CLC bio's RNA-Seq Analysis Pipeline

Anne Arens (aarens@clcbio.com)

X International Course of Massive Data Analysis

CLC bio A QIAGEN® Company

Agenda

- Introduction
- RNA-Seq Analysis using CLC Genomics Workbench
- Hands on RNA-Seq Analysis
- Wrap up, Questions & Answers

About us

- QIAGEN Aarhus, Denmark
- Center of Excellence in Bioinformatics
- CLC bio started operating in 2005
- Global presents
  - 100+ employees worldwide
  - 2400+ customers

Desktop solutions

Classical

CLC Main Workbench 

Next-generation sequencing

CLC Genomics Workbench 

CLC Assembly Cell 

CLC Cancer Research Workbench 

Protein modeling

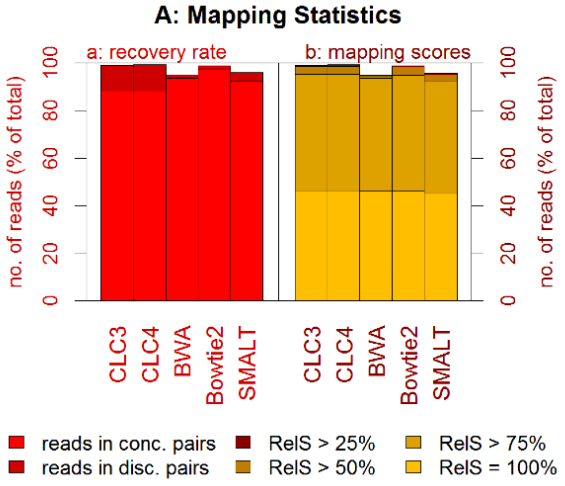
CLC Drug Discovery Workbench 

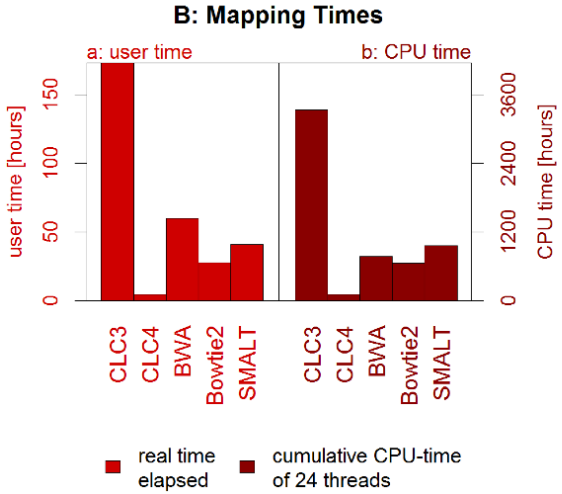
Read Mapper 4 Benchmark

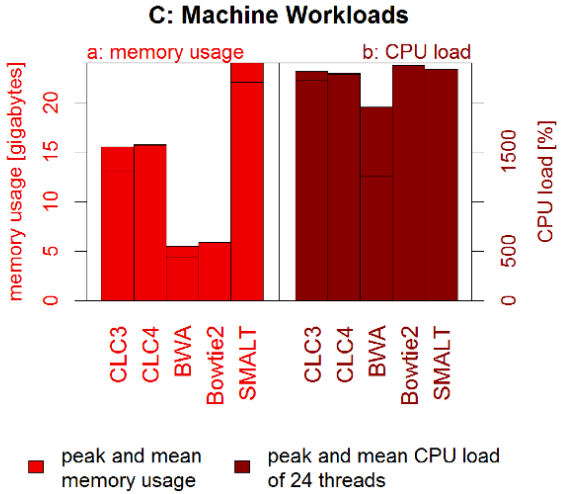
Platform	Reads	Protocol	Mean length
Illumina Genome Analyzer II	1,339,740,542	paired-end	101

Mapping hardware

Processors	2x Intel X5650 @ 2.66 GHz
Total physical cores	12
Total logical cores	24
Main memory	48 Gbyte







Mapper	unmapped
CLC3	0.81%
CLC4	0.66%
BWA	5.02%
Bowtie2	1.19%
SMALT	4.07%

user time	CPU time	mem. load	CPU load
173:07:19	3338:48:07	12.82	2,224
4:40:11	106:32:02	15.26	2,272
59:59:42	779:01:16	4.26	1,251
27:53:58	663:04:28	5.70	2,375
41:06:36	962:25:05	21.58	2,284

Genomics Server

Across Platforms

Microsoft®

illumina®

life technologies™

Roche

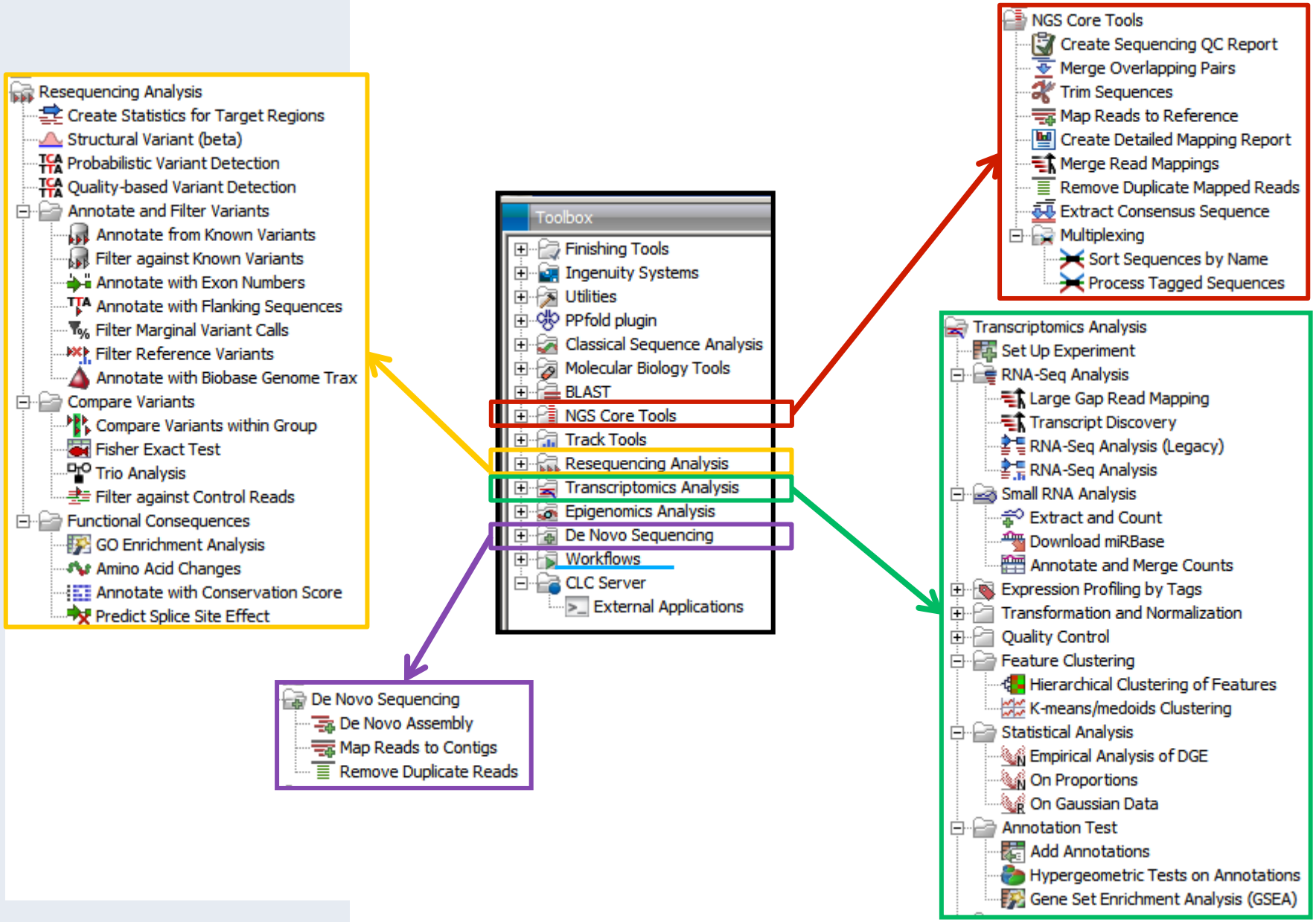
MiSeq

ion torrent

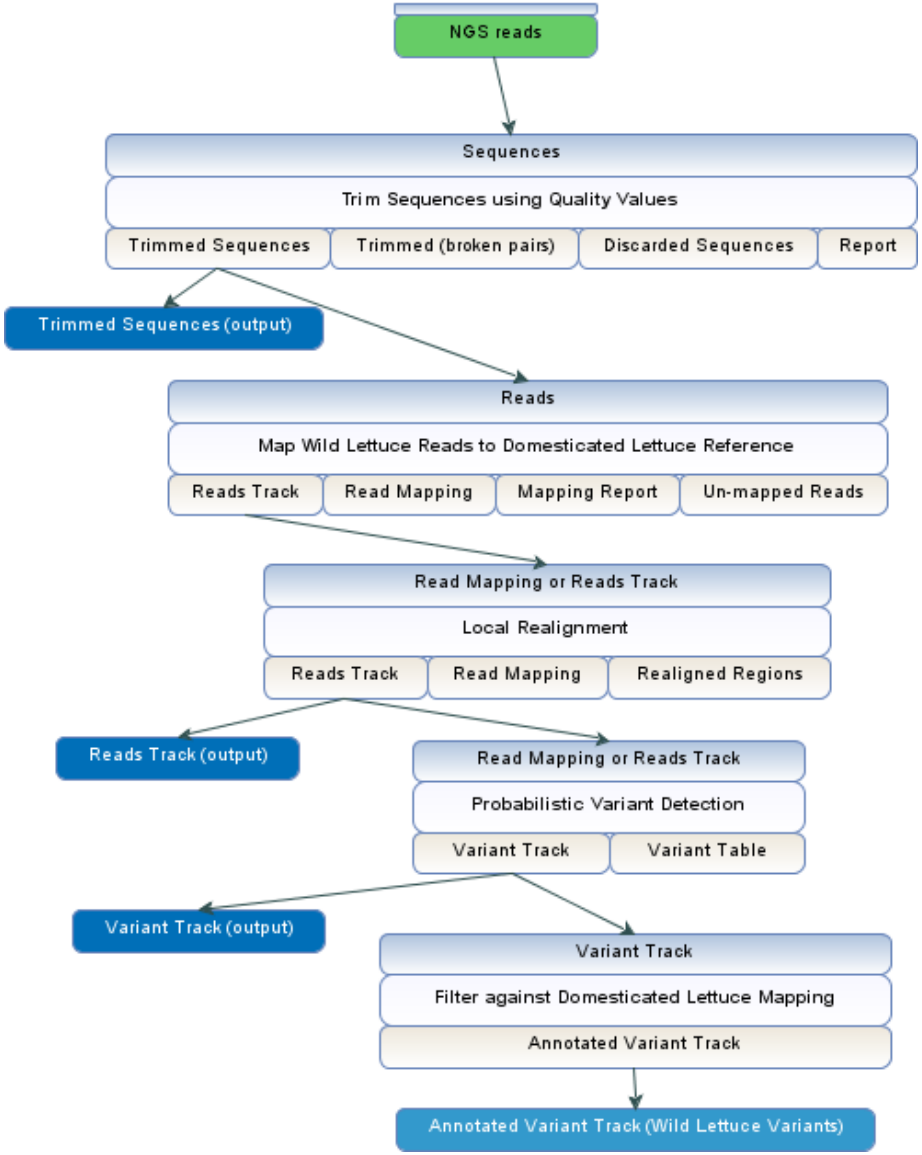
HiSeq

SOLiD™

Supported Applications and Tools



Workflows


NGS reads
Sequences
Trim Sequences using Quality Values
Trimmed Sequences
Trimmed (broken pairs)
Discarded Sequences
Report
Trimmed Sequences (output)
Reads
Map Wild Lettuce Reads to Domesticated Lettuce Reference
Reads Track
Read Mapping
Mapping Report
Un-mapped Reads
Read Mapping or Reads Track
Local Realignment
Reads Track
Read Mapping
Realigned Regions
Reads Track (output)
Read Mapping or Reads Track
Probabilistic Variant Detection
Variant Track
Variant Table
Variant Track (output)
Variant Track
Filter against Domesticated Lettuce Mapping
Annotated Variant Track
Annotated Variant Track (Wild Lettuce Variants)

External Plugins

Extending the Toolbox

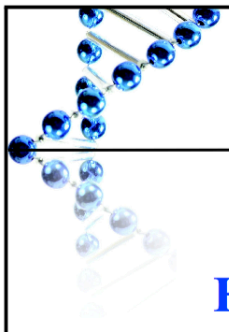
blast2go
CLC bio Workbench Plugin

Blast2GO® is a complete framework for functional annotation of (novel) sequences and the analysis of annotation data.

BIOBASE
BIOLOGICAL DATABASES

TRANSFAC®

Genome Trax™

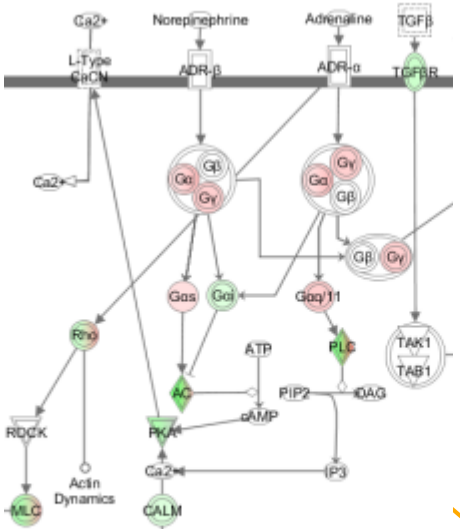


KLAST *for CLC Bio Workbench*

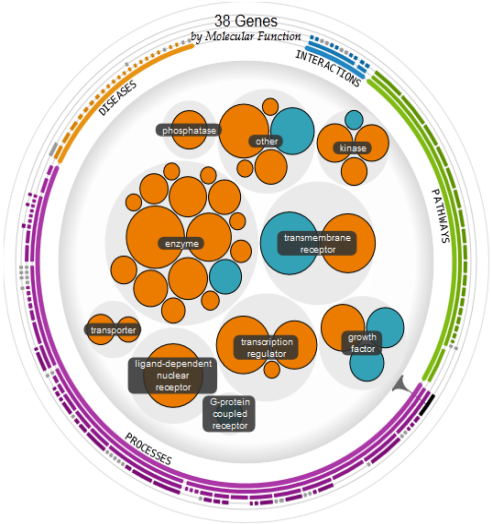
High-performance sequence similarity search tool

Ingenuity Solutions for NGS

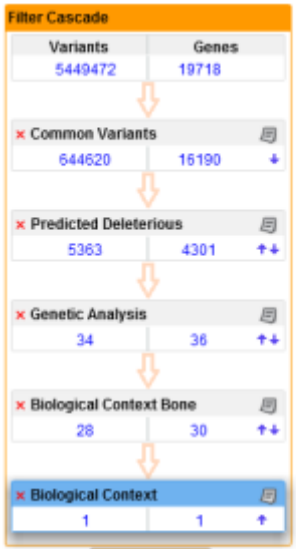
INGENUITY PATHWAY ANALYSIS

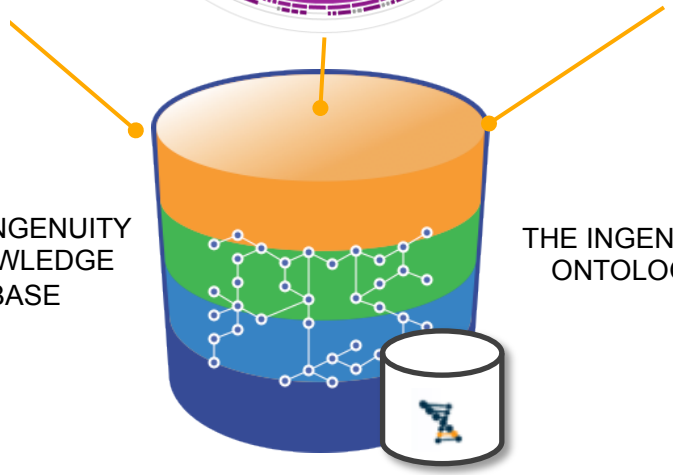


INGENUITY iREPORT



INGENUITY VARIANT ANALYSIS



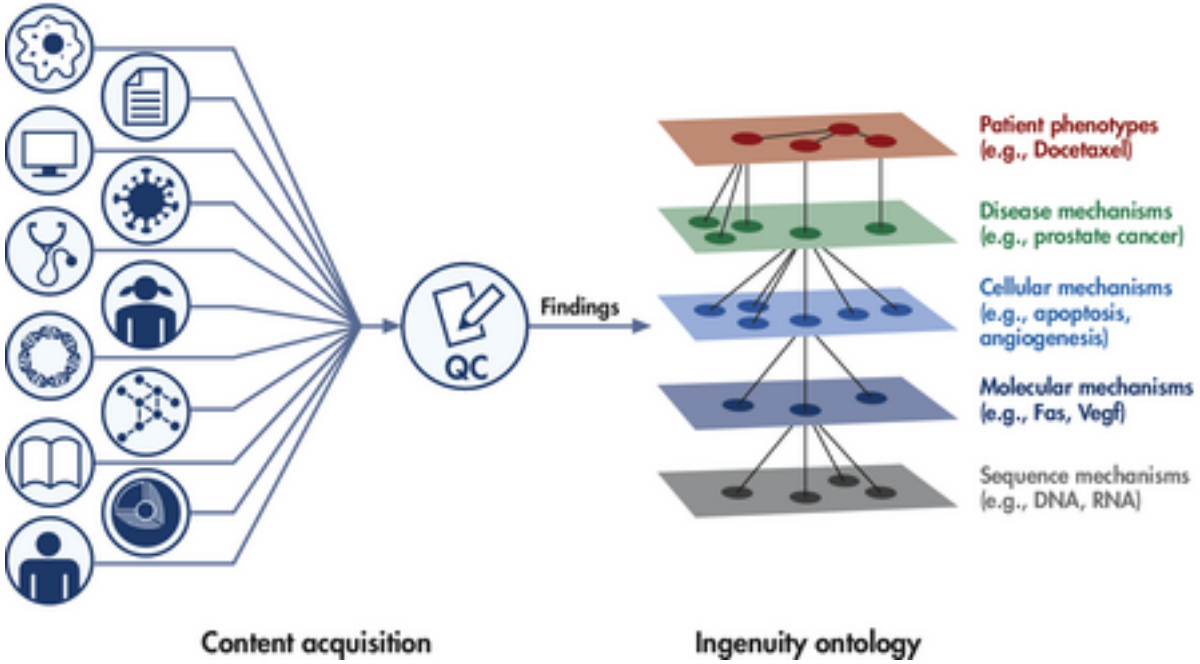


THE INGENUITY KNOWLEDGE BASE

THE INGENUITY ONTOLOGY

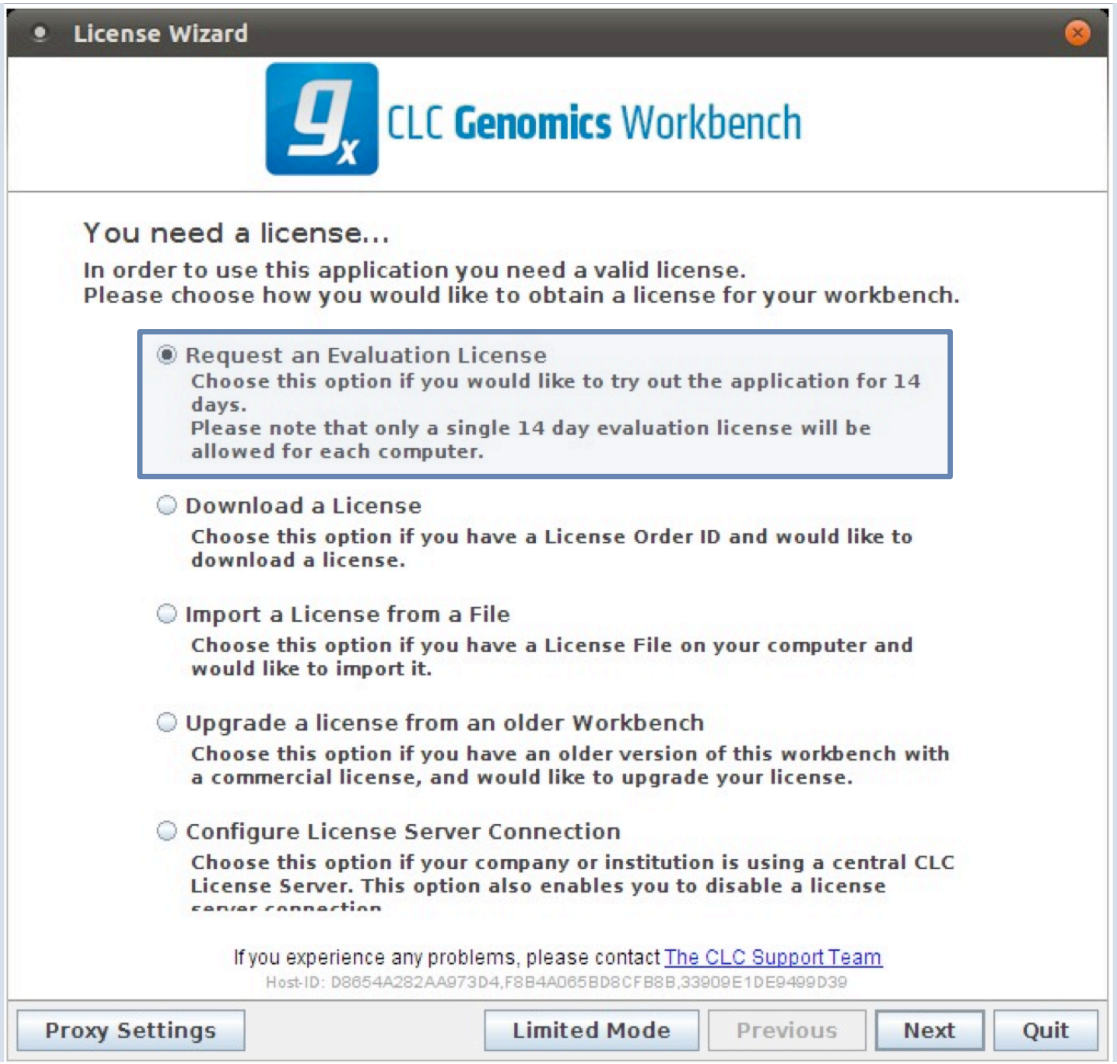
The Ingenuity® Knowledge Base

Rich, Structured, Accurate Content



- Provides contextual details such as:
  - Species specificity, cell type/tissue context, site and type of mutations, direction of change, post-translational modification sites, epigenetic modifications, and experimental methods
- Supports Computation
  - Ask questions across various types of connections
  - Make inferences from one concept to another, or find likely paths between molecular concepts (gene to disease, drug to gene, etc.).
- Provides Synonym Resolution

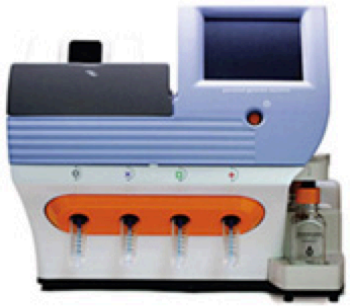
Evaluation License



Benchtop Sequencer Campaign

Benchtop Sequencers

90-day CLC Genomics Workbench trial license for benchtop sequencers



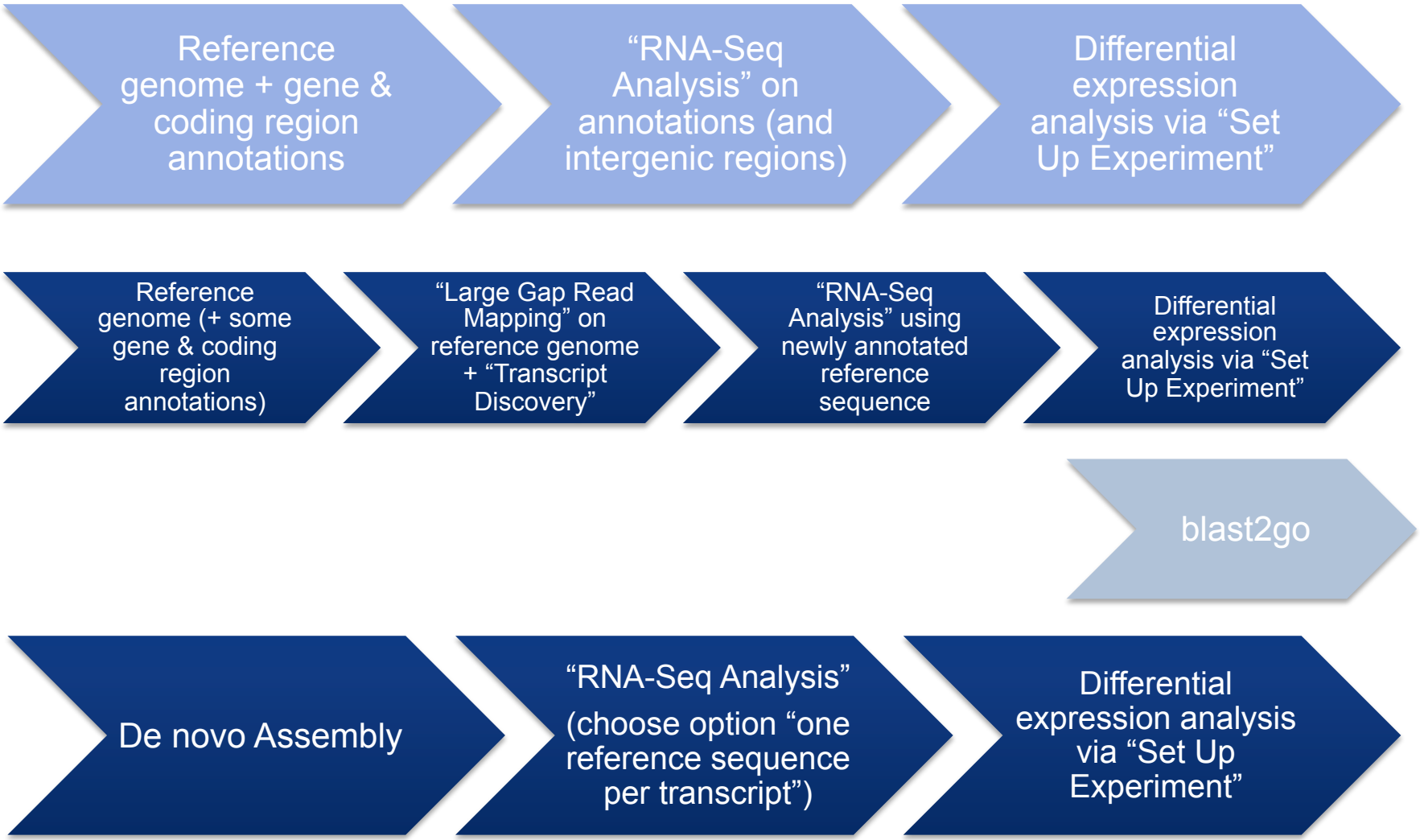

To help make sequencing analysis easier, CLC bio has decided to provide one fully functional 90-day trial license for CLC Genomics Workbench, including free customer support and all upgrades, to all labs or organizations that own a benchtop NGS sequencing instrument:

- Ion Torrent PGM
- Roche 454 GS Junior
- Illumina MiSeq
- Ion Proton

If your lab owns more than one instrument, we will provide you with multiple licenses — one license per instrument.

RNA-Seq Analysis Overview

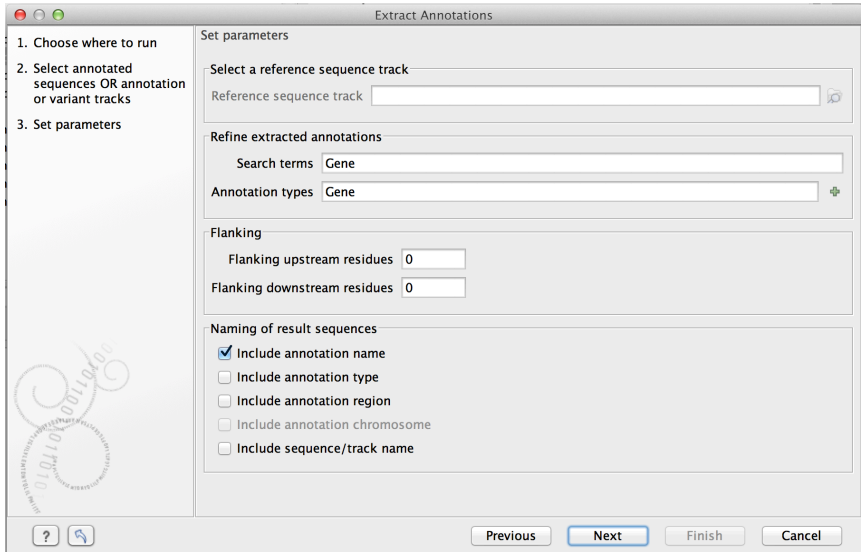
RNA-Seq reads +



Transcript Discovery + blast2go

Input files in TranscriptDiscovery folder

- Install plugin “Transcript discovery”
- Run Transcriptomics -> RNA-Seq Analysis -> Transcript discovery using “Large Gap Read Mapping”
- Extract new genes (Gene 1-5) from “annotated reference” using Classical Sequence Analysis -> General Sequence Analysis -> Extract Annotations:



- Blast extracted sequences (blastx)

- Install plugin “blast2GO Pro”
- Run Mapping, Annotation, Interpro
- …