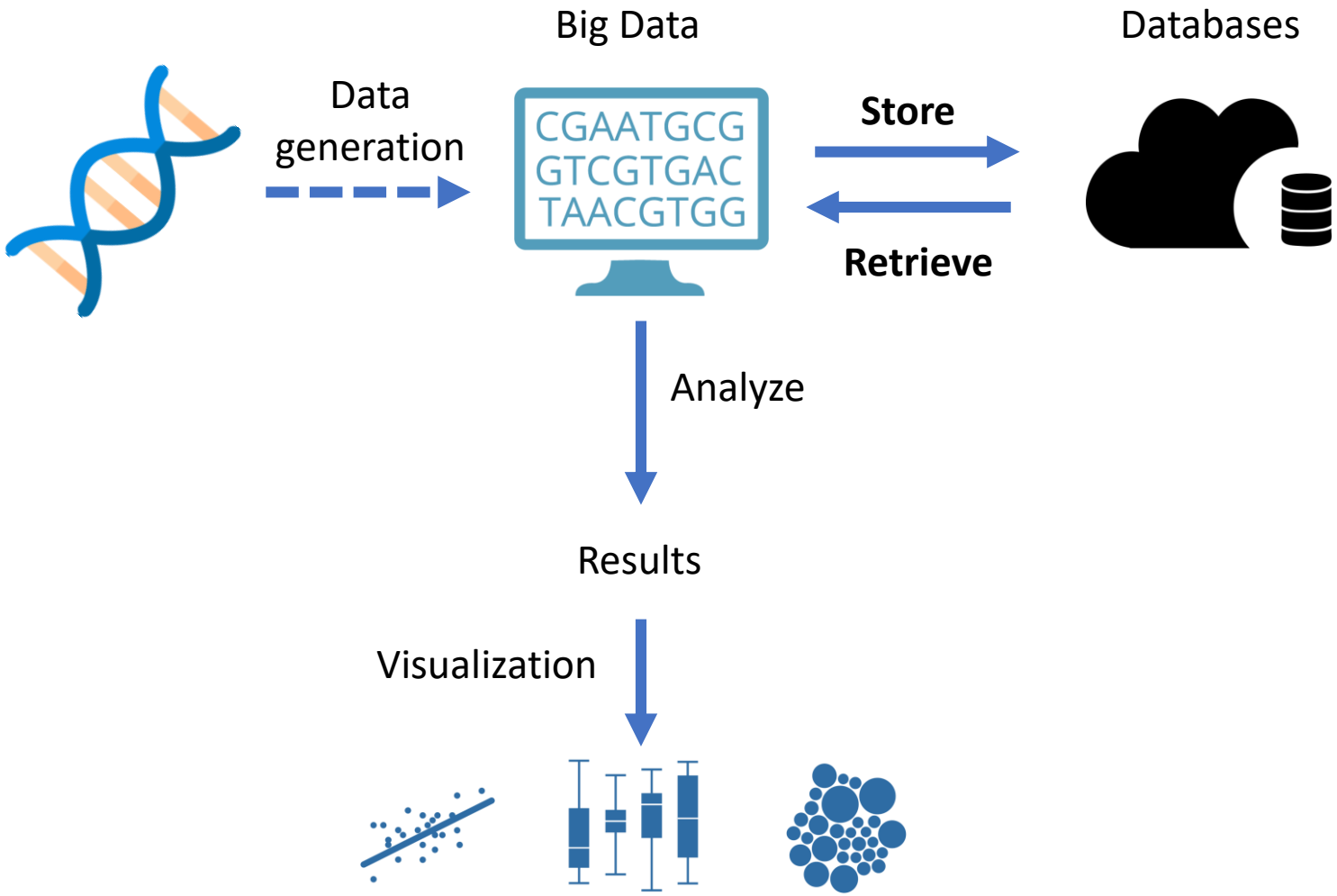
#NAMES	statistic	coefs.	p-value	adj. p-value
CFL1	3.16	1.05	0.0016	0.019
CCL5	3.12	1.04	0.0018	0.019
CBX3	3.04	1.06	0.0024	0.019
RPL19	2.99	1.02	0.0028	0.019
SNX3	2.95	0.85	0.0032	0.019
JTB	2.95	0.94	0.0032	0.019
GDI2	2.93	1.04	0.0034	0.019
RPS24	2.93	0.96	0.0034	0.019
UBA7	2.92	0.91	0.0034	0.019
MYST2	2.9	1.11	0.0037	0.019

46 Results

‹ 1 of 5 ›

Term	Term size	Term size(in genome)	annotated_genes lists	converged ids list	lor	adj_pvalue
positive regulation of developmental process(GO:0051094)	8	1937	THRA SART1 PAX8 RHOA EPHB3 BAD	true	-0.51	0.049
negative regulation of cellular biosynthetic process(GO:0031327)	11	2778	THRA CBX3 NONO TARDBP RPS27A KHDRBS1	true	-0.52	0.024
heterocycle metabolic process(GO:0046483)	64	20258	RPL18 ABCF1 RPL17 CHURC1 THRA RPL19	true	-0.4	0.02

Databases

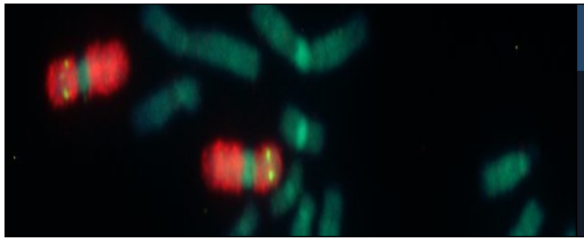

Big Data
Databases
Data generation
CGAATGCG
GTCGTGAC
TAACGTGG
Store
Retrieve
Analyze
Results
Visualization

Databases

NCBI Resources How To Sign in to NCBI

Gene Gene BRCA2 Search

Advanced Help



Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene	Gene Tools	Other Resources
Gene Quick Start	Submit GeneRIFs	HomoloGene
FAQ	Submit Correction	OMIM
Download/FTP	Statistics	RefSeq
RefSeq Mailing List	BLAST	RefSeqGene
Gene News	Genome Workbench	UniGene
Factsheet	Splign	Protein Clusters

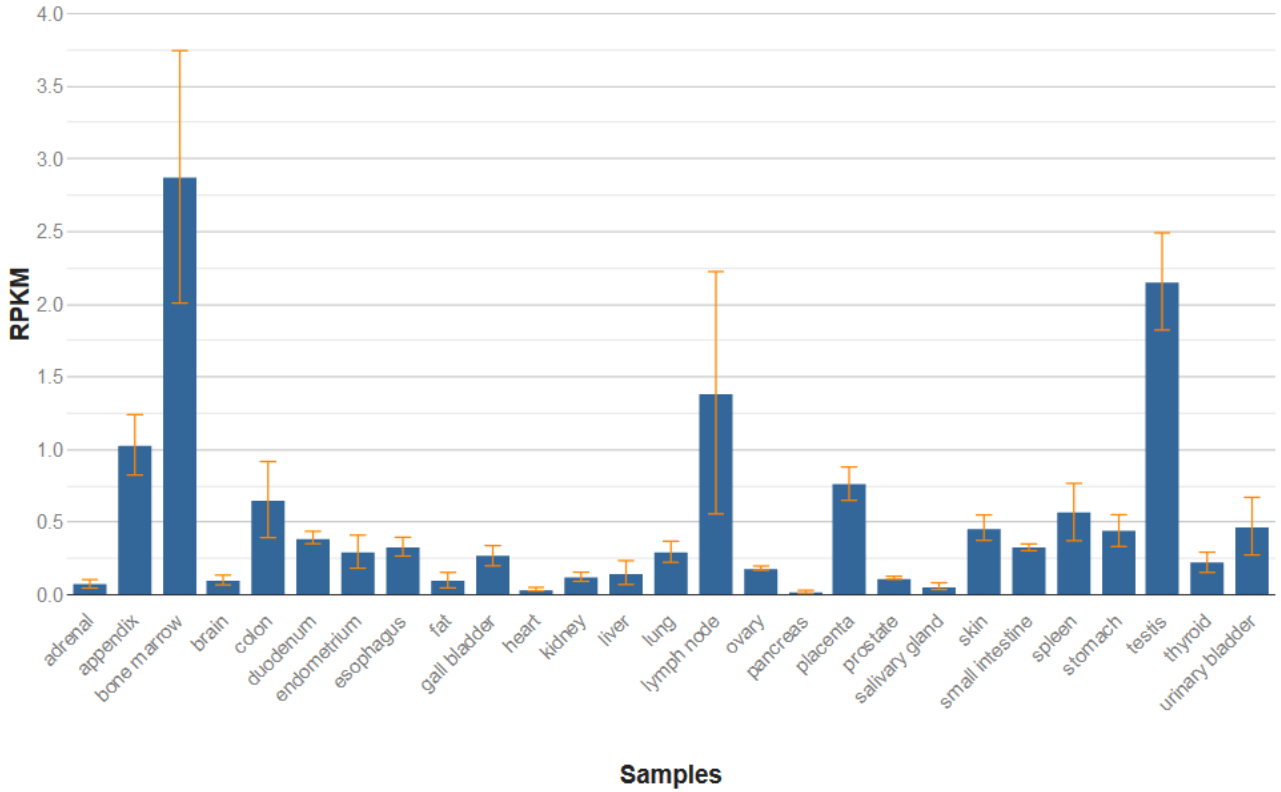
Databases

Summary

Official Symbol	BRCA2 provided by HGNC
Official Full Name	BRCA2 DNA repair associated provided by HGNC
Primary source	HGNC:HGNC:1101
See related	Ensembl:ENSG00000139618 MIM:600185
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	FAD; FACD; FAD1; GLM3; BRCC2; FANCD; PNCA2; FANCD1; XRCC11; BROVCA2
Summary	Inherited mutations in BRCA1 and this gene, BRCA2, confer increased lifetime risk of developing breast or ovarian cancer. Both BRCA1 and BRCA2 are involved in maintenance of genome stability, specifically the homologous recombination pathway for double-strand DNA repair. The BRCA2 protein contains several copies of a 70 aa motif called the BRC motif, and these motifs mediate binding to the RAD51 recombinase which functions in DNA repair. BRCA2 is considered a tumor suppressor gene, as tumors with BRCA2 mutations generally exhibit loss of heterozygosity (LOH) of the wild-type allele. [provided by RefSeq, Dec 2008]
Expression	Broad expression in bone marrow (RPKM 2.9), testis (RPKM 2.2) and 17 other tissues See more
Orthologs	mouse all

Databases

Expression



Genomic context

Bibliography

Variation

Databases


BIER
Collaborative Spanish Variant Server
BIER
CSVS
Start
Overview
Welcome to the Collaborative Spanish Variant Server. CSVS was created to provide information about the variability of the Spanish population to the scientific/medical community. It is useful for filtering polymorphisms and local variations in the process of prioritizing candidate disease genes. CSVS currently stores information on 1644 unrelated Spanish individuals. We accept submissions from WES or WGS.
Supported by
ciberer BIER
ciberer
Fundación Progreso y Salud
CONSEJERÍA DE SALUD
elixir
CSVS: created by Clinical Bioinformatics Area
Fundación Progreso y Salud
2015-2017

Databases

Position

Chromosomal location:

1:1-1000000,2:1-1000000

Gene:

BRCA2

Subpopulations

- [x] MGP (267 healthy controls)
- [x] IBS (107 Spanish individuals from 1000genomes)
- [x] Healthy controls
- [x] I Certain infectious and parasitic diseases
- [x] II Neoplasms
- [x] III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- [x] IV Endocrine, nutritional and metabolic diseases
- [x] V Mental and behavioural

Chr	Position	Alleles	Gene	Id	MAF							1000G MAF (phase 3)		ExAC	ESP 6500	
					Genotypes				Freq.							
					0/0	0/1	1/1	./.	0 Freq	1 Freq	MAF	ALL	EUR	ALL	ALL	Eur. An
13	32889792	A>G	ZAR1L,BRCA2	rs206118	1519	5	2	56	0.997	0.003	0.003	0.149	0.193	.	.	.
13	32889968	G>A	ZAR1L,BRCA2	rs206119	1581	0	1	0	0.999	0.001	0.001	0.259	0.216	.	.	.
13	32890404	A>G	BRCA2,ZAR1L	.	1581	1	0	0	1	0	0	.	.	.	.	.
13	32890523	C>G	ZAR1L,BRCA2	rs183997506	1579	3	0	0	0.999	0.001	0.001	0.000	.	.	.	.
13	32890572	G>A	ZAR1L,BRCA2	rs1799943	920	547	100	15	0.762	0.238	0.238	0.209	0.216	0.247	0.209	0.263
13	32890583	A>C	BRCA2,ZAR1L	rs138705202	1581	1	0	0	1	0	0	0.001	.	0.000	0.000	.
13	32890587	C>T	BRCA2,ZAR1L	rs76874770	1643	1	0	0	1	0	0	0.004	.	0.002	0.006	0.000
13	32890629	T>A	BRCA2	.	1643	0	0	1	1	0	0	.	.	.	.	.
13	32890726	T>G	BRCA2,ZAR1L	rs11571574	1436	1	0	145	1	0	0	0.003	0.002	.	0.004	0.004
13	32893197	AT>A,ATT	BRCA2,ZAR1L	.	1562	14	0	6	0.996	0.004	0.004	.	.	.	.	.

Page

Detailed information — Window size: 121 nts

32890490 — 32890550

Sequence

Gene

SNP

rs751297272 T > C
rs183997506 C > G
rs276174798 TCT >
rs276174797 T > C
rs757349122 T > C
rs276174795 T > C
rs276174799 G > A
rs276174800 T >
rs397507566 T > C
rs397507260 T > C
rs276174801 T > C
rs763296677 A > C
rs276174796 T > C
rs398122721 C > T
rs545355025 C > A
rs751741252 A > G
rs750780826 T >
rs764618190 A > G
rs754814638 G >
rs758732038 AG >
rs775945279 C > T
rs138705202 A > C

CSVS

13:32890523-32890523
13:32890572-32890572
13:32890587-3289
13:32890583-32890583

Databases

dbSNP Short Genetic Variations

Search for rs | Search

Example: rs268

Reference SNP (rs) Report

Download

Switch to classic site

rs1799943

Current Build 153
Released July 9, 2019

Organism	*Homo sapiens*	**Clinical Significance**	Reported in ClinVar
Position	chr13:32316435 (GRCh38.p12)	**Gene : Consequence**	BRCA2 : 5 Prime UTR Variant
Alleles	G>A / G>C / G>T	**Publications**	10 citations
Variation Type	SNV Single Nucleotide Variation	**Genomic View**	See rs on genome
Frequency	A=0.24553 (61394/250044, GnomAD_exome) A=0.21567 (27081/125568, TOPMED) A=0.24651 (29177/118358, ExAC) (+ 9 more)		

FEEDBACK

Databases

- Variant Details
- **Clinical Significance**
- Frequency
- Aliases
- Submissions
- History
- Publications

Allele: A (allele ID: 131503)

ClinVar Accession	Disease Names	Clinical Significance
RCV000112977.3	Breast-ovarian cancer, familial 2	Benign
RCV000114981.1	Familial cancer of breast	Not-Provided
RCV000246798.2	not specified	Benign
RCV000312794.1	Hereditary breast and ovarian cancer syndrome	Benign
RCV000397056.1	Fanconi anemia	Benign
RCV000580284.1	Hereditary cancer-predisposing syndrome	Benign
RCV000755477.1	not provided	Benign

Databases

NCBI Resources How To Sign in to NCBI

GEO Home | Documentation | Query & Browse | Email GEO

Gene Expression Omnibus

GEO Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Keyword or GEO Accession **Search**

Getting Started

- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

Tools

- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- Studies with Genome Data Viewer Tracks
- Programmatic Access
- FTP Site

Browse Content

Repository Browser

DataSets:	4348
Series:	114167
Platforms:	19798
Samples:	3098023

Databases

GEO DataSets

5. **BRCA2** abrogation triggers innate immune responses potentiated by treatment with PARP inhibitors

(Submitter supplied) Heterozygous germline mutations in **BRCA2** predispose to breast and ovarian cancer. Contrary to non-cancerous cells, where **BRCA2** deletion causes cell cycle arrest or cell death, **BRCA2** inactivation in tumors is associated with uncontrolled cell proliferation. We set out to investigate this conundrum by exploring modalities of cell adaptation to loss of **BRCA2** and focused on genome-wide transcriptome alterations. more...

Organism: **Homo sapiens**
Type: Expression profiling by high throughput sequencing
Platform: GPL20301 48 Samples
Download data: TXT
Series Accession: GSE123631 ID: 200123631
SRA Run Selector

6. Breast tumor subtypes correlate with prognosis

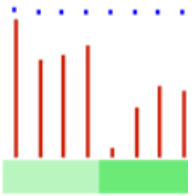
(Submitter supplied) To advance in our understanding of the biological pathways involved in **breast cancer** tumor progression we have analyzed a set of breast tumor biopsies in order to identify the genomic pathways in which tumor may develop. With this objective, a cDNA microarray platform containing 800 genes was constructed. These genes were chosen because they are in several representatives signaling pathways, namely estrogen and progesterone receptor related pathways, cell cycle, DNA repair, chromatin remodeling, cell proliferation, apoptosis, cell adhesion, cell invasion and angiogenesis. more...

Organism: **Homo sapiens**
Type: Expression profiling by array
Platform: GPL5953 111 Samples
Download data: GPR
Series Accession: GSE18908 ID: 200018908
Analyze with GEO2R

Databases

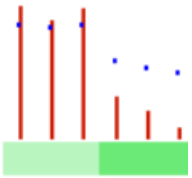
GEO Profiles

11. Bccip - Niacin-bound chromium effect on the subcutaneous fat tissue of a model for type 2 diabetes with obesity

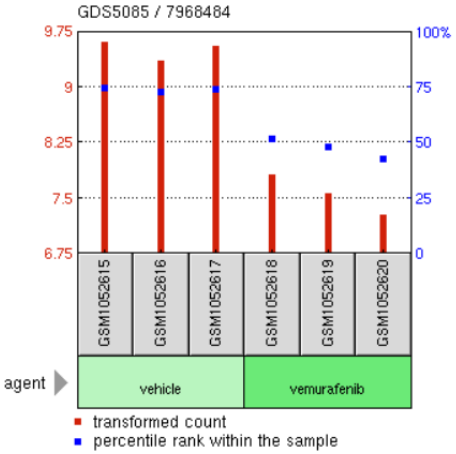


Annotation: Bccip, **BRCA2** and CDKN1A interacting protein
Organism: Mus musculus
Reporter: GPL1261, 1448542_at (ID_REF), GDS2605, 66165 (Gene ID), NM_025392
DataSet type: Expression profiling by array, count, 8 samples
ID: 35750542
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Homologene neighbors

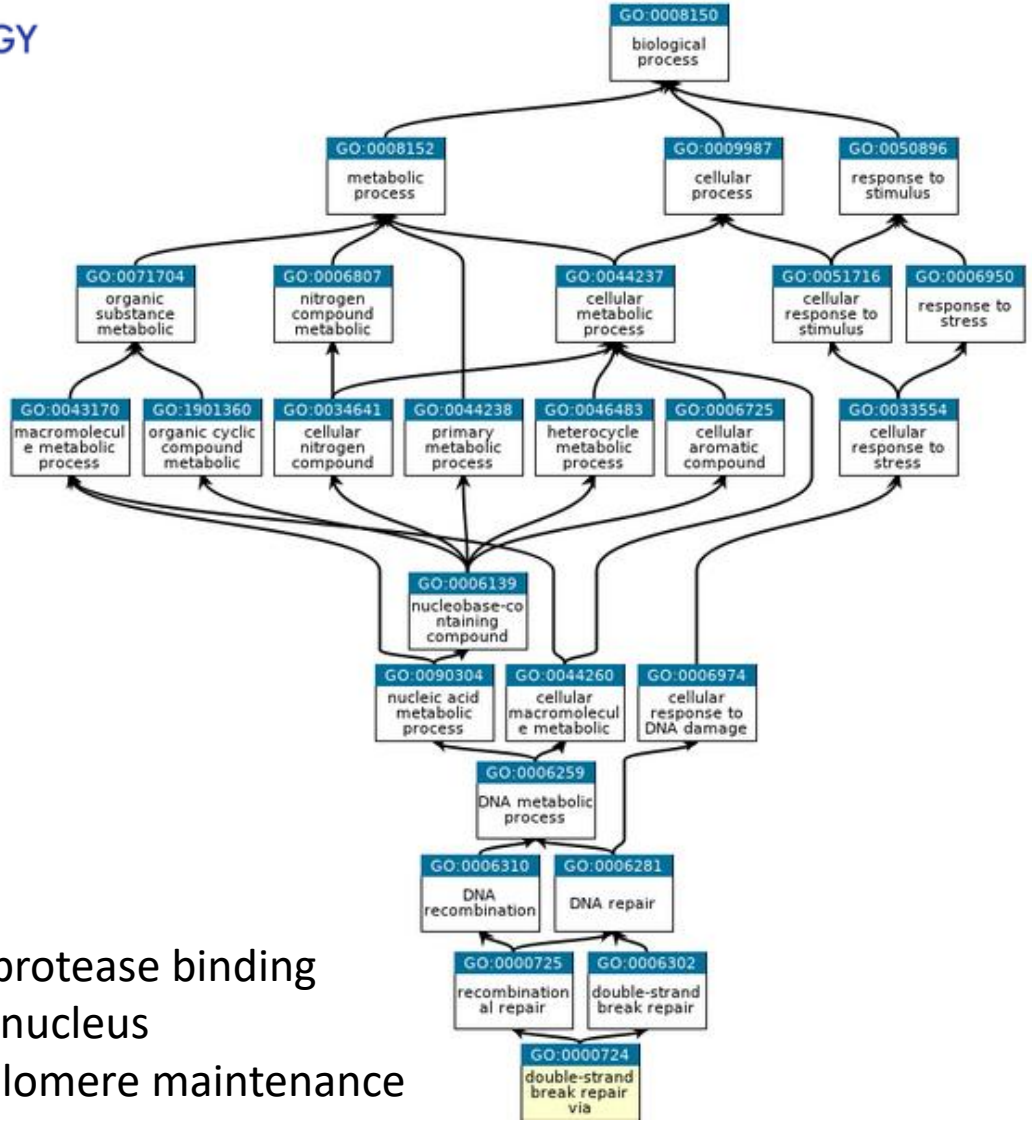
12. **BRCA2** - Oncogenic BRAF harboring melanoma cell line response to BRAF inhibition



Annotation: **BRCA2**, **BRCA2**, DNA repair associated
Organism: Homo sapiens
Reporter: GPL6244, 7968484 (ID_REF), GDS5085, NM_000059, DQ897648, U43746, chr13:32889617-32973809 (SPOT ID)
DataSet type: Expression profiling by array, transformed count, 6 samples
ID: 112035057
GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors



Databases

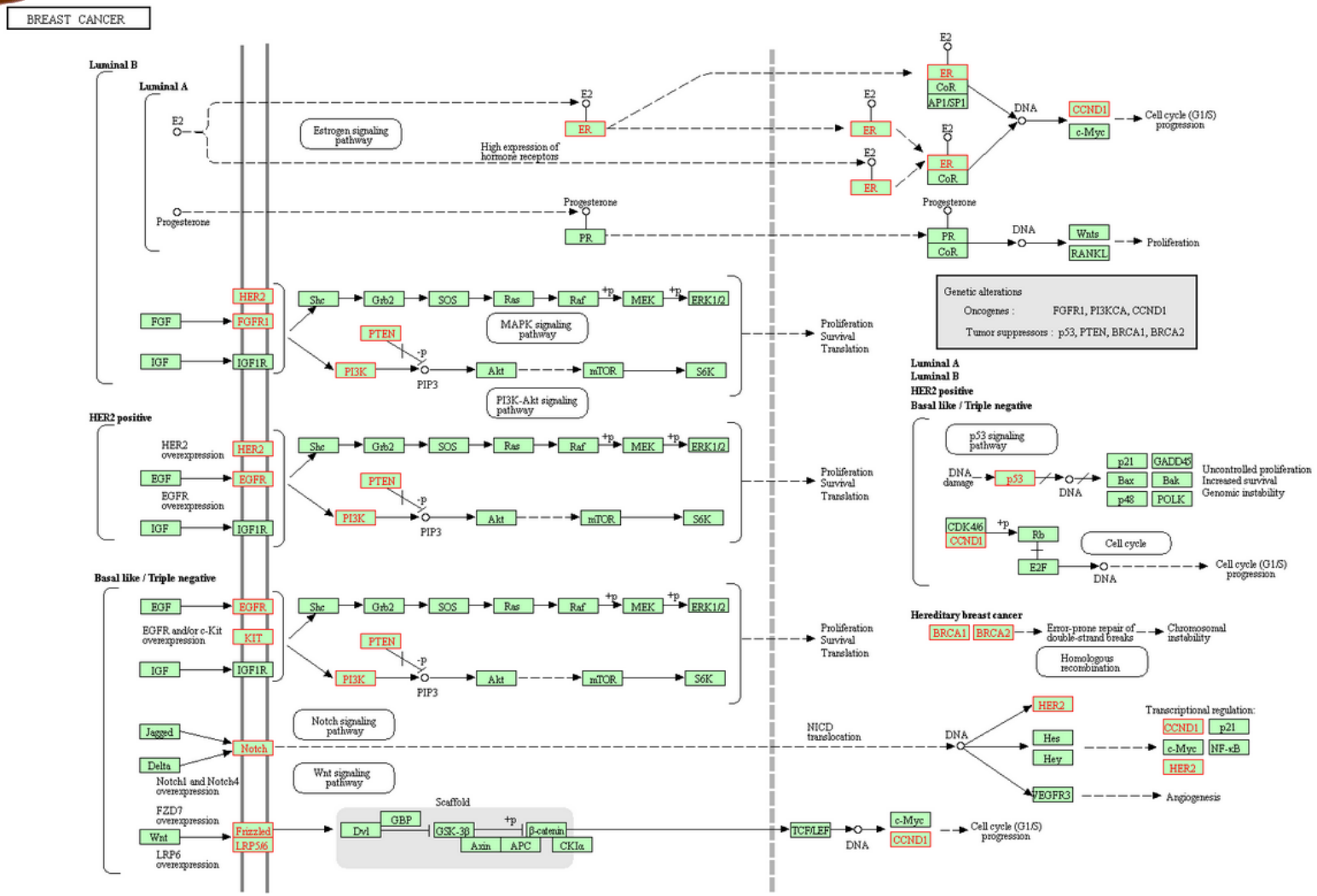


Molecular function: protease binding
Cellular component: nucleus
Biological process: telomere maintenance

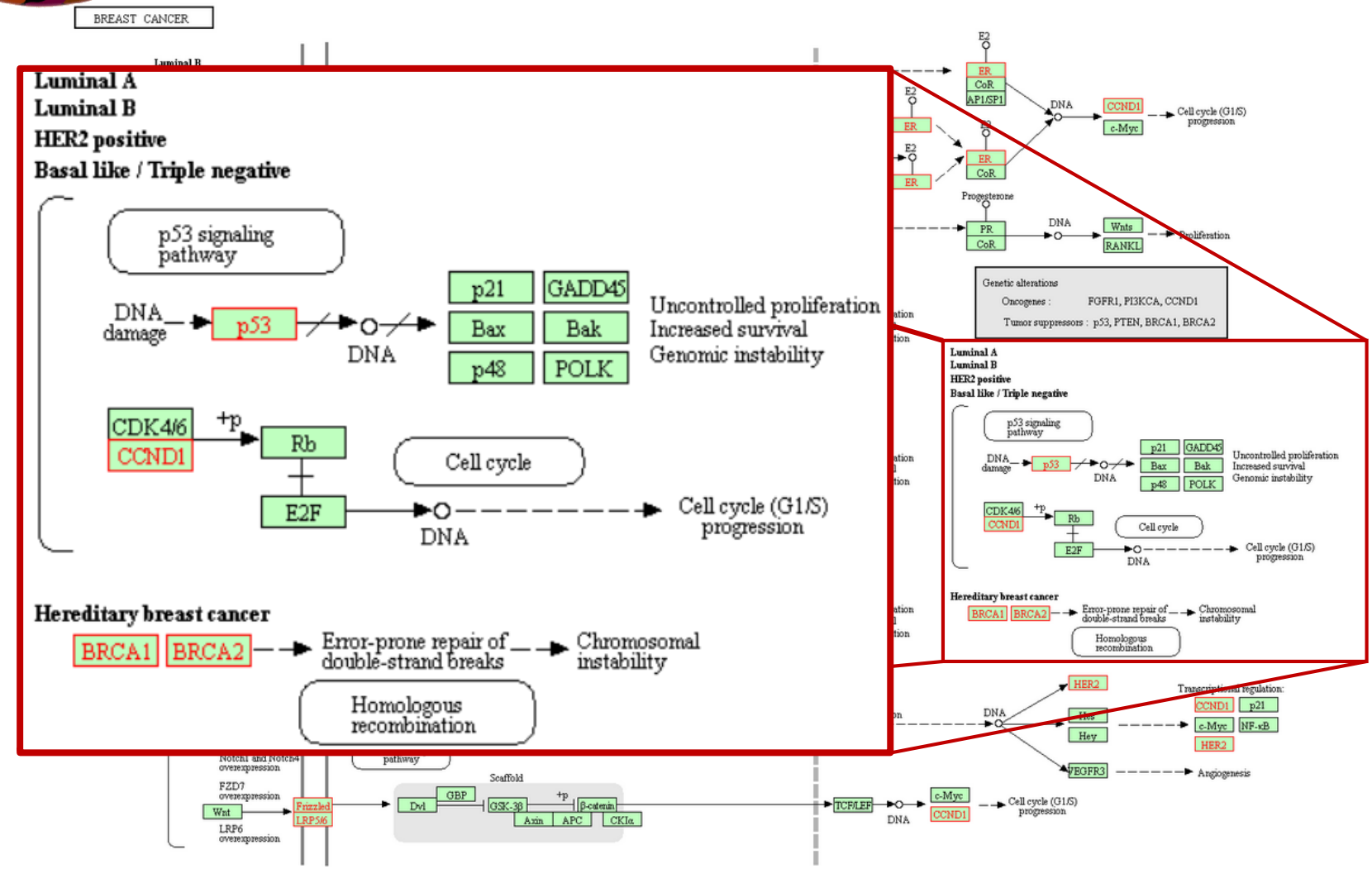
Databases

Entry	675 CDS T01001
Gene name	BRCA2, BRCC2, BROVCA2, FACD, FAD, FAD1, FANCD, FANCD1, GLM3, PNCA2, XRCC11
Definition	(RefSeq) BRCA2 DNA repair associated
KO	K08775 breast cancer 2 susceptibility protein
Organism	hsa Homo sapiens (human)
Pathway	hsa03440 Homologous recombination hsa03460 Fanconi anemia pathway hsa05200 Pathways in cancer hsa05212 Pancreatic cancer hsa05224 Breast cancer
Disease	H00019 Pancreatic cancer H00027 Ovarian cancer H00031 Breast cancer H00238 Fanconi anemia H01554 Fallopian tube cancer

Databases

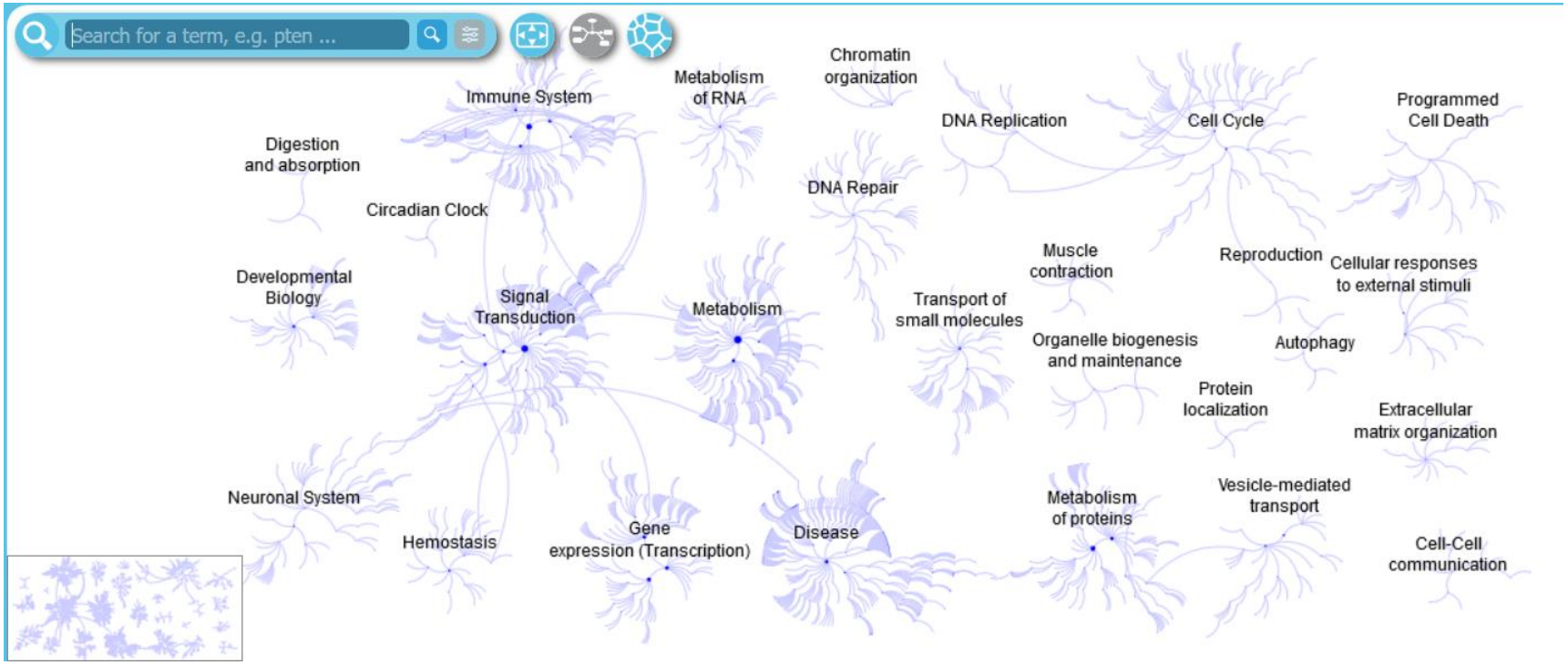

BREAST CANCER
Luminal B
Luminal A
E2
Estrogen signaling pathway
ER
High expression of hormone receptors
Progesterone
PR
CoR
AP1/SP1
DNA
CCND1
c-Myc
Cell cycle (G1/S) progression
Wnts
RANKL
Proliferation
Genetic alterations
Oncogenes : FGFR1, PI3KCA, CCND1
Tumor suppressors : p53, PTEN, BRCA1, BRCA2
HER2
FGF
FGFR1
IGF
IGF1R
Shc
Grb2
SOS
Ras
Raf
MEK
ERK1/2
MAPK signaling pathway
PTEN
PI3K
PIP3
Akt
mTOR
S6K
PI3K-Akt signaling pathway
Proliferation
Survival
Translation
HER2 positive
HER2 overexpression
EGF
EGFR
EGFR overexpression
Luminal A
Luminal B
HER2 positive
Basal like / Triple negative
p53 signaling pathway
DNA damage
p53
p21
GADD45
Bax
Bak
p48
POLK
Uncontrolled proliferation
Increased survival
Genomic instability
CDK4/6
CCND1
Rb
E2F
Cell cycle
Basal like / Triple negative
EGFR and/or c-Kit overexpression
KIT
Hereditary breast cancer
BRCA1
BRCA2
Error-prone repair of double-strand breaks
Chromosomal instability
Homologous recombination
Jagged
Delta
Notch
Notch1 and Notch4 overexpression
Notch signaling pathway
NICD translocation
HER2
Hes
Hey
VEGFR3
Angiogenesis
Transcriptional regulation:
CCND1
p21
c-Myc
NF-κB
HER2
FZD7 overexpression
Wnt
LRP6 overexpression
Frizzled
LRP5/6
Wnt signaling pathway
Scaffold
Dvl
GBP
GSK-3β
β-catenin
Axin
APC
CKIα
TCF/LEF
c-Myc
CCND1
Cell cycle (G1/S) progression

Databases


BREAST CANCER
Luminal A
Luminal B
HER2 positive
Basal like / Triple negative
p53 signaling pathway
DNA damage
p53
DNA
p21
GADD45
Bax
Bak
p48
POLK
Uncontrolled proliferation
Increased survival
Genomic instability
CDK4/6
CCND1
+p
Rb
E2F
Cell cycle
DNA
Cell cycle (G1/S) progression
Hereditary breast cancer
BRCA1
BRCA2
Error-prone repair of double-strand breaks
Chromosomal instability
Homologous recombination
Genetic alterations
Oncogenes : FGFR1, PI3KCA, CCND1
Tumor suppressors : p53, PTEN, BRCA1, BRCA2
Proliferation
Angiogenesis
Wnt
Frizzled
LRP5/6
Dvl
GBP
GSK-3β
β-catenin
Axin
APC
CKIα
Scaffold
TCF/LEF
c-Myc
NF-κB
HER2
Hey
VEGFR3

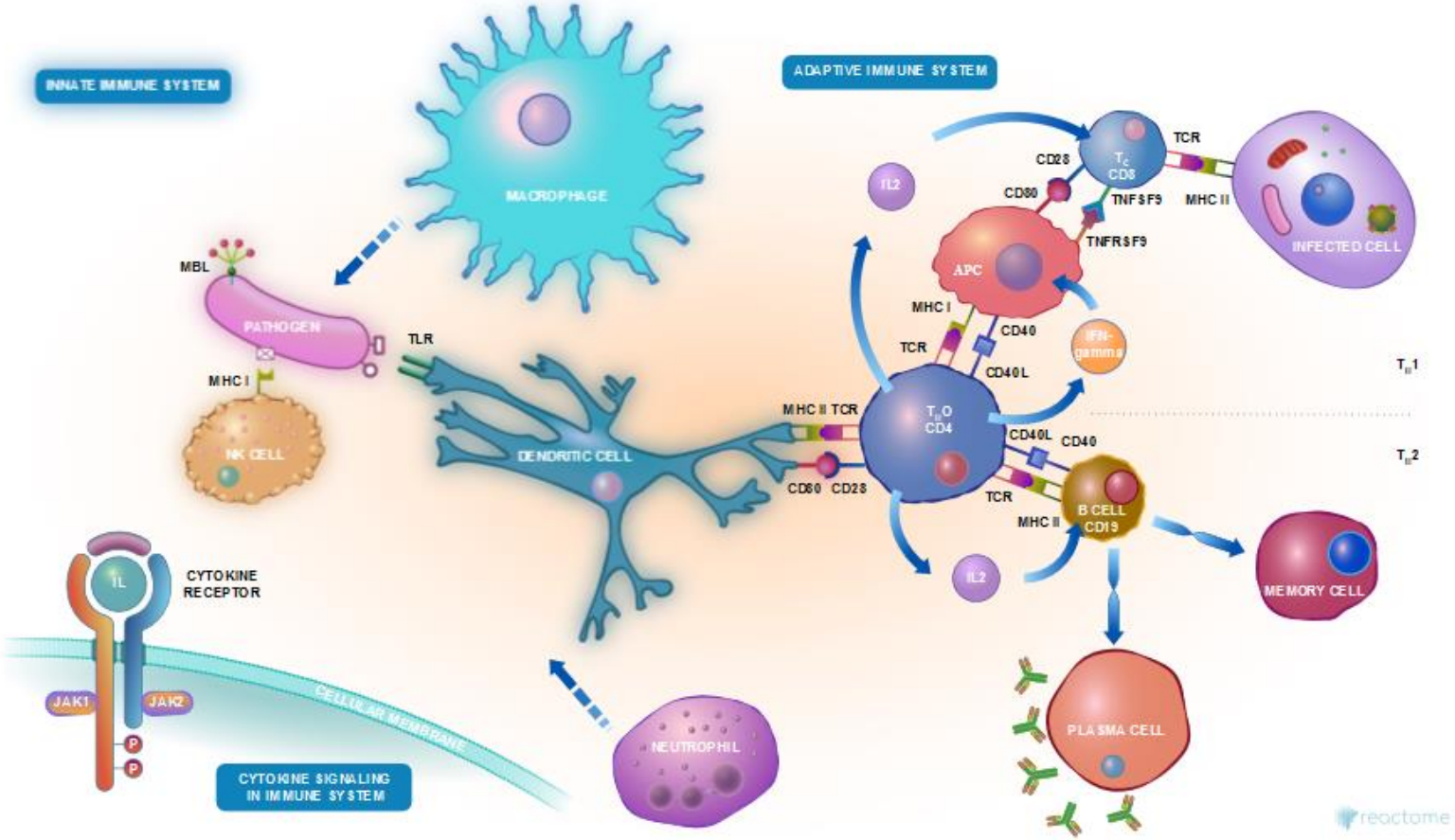
Databases

reactome



Databases

reactome



Databases

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (1+)
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

Protein **Breast cancer type 2 susceptibility protein**

Gene **BRCA2**

Organism *Homo sapiens (Human)*

Status Reviewed - Annotation score: ●●●●● - Experimental evidence at protein level

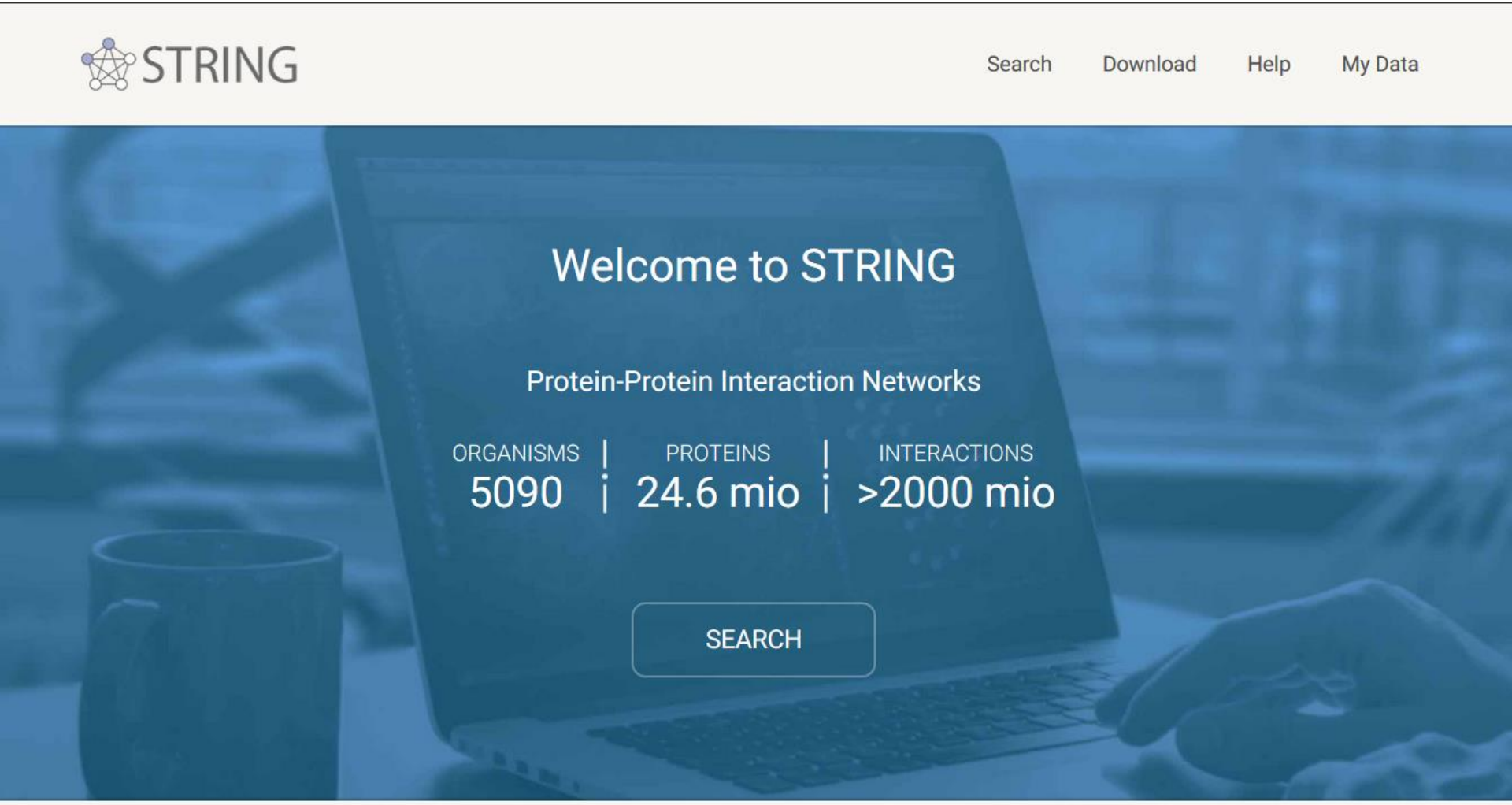
Keywords

Molecular function	DNA-binding
Biological process	Cell cycle, DNA damage, DNA recombination, DNA repair

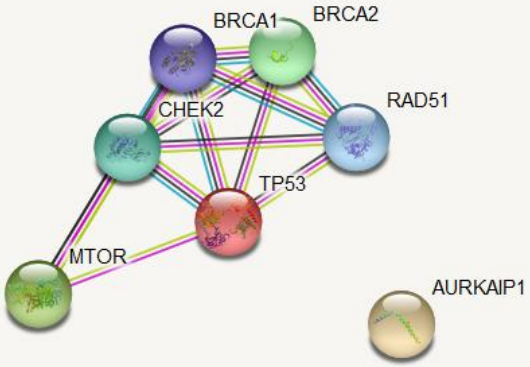
Enzyme and pathway databases

Reactome	R-HSA-5685942 HDR through Homologous Recombination (HRR) R-HSA-5693554 Resolution of D-loop Structures through Synthesis-Dependent Strand Annealing (SDSA) R-HSA-5693568 Resolution of D-loop Structures through Holliday Junction Intermediates R-HSA-5693579 Homologous DNA Pairing and Strand Exchange R-HSA-5693616 Presynaptic phase of homologous DNA pairing and strand exchange R-HSA-912446 Meiotic recombination

Databases


STRING
Search
Download
Help
My Data
Welcome to STRING
Protein-Protein Interaction Networks
ORGANISMS
5090
PROTEINS
24.6 mio
INTERACTIONS
>2000 mio
SEARCH
© STRING CONSORTIUM 2019
ABOUT
INFO
ACCESS
CREDITS

Databases



Viewers › | Legend ⌄ | Settings › | Analysis › | Exports › | Clusters › | More | Less

Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Color

colored nodes: query proteins and first shell of interactors

white nodes: second shell of interactors

Node Content

empty nodes: proteins of unknown 3D structure

filled nodes: some 3D structure is known or predicted

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.

Known Interactions

from curated databases

experimentally determined

Predicted Interactions

gene neighborhood

gene fusions

gene co-occurrence

Others

textmining

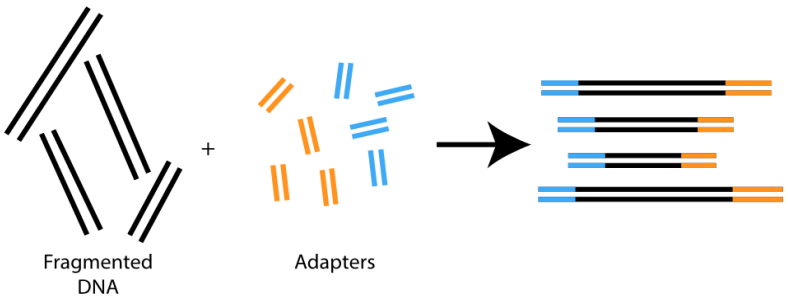
co-expression

protein homology

THANKS!

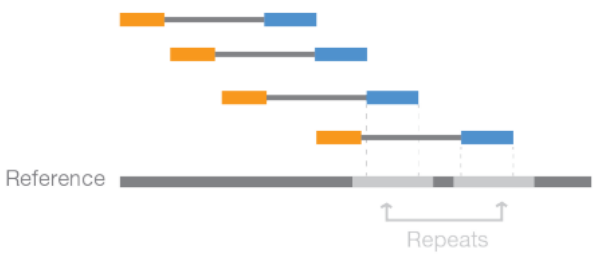
Types of libraries

Library preparation



Single-end sequencing

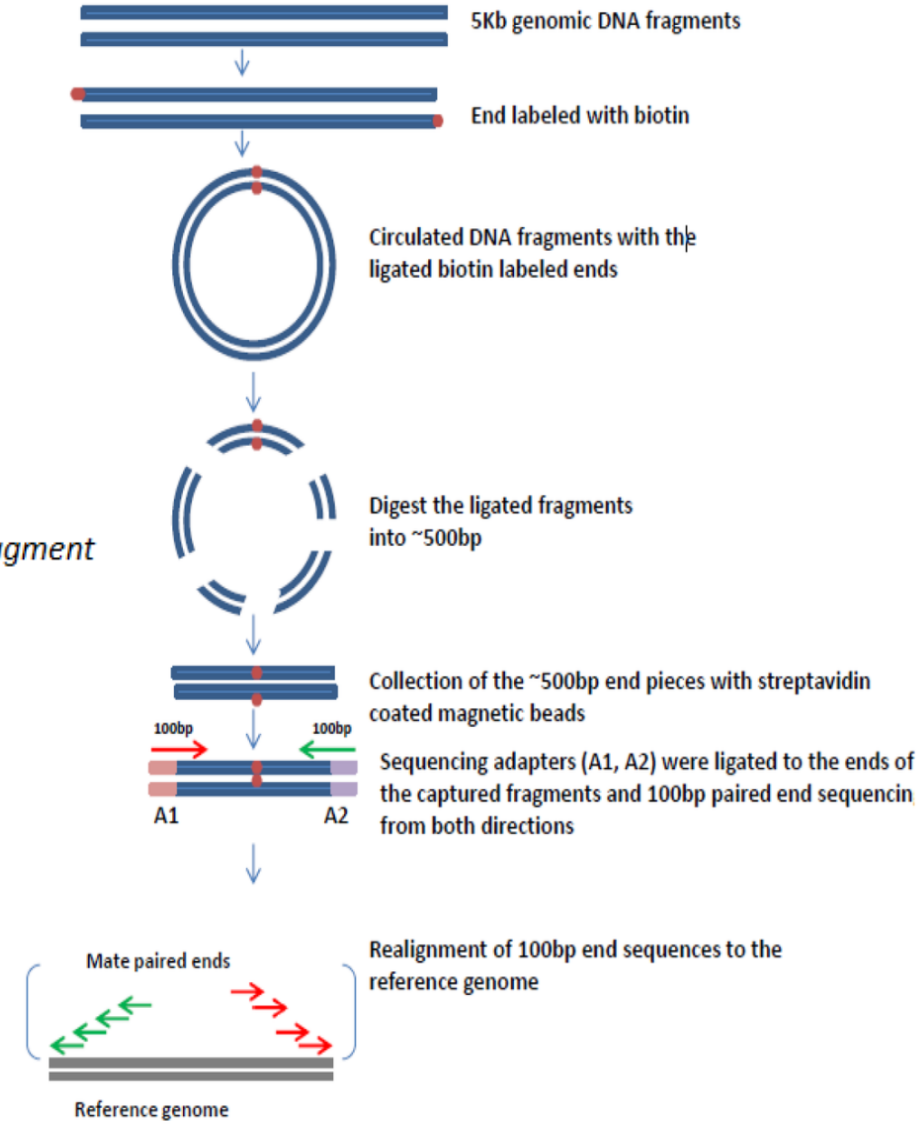
Paired-end sequencing



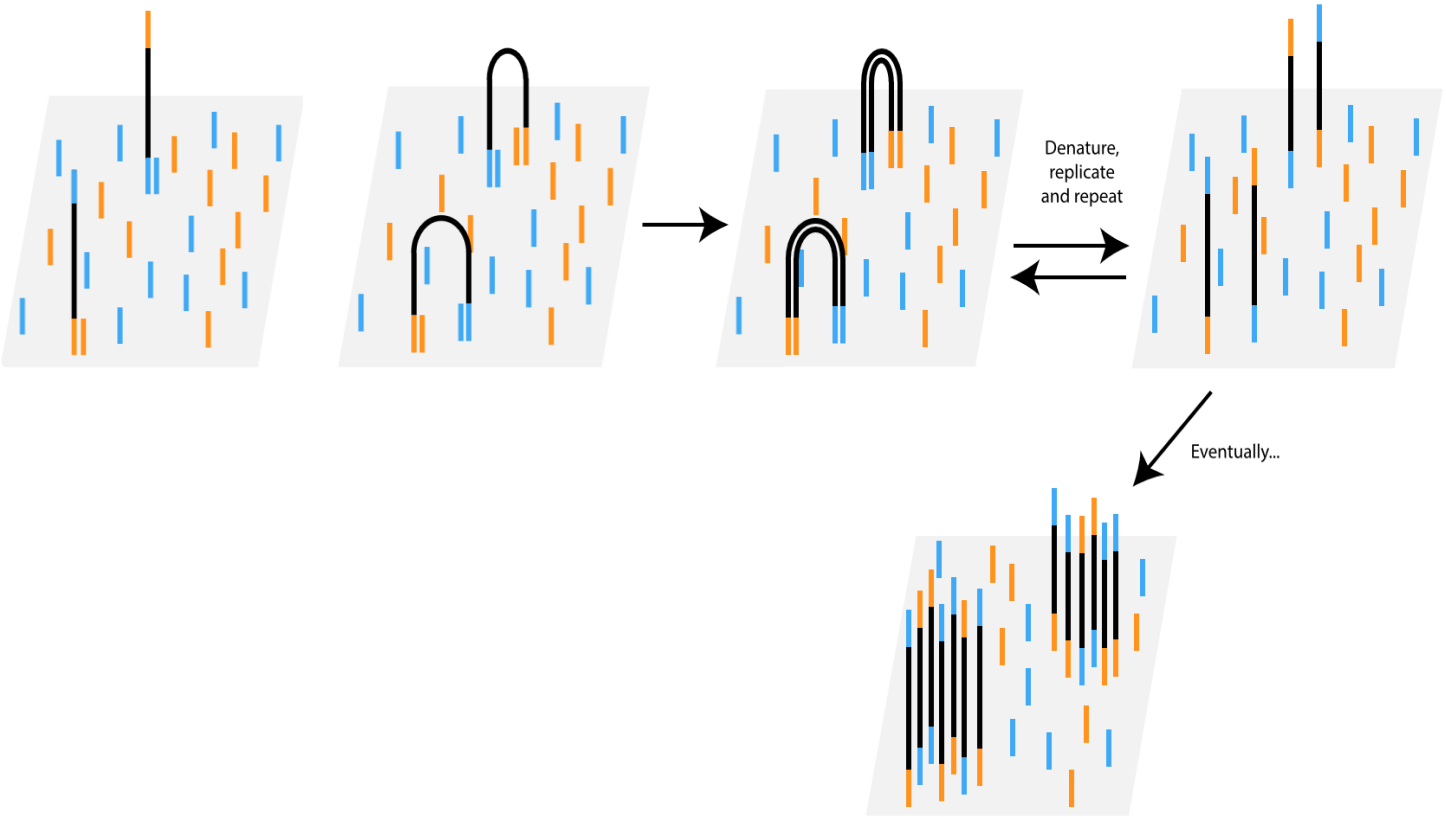
Types of libraries

Mate-pair sequencing

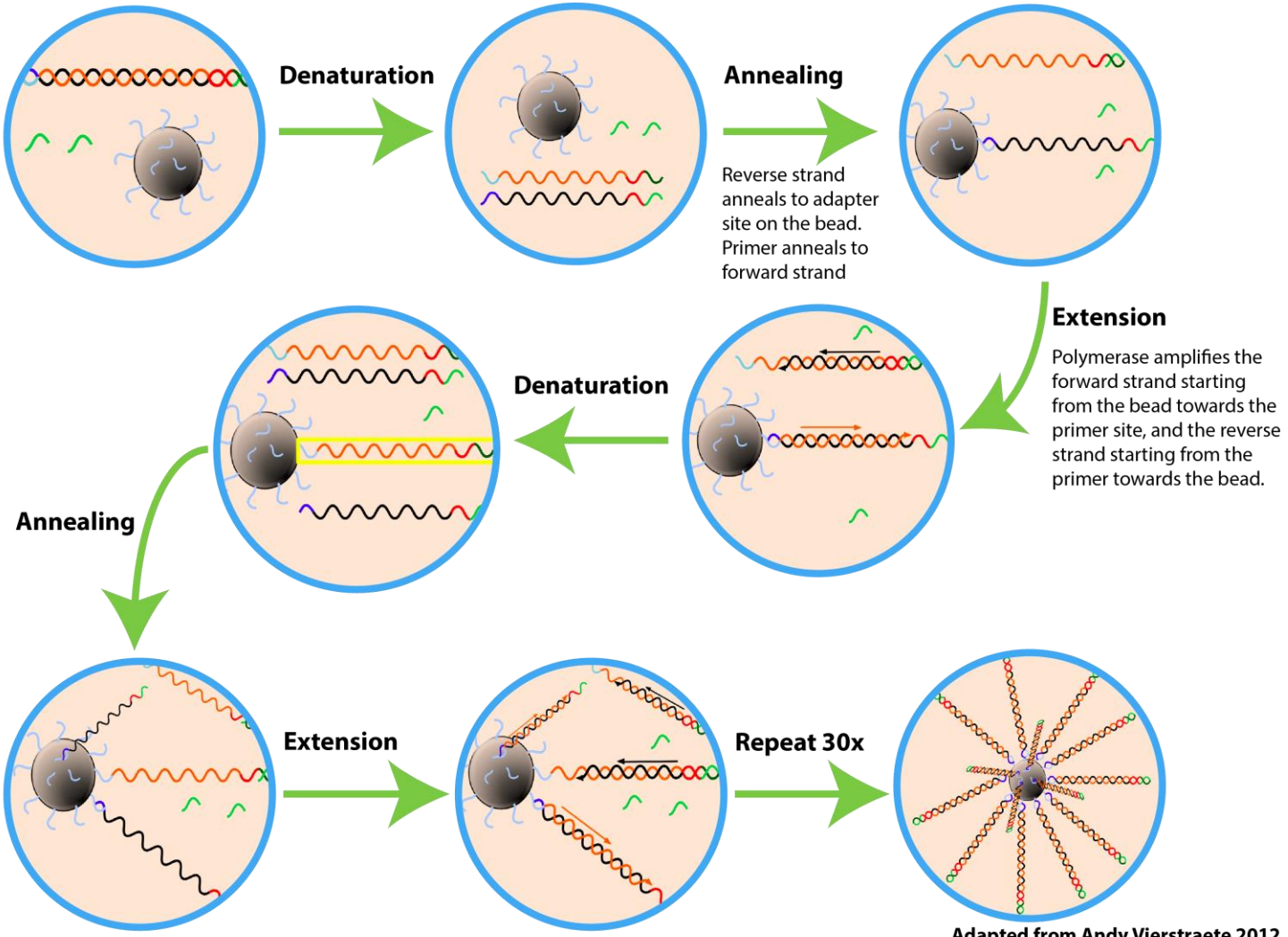
DNA fragment



Bridge PCR



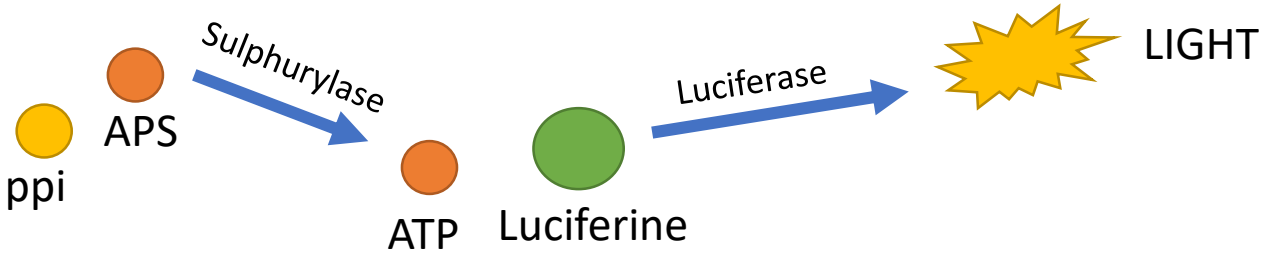
Emulsion PCR

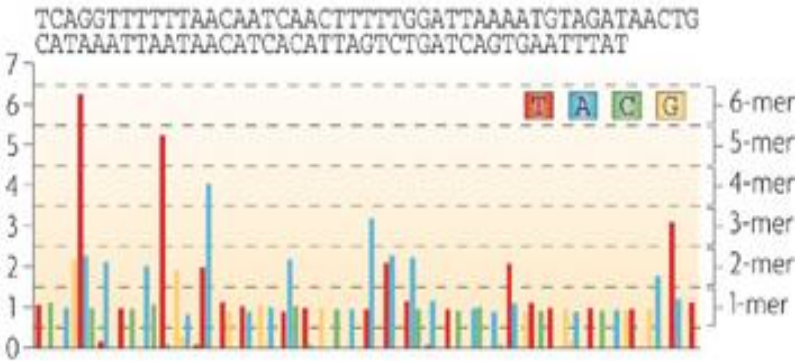


Adapted from Andy Vierstraete 2012

Roche

Pyrosequencing

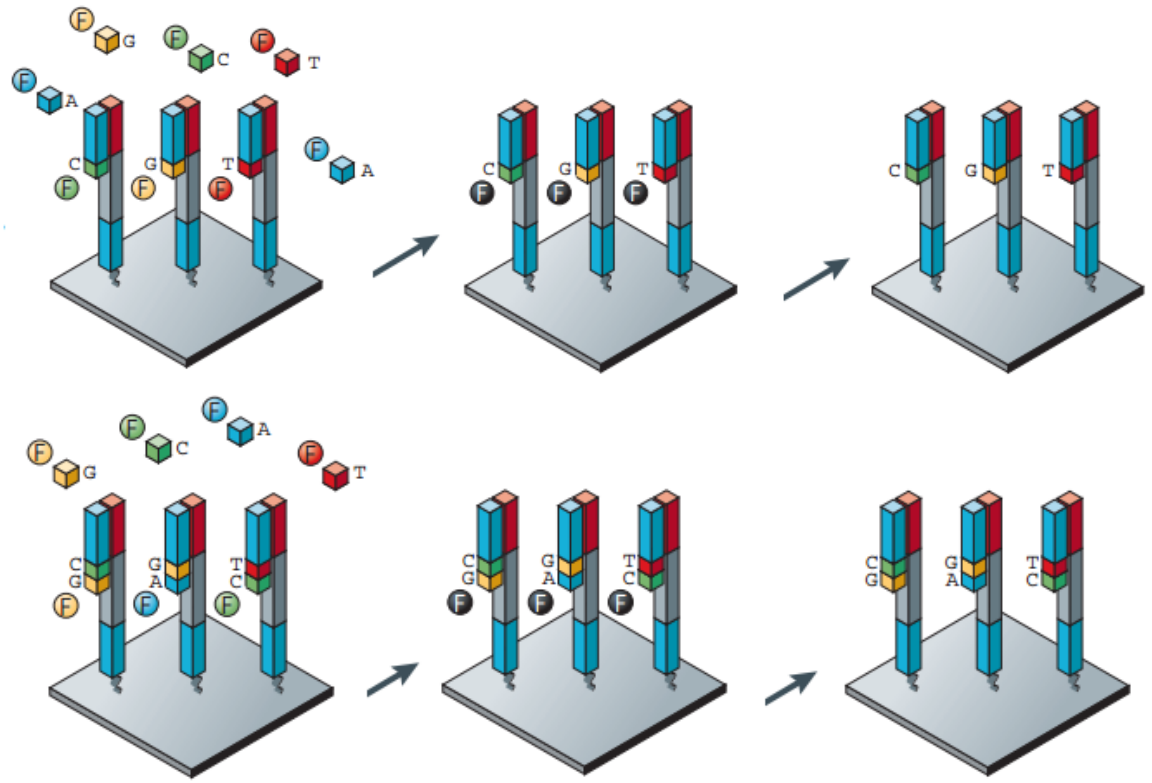




- Large read lenghts generation
- High reagent cost
- High error rate over strings of 6+ homopolymer

Illumina

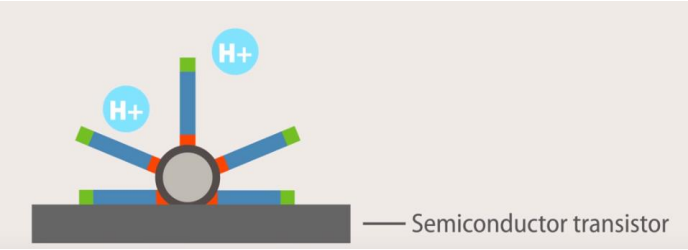
Sequencing by synthesis

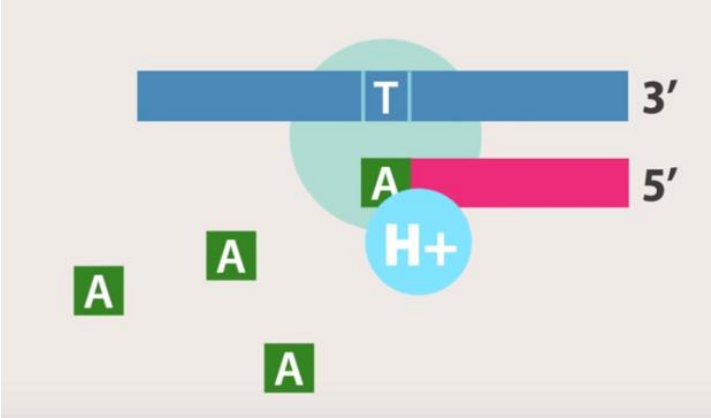


- Overcomes homopolymer issue due to terminated nucleotides
- Increased error rate with read length

Ion Torrent - Thermofisher

Ion semiconductor

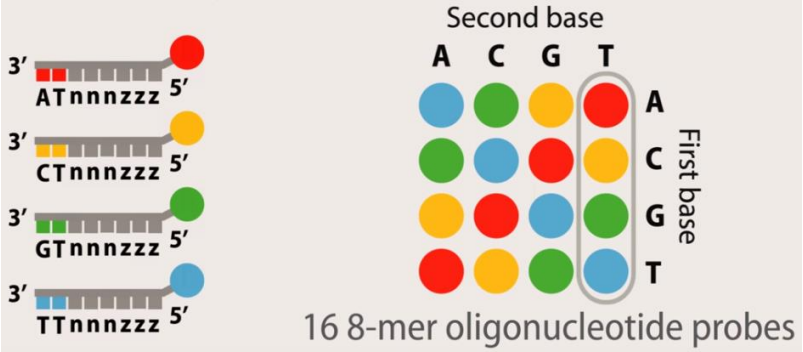


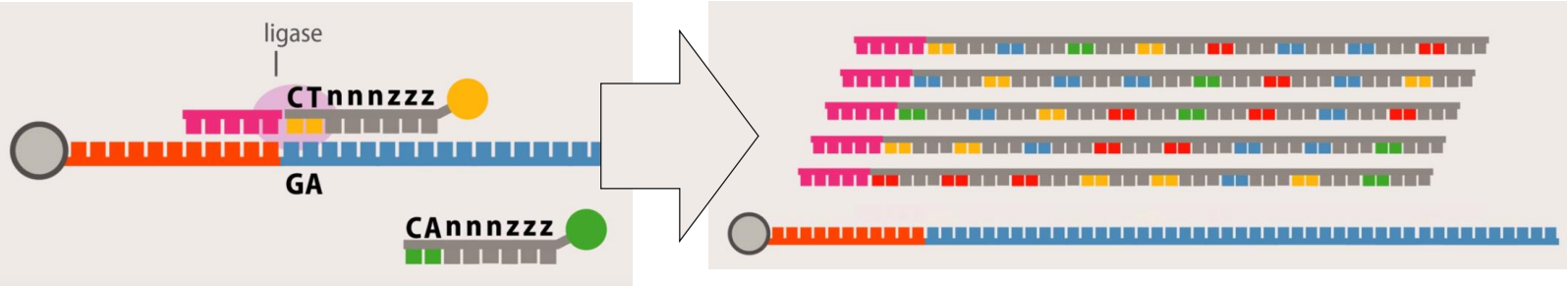


- Similar to pyrosequencing, but measures the release of H+ instead of pyrophosphate
- More cost-effective and time-efficient

SOLiD- Thermofisher

Sequencing by ligation

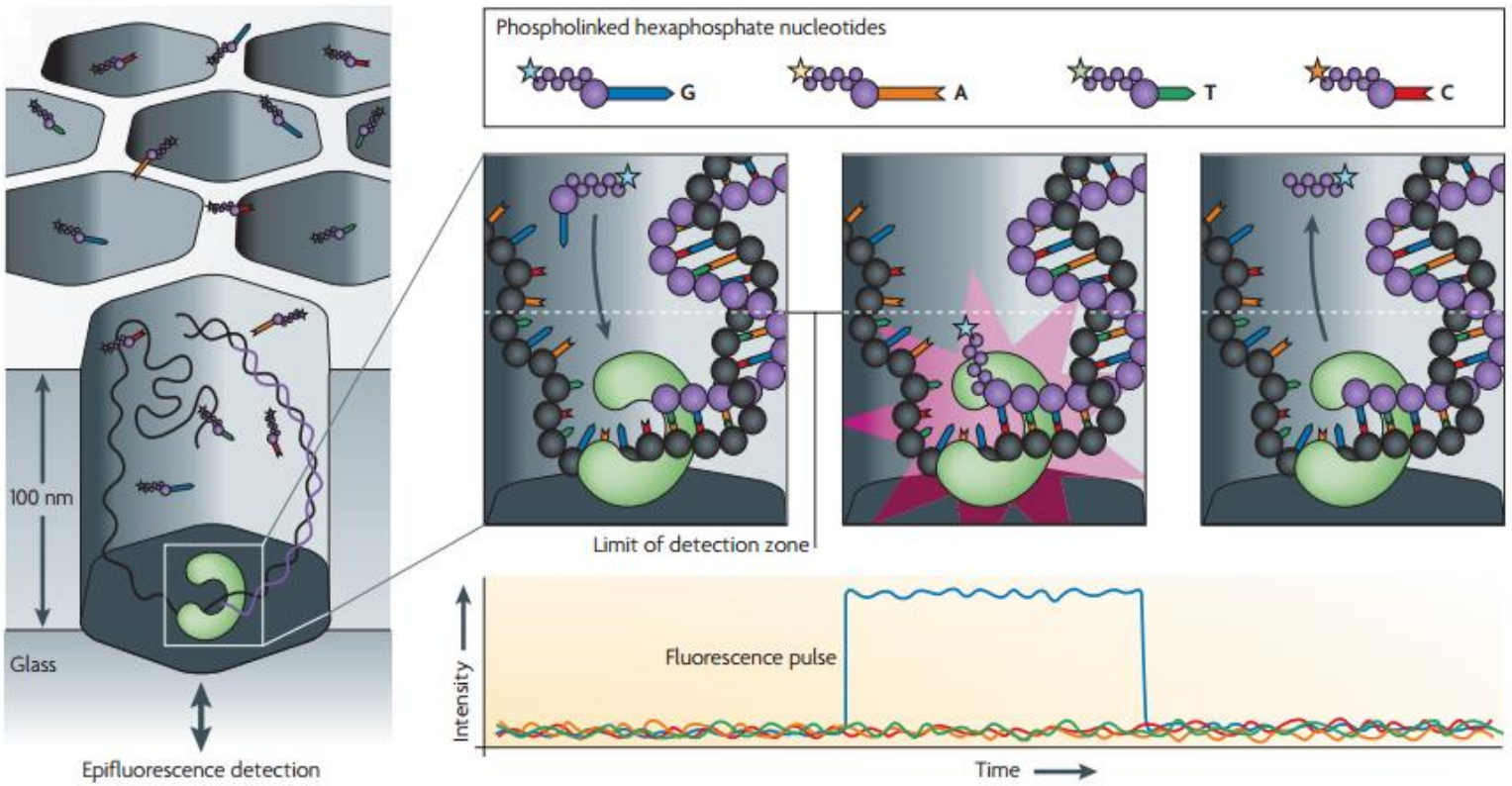




- Oligonucleotide probes used rather than DNA polymerase
- Very short read lengths

Pac Bio

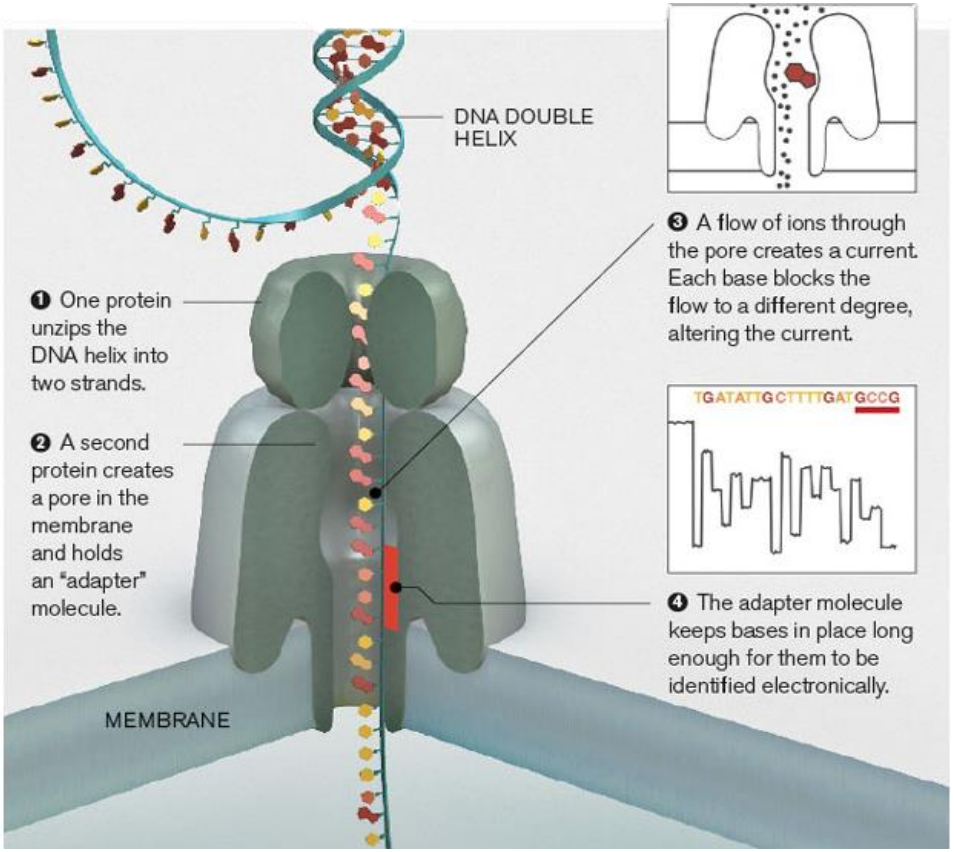
Real-time SMS



- Non-stop sequencing, no need to “wash and scan”
- DNApol fixed at the bottom of the well, the laser detector aims at the active site

Oxford Nanopore

Real-time SMS



- Non-stop sequencing, no need to "wash and scan"
- High error rate, this technology is still improving