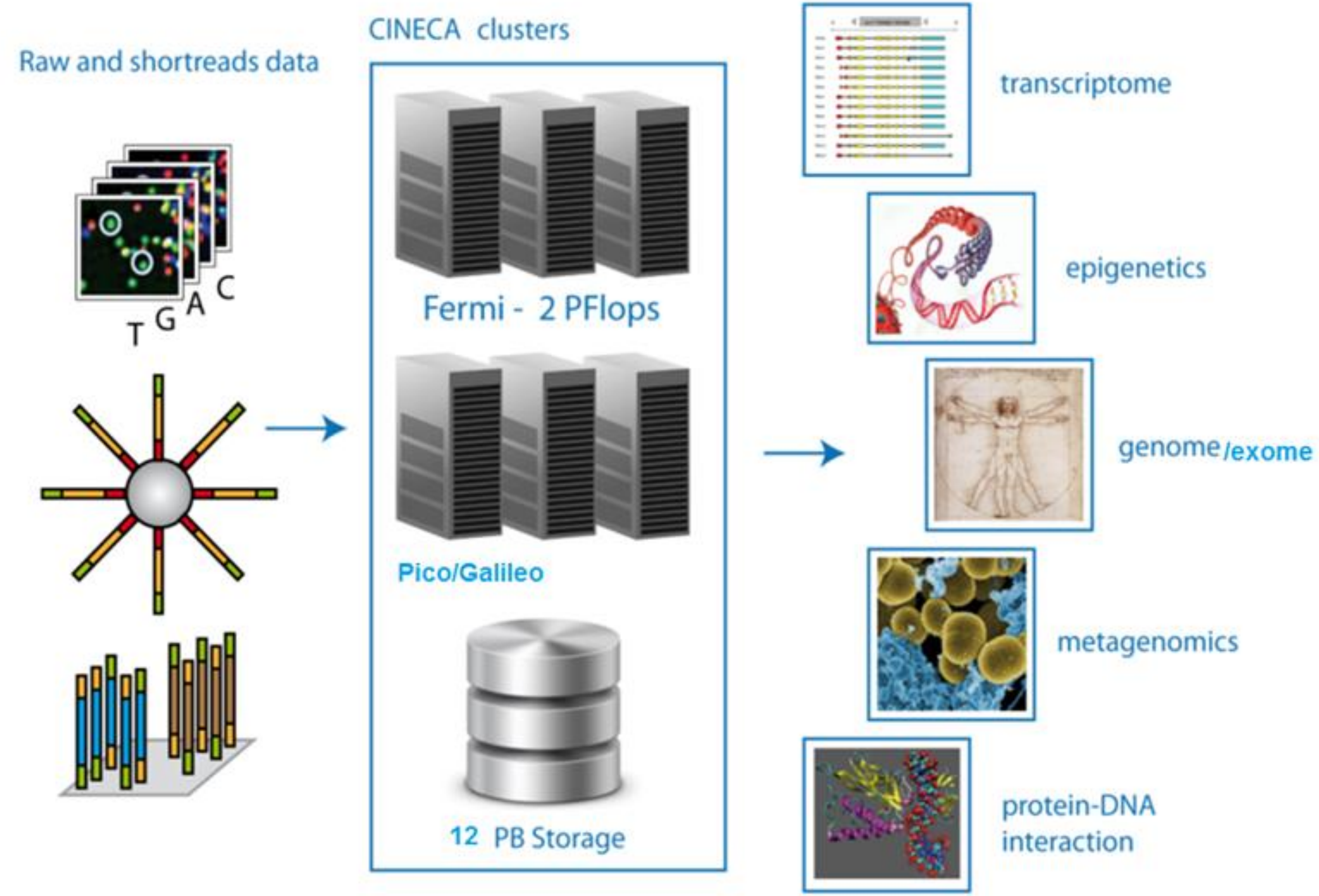
High-Performance Computing Bioinformatics data analysis environment @ CINECA

Tiziana Castrignanò (Cineca)

What does CINECA offer for bioinformatics?



Raw and shortreads data

CINECA clusters

Fermi - 2 PFlops

Pico/Galileo

12 PB Storage

transcriptome

epigenetics

genome/exome

metagenomics

protein-DNA interaction

What does CINECA offer for bioinformatics?

1. **Computing resources**

   Bioinformatics software available through command line

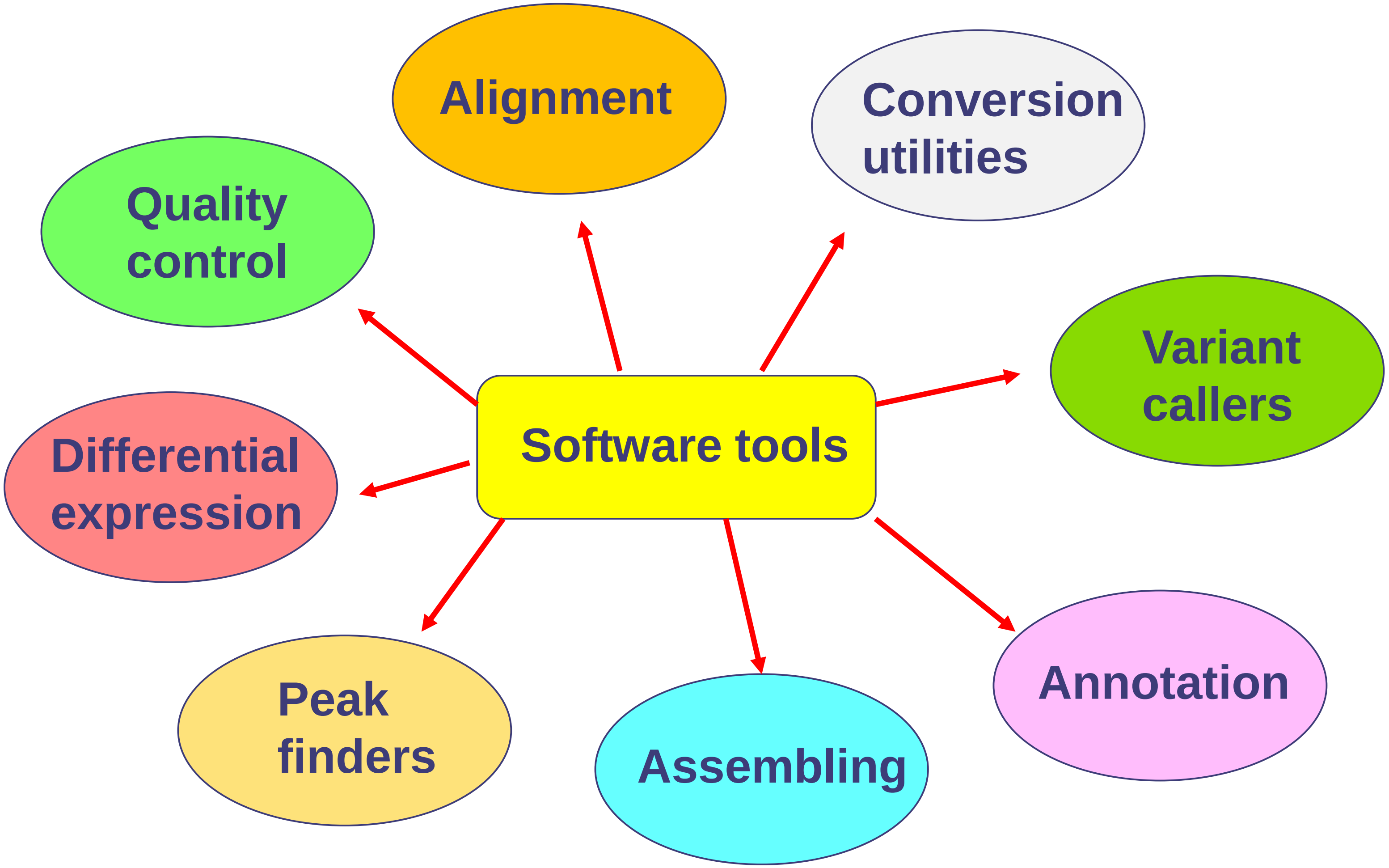
2. **Advanced services**

   Automated web workflows for Next Generation Sequencing

3. **Bioinformatics Expertise**

   To customize solutions or implement new systems and tools

Bioinformatics software available through command line



Software tools

Alignment

Conversion utilities

Variant callers

Annotation

Assembling

Peak finders

Differential expression

Quality control

Bioinformatics software available through command line

Quality control

fastqc
ngsqctoolkit
trimmomatic

Annotation

annovar
snpeff
ngsrich

Conversion utilities

samtools
bedtools
vcftools
sra
picard

Alignment

abra
diamond
bowtie
bwa
shrimp
tophat
blast+
mosaik
mauve
mummer
star
bismark

General Purpose

bioconductor
biopython
cluto
igvtools
idl
mrjob
r
emboss

Bioinformatics software available through command line

RNA-Seq

cufflinks
htseq
splicetrap
chimerascan
reditools

Peak finders

macs
peakranger
sicer

Variant callers

gatk
mutect
varscan2
lofreq

Assembling

spades
velvet
ray
cisa
pagit

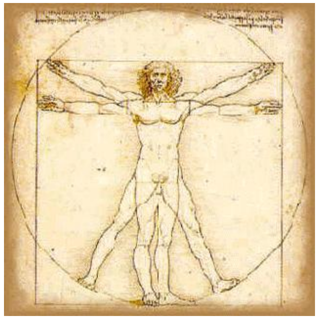
Metagenomics

concoct
qiime

Cineca can add new software under user requests

Available released genomes

Homo sapiens



hg18
hg19

Drosophila melanogaster

dm3

Zea mays

Maize3
ZmB73

Rattus norvegicus

rn4

Mus musculus

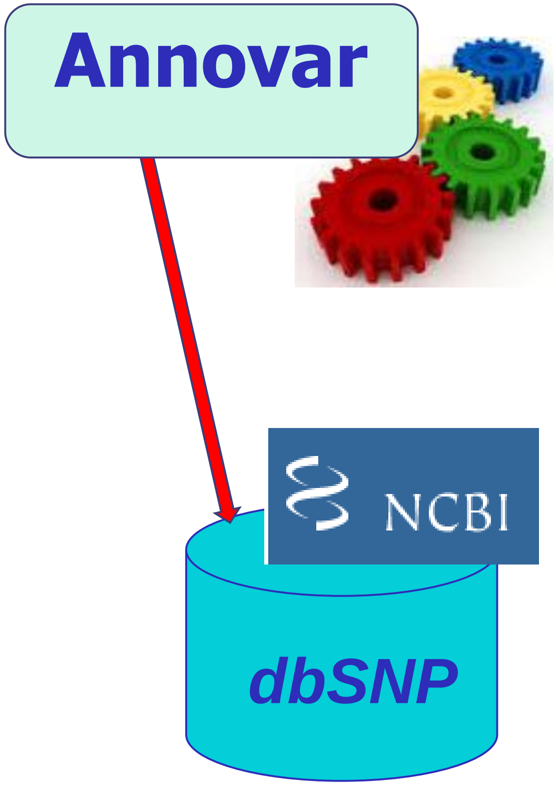
mm9
mm10

sacCer3

Saccharomyces cerevisiae

Annotation databases

ANNOVAR is an efficient software tool to utilize update-to-date information to functionally annotate genetic variants detected from diverse genomes



Cineca can add new genomes and annotation databases under user requests

ExAC Data Set:
exome sequencing data
from a wide variety of
large-scale sequencing projects

A free public archive
for short genetic variation
within and across different species

How to become Cineca user

- Please fill out the form on:

  https://userdb.hpc.cineca.it/user/register

- You'll receive userdb credentials: Then

➔ Click on “HPC Access” and follow the on-screen instructions

➔You'll be asked to upload an image of a valid ID document

➔ Ask your PI or send an email to superc@cineca.it to be included on an active project.

- When everything is done an automatic procedure sends you (via 2 separate emails) the username/password to access HPC systems

How to log in on PICO

All cluster HPC infrastructures are available for bioinformatics.

PICO is the infrastructure dedicated to NGS bioinformatics applications and big data.

Users can access trough command line

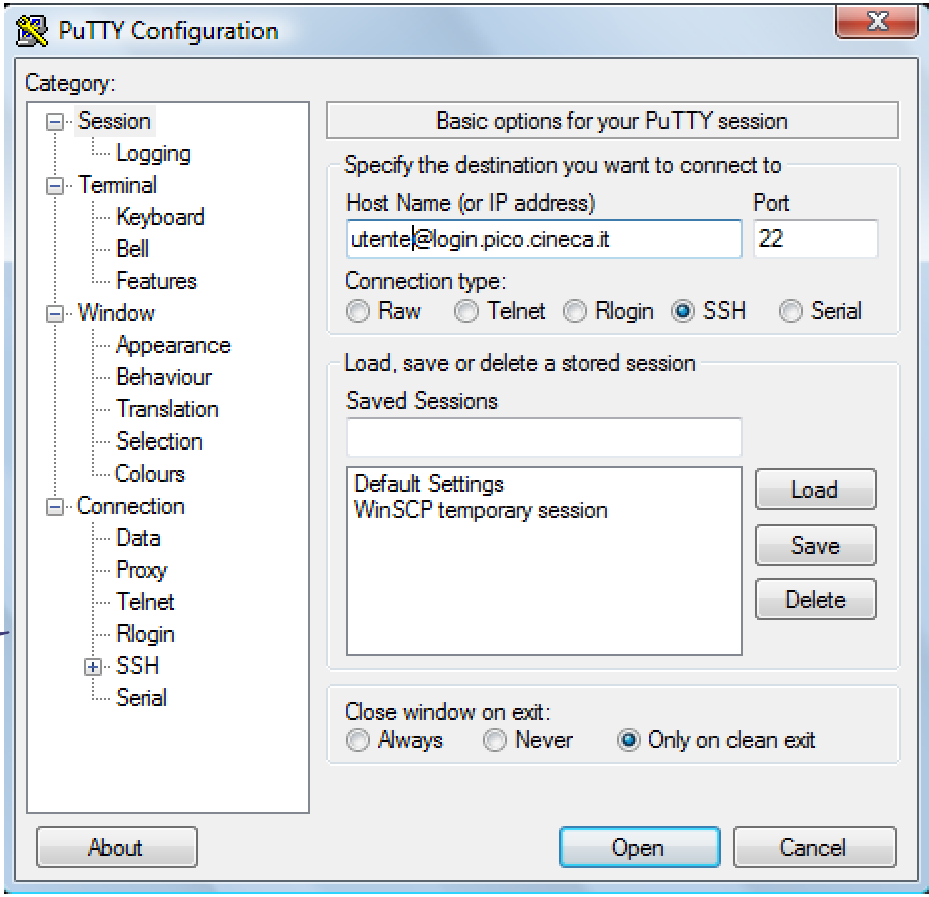
- **scp, ssh for linux users**

(ssh username@login.pico.cineca.it**)**

- **putty, winscp, TECTIA**

  for windows users

Example of connection on the front-end *PICO* through putty application



PuTTY Configuration

Basic options for your PuTTY session

Specify the destination you want to connect to

Host Name (or IP address): utente@login.pico.cineca.it

Port: 22

Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Load, save or delete a stored session

Saved Sessions

Default Settings
WinSCP temporary session

Close window on exit: ○ Always ○ Never ● Only on clean exit

Storage and file system

$HOME (librerie e eseguibili personalizzati dell'utente):

- Permanent, backed-up, and local.
- Quota = 5GB.
- For source code or important input files.

$CINECA_SCRATCH (area indicata per la prototipazione e controllo di validità dell'eseguibile):

- Large, parallel filesystem (GPFS).
- Temporary (files older than 30 days automatically deleted), no backup.
- No quota max. A cleaning procedure for files older than 30 days

$WORK (area per lo storage dei dati/risultati ai fini del progetto):

- Permanent, backed-up, project specific, 1 Tb quota by default.

Acconting: saldo

Accounting philosophy is based on the resources requested for the time of the batch job:

cost = no. of cores **requested** x job duration

In the CINECA system it is possible to have more than 1 budget (“account”) from which you can use time. The accounts available to your UNIX username can be found from the `saldo` command.

```
[mcestari@node342]$ saldo -b

--------------------------------------------------------------------------------------------------------
account                start          end        total        localCluster  totConsumed      totConsumed
                                             (local h)  Consumed(local h)    (local h)                %
--------------------------------------------------------------------------------------------------------
try11_test          20110301     20111201        10000                   0            2              0.0
cin_staff           20110323     20200323    200000000               64581      6689593              3.3
ArpaP_prod          20130130     20131101      1500000                   0            0              0.0
```

Modules

- CINECA’ s work environment is organized in modules, a set of installed libs, tools and applications available for all users.

- “loading” a module means that a series of (useful) shell environment variables wil be set

- E.g. after a module is loaded, an environment variable of the form “<MODULENAME>_HOME” is set

Module Environment

Bioinformatics applications, public databases and annotations are pre-installed on *PICO* cluster using the "*module*" environment.

***"module" environment* allows the user, by using a single command, to:**

- **list all the installed programs**
- **list all the genomes, indexes, and annotation databases**
- **get all the configured path (set environmental variables)**
- **automatic load the program in any directory**
- **launch the program**

Module environment: usage

Command to initialize the module environment

```
$ module load profile/advanced
```

Command to list the installed modules

```
$ module available
```

Command to load a module program

```
$ module load autoload name_program
```

Module commands

> module available (or just “> module av”)

Shows the full list of the modules available in the profile you’re into, divided by: environment, libraries, compilers, tools, applications

> module (un)load <module_name>

(Un)loads a specific module

> module show <module_name>

Shows the environment variables set by a specific module

> module help <module_name>

Gets all informations about how to use a specific module

> module purge

Gets rid of all the loaded modules

Module environment: usage

Command example to list available modules in «profile bio»

```
$ module available

------------ /cineca/prod/modulefiles/base/biodata ------------

D_melanogaster/dm3      Mus_musculus/mm9        Z_mays/ZmB73
Homo_Sapiens/hg18       R_norvegicus/rn4        Z_mays/maize3
Homo_Sapiens/hg19       S_cerevisiae/sacCer3    Mus_musculus/mm10
Z_mays/Mo17_v1(default)

------------ /cineca/prod/modulefiles/base/applications ------------

annovar/2014Sep15      cufflinks/2.2.1        snpeff/4.1b
bedtools/2.21.0        fastqc/0.11.2          star/2.4.0d
bowtie/1.0.1           idl/8.1                tophat/2.0.11(default)
bowtie2/2.2.3          picard/1.119           tophat/2.0.12
bwa/0.7.10             samtools/0.1.19        vcftools/0.1.12b
chimerascan/0.4.5a     samtools/1.1
```

Module commands

```
> module available (or just "> module av")
```

Examples

```
------------ /cineca/prod/modulefiles/advanced/applications ------------

bioconductor/2.14    gmql/2.2         peakranger/1.17  tabix/0.2.6
bioconductor/3.0     homer/4.7        picard/1.119     tophat/2.0.11
biopython/1.65       htseq/0.6.1      pintron/1.3.0    tophat/2.0.12
bismark/0.14.2       idl/8.1          qiime/1.9.0      treetagger/3.2
blast+/2.2.30        igvtools/2.3.40  r/3.1.2          trimmomatic/0.33
bowtie/1.0.1         lofreq/2.1.1     r/3.2.2          ucsc/1.0
bowtie2/2.2.3        macs/1.4.0       racket/6.2.1     varscan2/2.3.7
….
```

```
> module available bowtie*
------------ /cineca/prod/modulefiles/advanced/applications ------------

bowtie/1.0.1  bowtie2/2.2.3
```

Module commands

```
> module load bowtie2/2.2.3


> module list
  Currently Loaded Modulefiles:
  1) profile/advanced  2) bowtie2/2.2.3

> module show bowtie2/2.2.3
-------------------------------------------------------------------
/cineca/prod/modulefiles/advanced/applications/bowtie2/2.2.3:

module-whatis  Fast and sensitive read alignment
setenv  BOWTIE2_HOME   /cineca/prod/applications/bowtie2/2.2.3/binary
prepend-path    PATH    /cineca/prod/applications/bowtie2/2.2.3/binary/bin    :
-------------------------------------------------------------------
```

Module commands

```
> module help bowtie2/2.2.3
-----------------------------------------------------------------------
Module Specific Help for /cineca/prod/modulefiles/advanced/applications/bowtie2/2.2.3:

modulefile "bowtie2/2.2.3"

bowtie2-2.2.3
Fast and sensitive read alignment

-------------------------------------------------------------------------------------
License type: gpl
Web site:     http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Download url: http://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.2.3/
-------------------------------------------------------------------------------------

Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference
sequences. It is particularly good at aligning reads of about 50 up to 100s or 1,000s of characters, and
particularly good at aligning to relatively long (e.g. mammalian) genomes. Bowtie 2 indexes the
genome with an FM Index to keep its memory footprint small: for the human genome, its memory
footprint is typically around 3.2 GB. Bowtie 2 supports gapped, local, and paired-end alignment modes.


-----------------------------------------------------------------------
```

Module commands: dependencies

```
> module load biopython/1.65
    WARNING: biopython/1.65 cannot be loaded due to missing prereq.
     HINT: the following modules must be loaded first: python/2.7.8
```

- What happens?

```
> module show biopython /1.65
-------------------------------------------------------------------
/cineca/prod/modulefiles/advanced/applications/biopython/1.65:

module-whatis Biopython is a set of freely available tools for biological computation
written in Python by an international team of developers.
prereq      python/2.7.8
setenv      BIOPYTHON_HOME      /cineca/prod/applications/biopython/1.65/gnu--4.8.3
prepend-path   PYTHONPATH       /cineca/prod/applications/biopython/1.65/gnu--
4.8.3/lib/python2.7/site-packages   :
-------------------------------------------------------------------
```

Module commands:dependencies

```
> module load autoload biopython/1.65


> module list
  Currently Loaded Modulefiles:
  1) profile/advanced      3) gnu/4.8.3       5) biopython/1.65
  2) autoload/0.1        4) python/2.7.8


> module show python/2.7.8
-------------------------------------------------------------------
/cineca/prod/modulefiles/advanced/tools/python/2.7.8:

module-whatis python language
prereq     gnu/4.8.3
conflict   python
setenv     PYTHON_HOME    /cineca/prod/tools/python/2.7.8/gnu--4.8.3
prepend-path   PYTHONPATH    /cineca/prod/tools/python/2.7.8/gnu--4.8.3/lib/python2.7/site-packag
prepend-path   PATH      /cineca/prod/tools/python/2.7.8/gnu--4.8.3/bin        :
prepend-path   LD_LIBRARY_PATH      /cineca/prod/tools/python/2.7.8/gnu--4.8.3/lib :
-------------------------------------------------------------------
```

Module environment: load biodata

Command example to load available data and indexes

```
$ module load Homo_Sapiens/hg19

several environment variables are defined:

$ module show Homo_Sapiens/hg19
-------------------------------------------------------------------
/cineca/prod/modulefiles/base/biodata/Homo_Sapiens/hg19:
module-whatis   Human Sapiens genome hg19
setenv  GENOME  /cineca/prod/biodata/Homo_Sapiens/hg19/
setenv  ANNOT   /cineca/prod/biodata/Homo_Sapiens/hg19/annotation
setenv  GFASTA  /cineca/prod/biodata/Homo_Sapiens/hg19/genome
setenv  GINDEX  /cineca/prod/biodata/Homo_Sapiens/hg19/indexes
setenv  BWINDEX /cineca/prod/biodata/Homo_Sapiens/hg19/indexes/bowtie-1.0.1
setenv  BW2INDEX /cineca/prod/biodata/Homo_Sapiens/hg19/indexes/bowtie2-2.2.3
-------------------------------------------------------------------


that point to raw or indexed genomic data
```

Module environment: usage

Command example to launch a program using environmental variables

```
1)  Command example to launch bowtie (using bowtie2 index)


$ module load autoload bowtie2



$  bowtie2 $BW2INDEX/name_index  -un output.unmapped.fastq --chunkmbs 128 -p 8
-k 1 --best -S input.sam --phred64-quals
```

Lauching a job

- Now that we have our executable, it’s time to learn how to prepare a job for its execution

- Pico has the **PBS** scheduler.

- The job script scheme is:

```
- #!/bin/bash
- #PBS keywords
- variables environment
- execution line
```

Environment setup and execution line

The execution line starts with *./myexe arg_1 arg_2:*

```
./myexe arg_1 arg_2
```

`arg_1 arg_2` are the normal arguments of myexe

The environment setting usually starts with “**cd $PBS_O_WORKDIR**”.

That’s because by default you are launching on your home space the executable may not be found.

$PBS_O_WORKDIR points to the directory from where you’re submitting the job

.

PBS keywords

```
#PBS -N jobname                                              # name of the job
#PBS -o job.out                                              # output file
#PBS -e job.err                                              # error file
#PBS -l select=1:ncpus=20:mpiprocs=20:mem=122GB              # resources
#PBS -l walltime=1:00:00                                     # hh:mm:ss
#PBS -q <queue>                                              # chosen queue
#PBS -A <my_account>                                         # name of the account
#PBS -W group_list=<group>                                   # name of effective group
                                                              for reservation
```

select = number of node requested

ncpus = number of cpus per node requested

mpiprocs = number of mpi tasks per node

mem = RAM memory per node

Example of inteactive submisison of fastqc/0.11.3 on PICO

```
username@node013.pico:[~]$
qsub -I -l select=1:ncpus=2:mpiprocs=1:mem=8GB -l walltime=5:00:00 -A
train_RNAseq15 -W group_list=train_RNAseq15 -q R121546
qsub: waiting for job 123456.node001 to start
qsub: job 123456.node001 ready
```

select = number of nodes requested

ncpus = number of cpus per node requested

mpiprocs = number of MPI tasks per node

mem = RAM memory per node

walltime = wall time limit

parallel = name of queue for parallel job (multithread too)

train... = account namec

Example of inteactive submisison of fastqc/0.11.3 on PICO

```
username@node013.pico:[~]$
qsub -I -l select=1:ncpus=2:mpiprocs=1:mem=8GB -l walltime=5:00:00 -A
train_RNAseq15 -W group_list=train_RNAseq15 -q R121546
qsub: waiting for job 123456.node001 to start
qsub: job 123456.node001 ready

username@node009.pico:[~]$ module load profile/advanced

username@node009.pico:[~]$ module load fastqc/0.11.3

username@node009.pico:[~]$ fastqc --nogroup -t 2 --extract input.R1 input.R2 -o
output 2>&1 | tee input.log
```

Example of batch-script to launch fastqc/0.11.3 on PICO

```
#!/bin/bash
#PBS -N fastqc
#PBS -l select=1:ncpus=2:mpiprocs=1:mem=8GB
#PBS -q R121546
#PBS -l walltime=5:00:00
#PBS -A train_RNAseq15
#PBS -W group_list=train_RNAseq15

cd $PBS_O_WORKDIR                    ==> change to current dir

module load profile/advanced
module load fastqc/0.11.3

fastqc --nogroup -t 2 --extract input.R1 input.R2 -o output 2>&1 | tee input.log
```

Example of batch-script to launch fastqc/0.11.3 on PICO

```
username@node013.pico:[~] qsub launch_fastqc.sh

    123456.node001
```

Example of batch-script to launch fastqc/0.11.3 on PICO

```
#!/bin/bash
#PBS -N fastqc
#PBS -l select=1:ncpus=2:mpiprocs=1:mem=8GB
#PBS -q R121546
#PBS -l walltime=5:00:00
#PBS -A train_RNAseq15
#PBS -W group_list=train_RNAseq15

INPUT_HOME="/pico/home/userinternal/tcastign/test/input"
OUTPUT_HOME="/pico/home/userinternal/tcastign/test/output"
OUTPUT_FASTQC="/pico/home/userinternal/tcastign/test/output/fastqc“

echo $INPUT_HOME;
echo $OUTPUT_HOME;
echo $OUTPUT_FASTQC;


………..

 fastqc --nogroup -t 2 --extract $INPUT_HOME/$fastq -o $OUTPUT_FASTQC 2>&1 |tee input.log
```

What does CINECA offer for bioinformatics?

1. Computing resources

Bioinformatics software available through command line

PROS	CONS
Rich environment: bioinformatics resources continuosly updated	Basic Unix/Linux Knowledge needed
Flexible environment: Resources can be added under request depending on user needs	
Simple usage through «module» environment	

What does CINECA offer for bioinformatics?

1. Computing resources

   Bioinformatics software available through command line

2. **Advanced services**

   Automated web workflows for Next Generation Sequencing

3. Bioinformatics Expertise

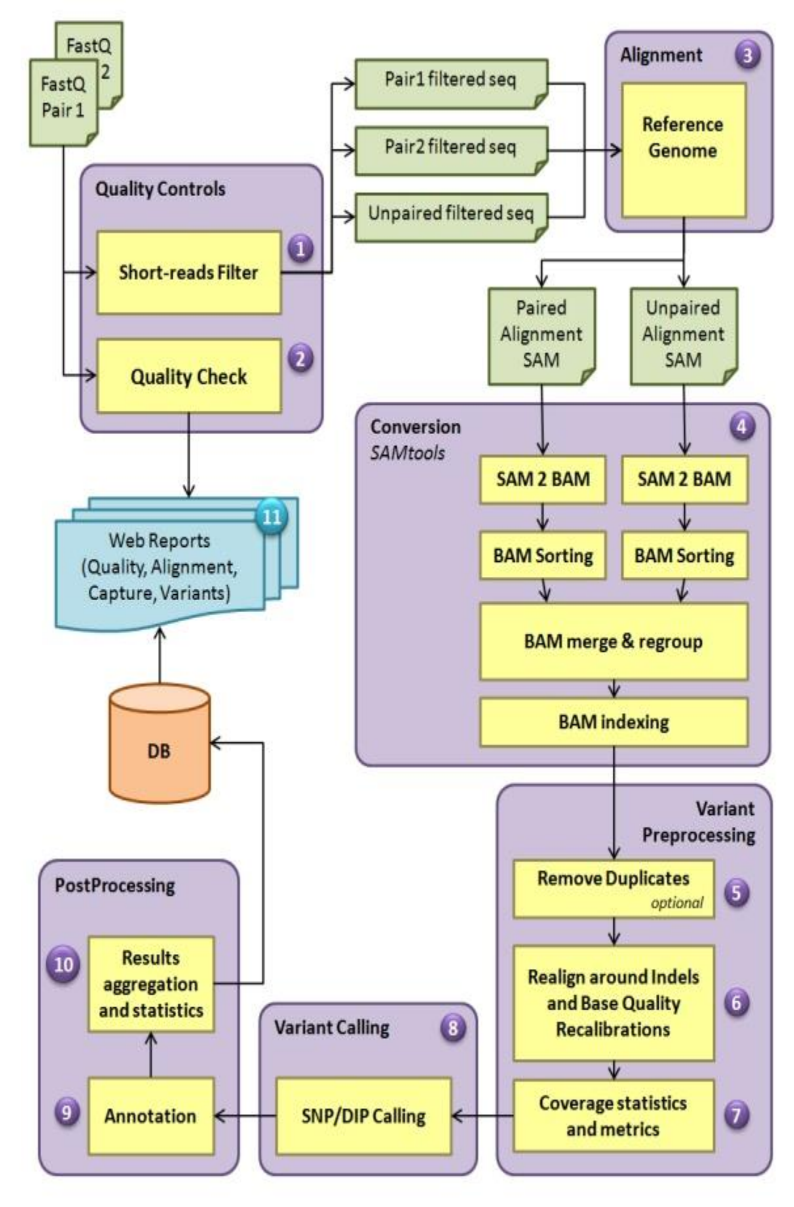
   To customize solutions or implement new systems and tools

Bioinformatics NGS Pipelines

Automated workflows (pipelines) for Next Generation Sequencing are available through a web interface and are able to perform analyses for several NGS application fields:

- Deep targeted exome sequencing;
- RNA sequencing (trascriptome analysis);
- Whole exome sequencing;
- Identification of DNA protein interactions by ChIP-seq;

Ultra Deep Exome Sequencing Pipeline



FastQ Pair 1

FastQ 2

Quality Controls

Short-reads Filter 1

Quality Check 2

Pair1 filtered seq

Pair2 filtered seq

Unpaired filtered seq

Alignment 3

Reference Genome

Paired Alignment SAM

Unpaired Alignment SAM

Conversion *SAMtools* 4

SAM 2 BAM

SAM 2 BAM

BAM Sorting

BAM Sorting

BAM merge & regroup

BAM indexing

Web Reports (Quality, Alignment, Capture, Variants) 11

DB

Variant Preprocessing

Remove Duplicates *optional* 5

Realign around Indels and Base Quality Recalibrations 6

Coverage statistics and metrics 7

Variant Calling 8

SNP/DIP Calling

PostProcessing

Results aggregation and statistics 10

Annotation 9

Online Deep Exome Sequencing Software Analysis (ODESSA)

Handles genes targeted at high coverage

Specifically focused for clinical diagnostics

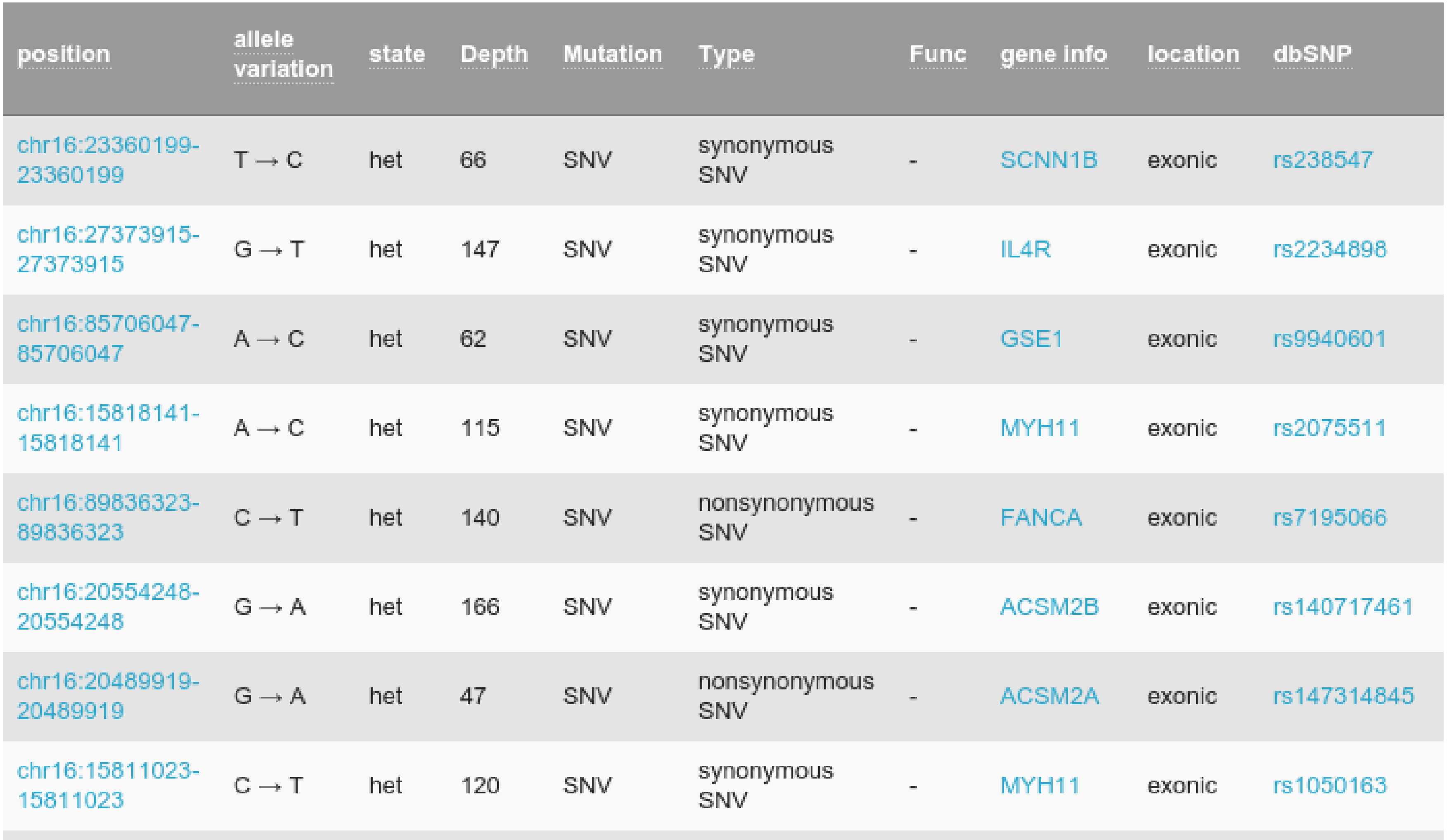
Identifies (SNPs) and (DIPs) classified by different scores (e.g. depth, SIFT, MAV, MEQ).

Results are supported with genomic information, functional annotations, cross-linking databases and quality and relevance scores, graphics, tables and browsing, filtering and download.

Optimized for MiSeq Illumina platform

Ultra Deep Exome Sequencing Pipeline

Example of output: variant results

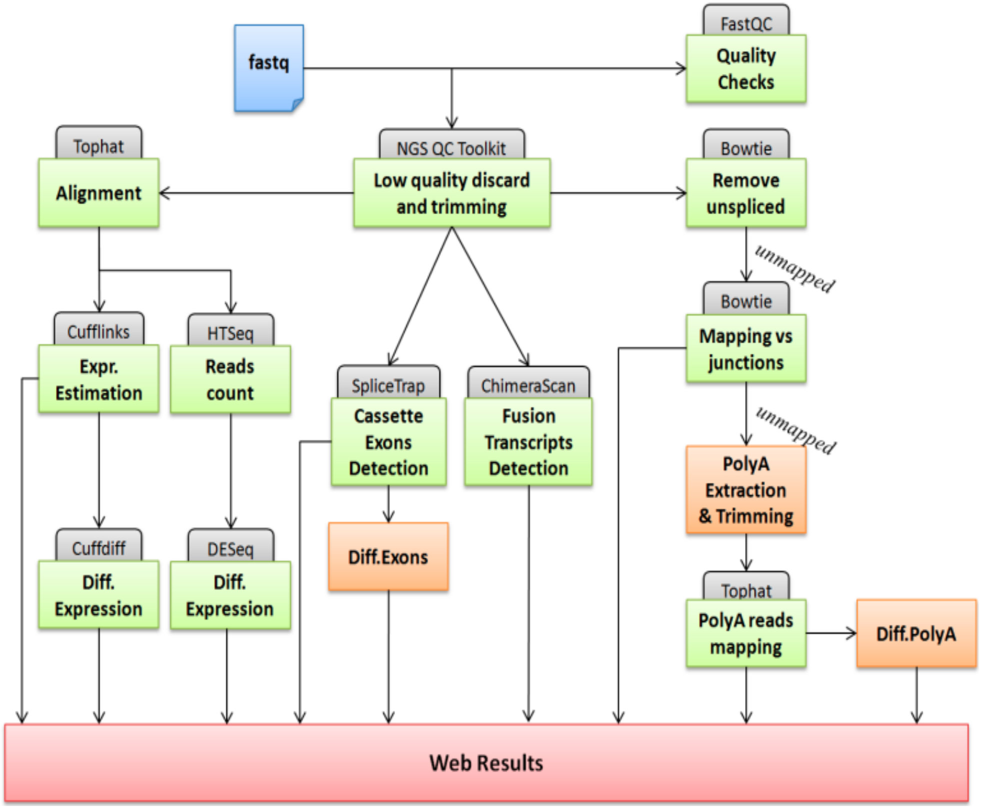


position	allele variation	state	Depth	Mutation	Type	Func	gene info	location	dbSNP
chr16:23360199-23360199	T → C	het	66	SNV	synonymous SNV	-	SCNN1B	exonic	rs238547
chr16:27373915-27373915	G → T	het	147	SNV	synonymous SNV	-	IL4R	exonic	rs2234898
chr16:85706047-85706047	A → C	het	62	SNV	synonymous SNV	-	GSE1	exonic	rs9940601
chr16:15818141-15818141	A → C	het	115	SNV	synonymous SNV	-	MYH11	exonic	rs2075511
chr16:89836323-89836323	C → T	het	140	SNV	nonsynonymous SNV	-	FANCA	exonic	rs7195066
chr16:20554248-20554248	G → A	het	166	SNV	synonymous SNV	-	ACSM2B	exonic	rs140717461
chr16:20489919-20489919	G → A	het	47	SNV	nonsynonymous SNV	-	ACSM2A	exonic	rs147314845
chr16:15811023-15811023	C → T	het	120	SNV	synonymous SNV	-	MYH11	exonic	rs1050163

RNA Sequencing Pipeline

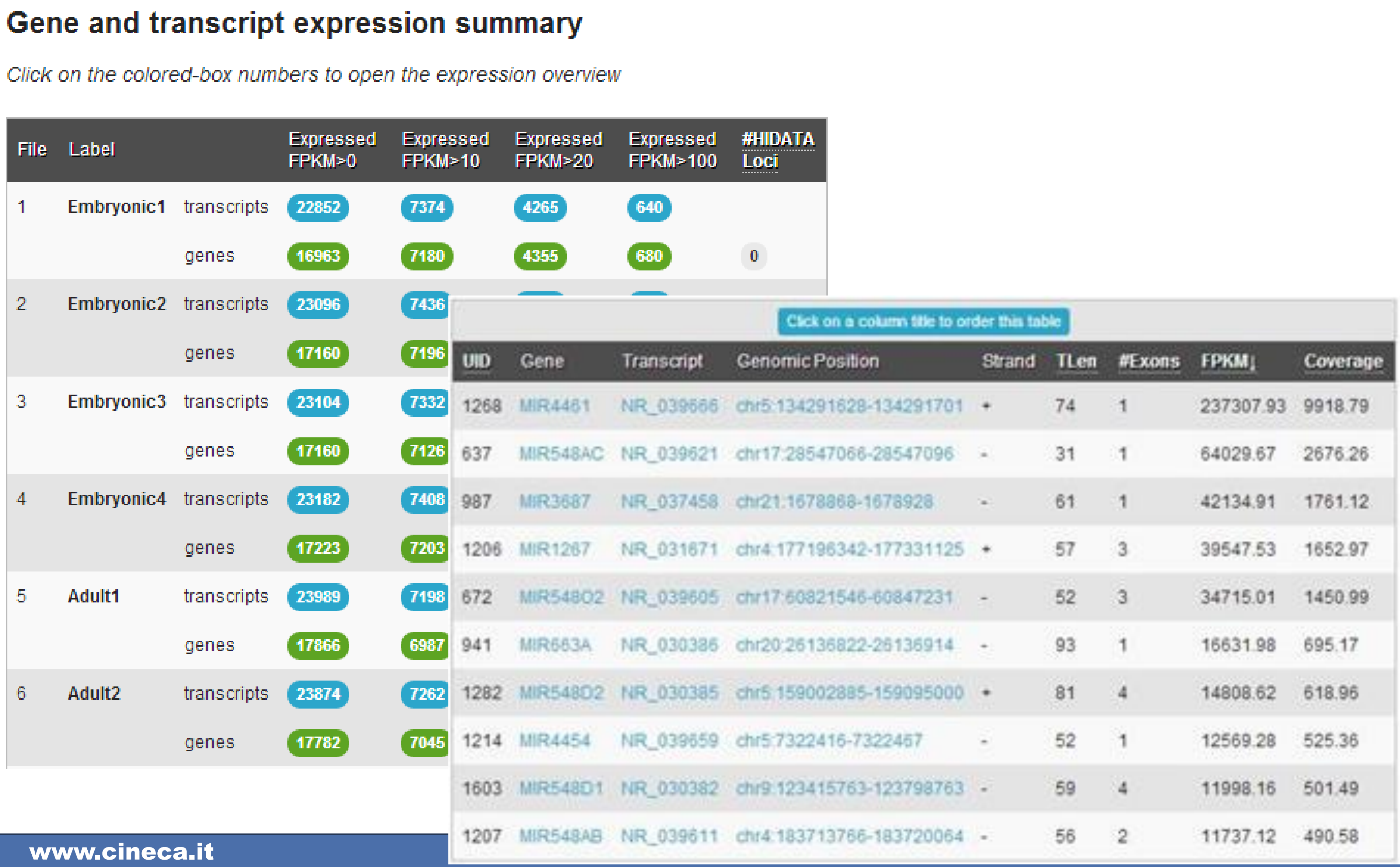
The RNA-Seq Analysis Pipeline (RAP)

Performs a complete and customizable RNA-seq pipeline, allowing users to examine NGS data under many points of view:



- Gene and transcript expression
- Differential expression
- Splicing junctions
- Cassette exons
- Poly(A) sites
- Fusion transcripts
- RNA editing

RNA Sequencing Pipeline



Gene and transcript expression summary

Click on the colored-box numbers to open the expression overview

File	Label		Expressed FPKM>0	Expressed FPKM>10	Expressed FPKM>20	Expressed FPKM>100	#HIDATA Loci
1	Embryonic1	transcripts	22852	7374	4265	640	
		genes	16963	7180	4355	680	0
2	Embryonic2	transcripts	23096	7436			
		genes	17160	7196			
3	Embryonic3	transcripts	23104	7332			
		genes	17160	7126			
4	Embryonic4	transcripts	23182	7408			
		genes	17223	7203			
5	Adult1	transcripts	23989	7198			
		genes	17866	6987			
6	Adult2	transcripts	23874	7262			
		genes	17782	7045			

UID	Gene	Transcript	Genomic Position	Strand	TLen	#Exons	FPKM↓	Coverage
1268	MIR4461	NR_039666	chr5:134291628-134291701	+	74	1	237307.93	9918.79
637	MIR548AC	NR_039621	chr17:28547066-28547096	-	31	1	64029.67	2676.26
987	MIR3687	NR_037458	chr21:1678868-1678928	-	61	1	42134.91	1761.12
1206	MIR1267	NR_031671	chr4:177196342-177331125	+	57	3	39547.53	1652.97
672	MIR548O2	NR_039605	chr17:60821546-60847231	-	52	3	34715.01	1450.99
941	MIR663A	NR_030386	chr20:26136822-26136914	-	93	1	16631.98	695.17
1282	MIR548D2	NR_030385	chr5:159002885-159095000	+	81	4	14808.62	618.96
1214	MIR4454	NR_039659	chr5:7322416-7322467	-	52	1	12569.28	525.36
1603	MIR548D1	NR_030382	chr9:123415763-123798763	-	59	4	11998.16	501.49
1207	MIR548AB	NR_039611	chr4:183713766-183720064	-	56	2	11737.12	490.58

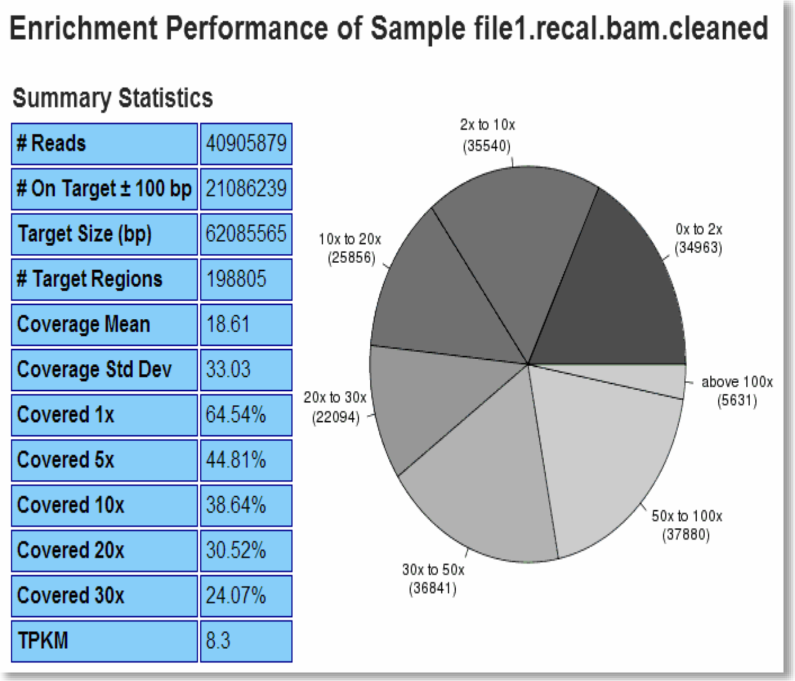
Whole-exome Sequencing Pipeline

Whole-Exome sequencing Pipeline (WEP)

SNP and DIP detection and annotation

gapped alignment, duplicates removal, quality scores recalibration

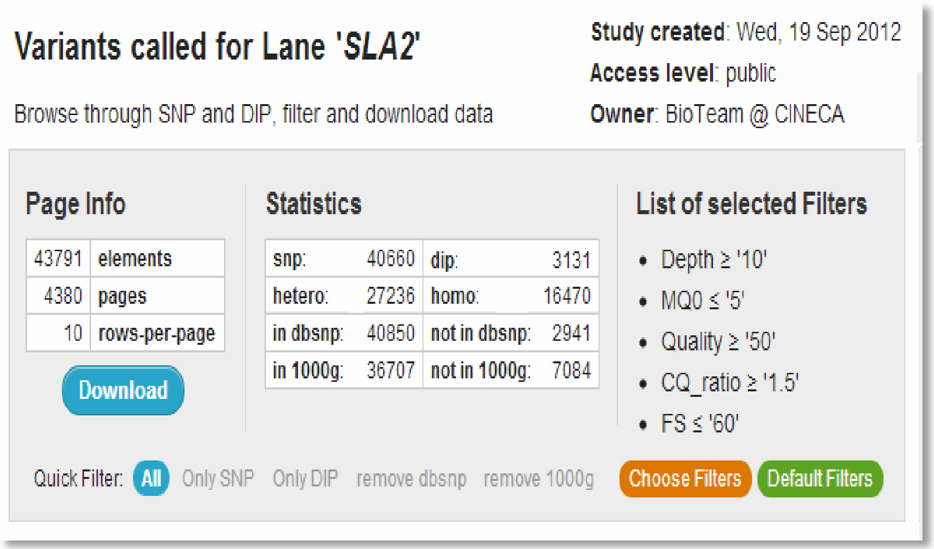
cross-linking, intersections, trio analyses, statistics



Enrichment Performance of Sample file1.recal.bam.cleaned

Summary Statistics

# Reads	40905879
# On Target ± 100 bp	21086239
Target Size (bp)	62085565
# Target Regions	198805
Coverage Mean	18.61
Coverage Std Dev	33.03
Covered 1x	64.54%
Covered 5x	44.81%
Covered 10x	38.64%
Covered 20x	30.52%
Covered 30x	24.07%
TPKM	8.3



Variants called for Lane '*SLA2*'

Browse through SNP and DIP, filter and download data

Study created: Wed, 19 Sep 2012
Access level: public
Owner: BioTeam @ CINECA

Page Info

43791	elements
4380	pages
10	rows-per-page

Statistics

snp:	40660	dip:	3131
hetero:	27236	homo:	16470
in dbsnp:	40850	not in dbsnp:	2941
in 1000g:	36707	not in 1000g:	7084

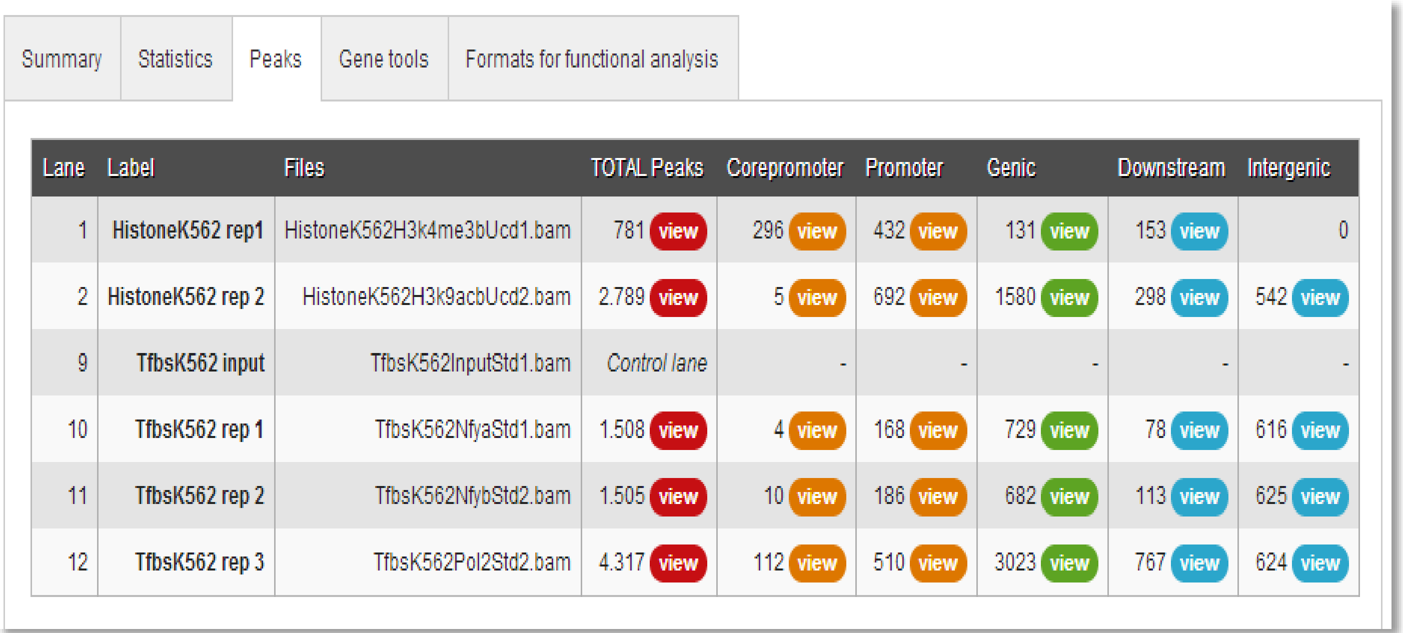
List of selected Filters

- Depth ≥ '10'
- MQ0 ≤ '5'
- Quality ≥ '50'
- CQ_ratio ≥ '1.5'
- FS ≤ '60'

ChiP-seq Pipeline

Chip-seq analysis pipeline (**CAST**)

- peaks detection
- peaks filtering
- peaks visualization on UCSC Genome Browser
- peaks annotation with genomic features



Lane	Label	Files	TOTAL Peaks	Corepromoter	Promoter	Genic	Downstream	Intergenic
1	**HistoneK562 rep1**	HistoneK562H3k4me3bUcd1.bam	781 view	296 view	432 view	131 view	153 view	0
2	**HistoneK562 rep 2**	HistoneK562H3k9acbUcd2.bam	2.789 view	5 view	692 view	1580 view	298 view	542 view
9	**TfbsK562 input**	TfbsK562InputStd1.bam	*Control lane*	-	-	-	-	-
10	**TfbsK562 rep 1**	TfbsK562NfyaStd1.bam	1.508 view	4 view	168 view	729 view	78 view	616 view
11	**TfbsK562 rep 2**	TfbsK562NfybStd2.bam	1.505 view	10 view	186 view	682 view	113 view	625 view
12	**TfbsK562 rep 3**	TfbsK562Pol2Std2.bam	4.317 view	112 view	510 view	3023 view	767 view	624 view

What does CINECA offer for bioinformatics?

2. Advanced services

Automated web workflows for Next Generation Sequencing

PROS	CONS
User-friendly graphic interface: The pipeline is completely automatized at each stage and doesn't require any computational knowledge by the user	**Low flexibility:** changes are allowed only with a specific project agreement with Cineca
Any knowledge of the underlying high-performance computing infrastructure is not needed by the user	
Automation avoids human errors introduced by hand-made scripts and also eases the processing of Big Data NGS experiments	

What does CINECA offer for bioinformatics?

1. Computing resources

   Bioinformatics software available through command line

2. Advanced services

   Automated web workflows for Next Generation Sequencing

3. **Bioinformatics Expertise**

   To customize solutions or implement new systems and tools

Bioinformatics specialistic support

Cineca offers bioinformatics specialistic support to develop and optimize

- configuration parameters
- command-line programs
- complex bash scripts

on hundreds of computing cores

For further information write to:
hpc-bioinformatics@cineca.it

Links

General Information

Official web site	http://www.hpc.cineca.it
Bio & Genomics	http://www.hpc.cineca.it/content/hpc-bioinformatics

How to get computational resources?

ISCRA initiative	http://www.hpc.cineca.it/services/iscra
PRACE:	http://www.prace-ri.eu/

Automated analysis workflows

Target Exome	https://bioinformatics.cineca.it/odessa
RNA-Seq	https://bioinformatics.cineca.it/rap
Whole Exome	https://bioinformatics.cineca.it/wep
ChIP-Seq	https://bioinformatics.cineca.it/cast