

GWB CLC Genomics Workbench

다양한 NGS 데이터 분석을 내 손으로 직접!



GENOMICS

- De novo assembly of any genome size
- Read mapping for reference genome
- Variant detection
- Support NGS hybrid assembly
- Visualization of genome information

TRANSCRIPTOMICS

- RNA-seq analysis
- Small RNA analysis
- Expression profiling by tags
- Gene expression analysis

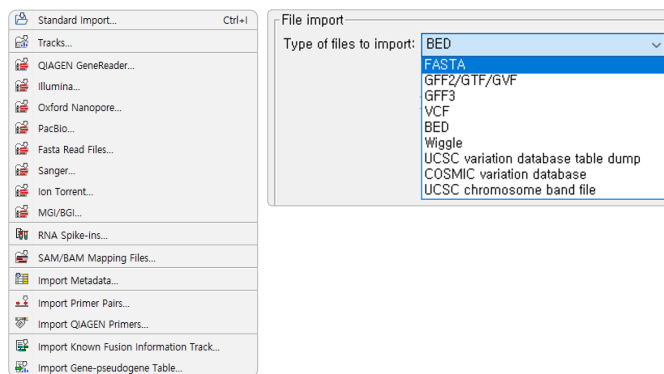
EPIGENOMICS

- ChIP-seq analysis
- Peak finding and peak refinement
- Case/control analysis
- Bisulfite sequencing analysis
- Histone ChIP-seq analysis

Support NGS multi platform

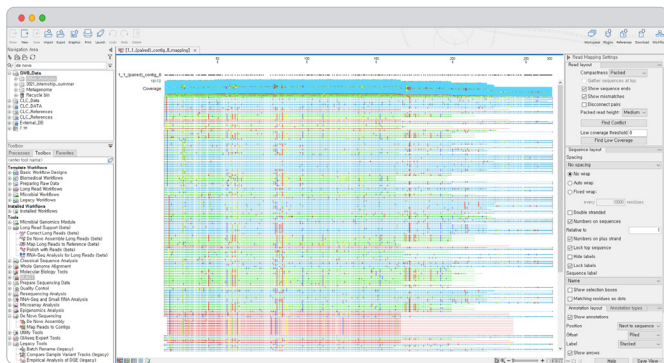
CLC Genomics Workbench는 GUI 기반의 NGS 데이터를 분석하기 위한 데스크톱 솔루션으로 기존 sanger 방식의 시퀀싱 데이터뿐만 아니라 Illumina, PacBio, Nanopore, Ion Torrent 등의 장비에서 생산되는 모든 시퀀싱 데이터의 분석을 지원합니다.

GFF3/GFF2/GTF/GVF 포맷의 annotation, BED, WIGGLE, VCF, UCSC chromosome band 파일 등을 import 할 수 있으며, txt or csv 포맷 등의 발현 정보 파일이나 annotation 파일도 분석에 사용할 수 있습니다.



De novo assembly

CLC Genomics Workbench는 NGS 멀티 포맷 데이터의 hybrid assembly를 지원하여 새로운 유전체 서열을 제작하기 위한 de novo assembly 분석을 수행합니다. 높은 퀄리티의 assembly 결과를 얻기 위한 word size 조절 및 유전체 크기에 제한 없는 분석을 수행할 수 있으며, mate paired end 시퀀싱 데이터를 이용한 scaffolding 기술을 지원하고 있습니다.



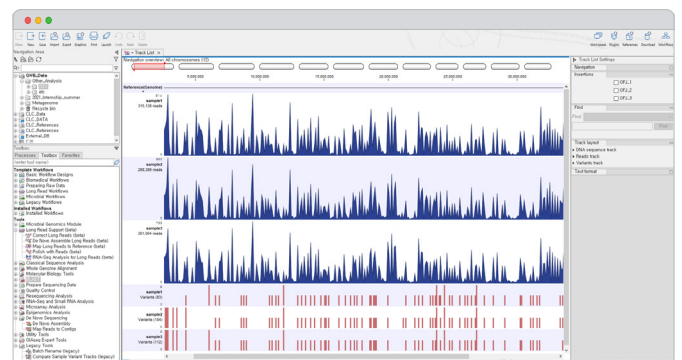
Reference assembly

Reference 서열에 시퀀싱 한 데이터를 맵핑함으로써 이후 variation 분석을 진행할 수 있는 기초 데이터를 제공합니다. 모델 organism의 유전체 데이터는 데이터베이스와 연동되어 쉽게 내려받아 분석에 활용할 수 있습니다.



Variation detection

Assembly에 기초한 SNP 및 small Indel 분석을 수행하며, 기본적으로 frequency를 이용한 알고리즘, low frequency에 적합한 알고리즘, germline variant를 분석하기에 적합한 fixed ploidy variant 분석 알고리즘 등 총 3종류의 variant detector가 탑재되어 확률적으로 수준 높은 potential variant 들을 찾을 수 있습니다. 또한 시퀀싱 중에 일어난 오류에 기인한 variation 등을 구별하기 위한 여러 가지 파라미터를 설정할 수 있고, 특정 SNP가 단백질 서열까지 변화되는 SNP인지 그 여부도 판단할 수 있습니다.



그리고 reference assembly에 기반하여 맵핑된 read 들의 통계 분석을 통해 다양한 structure variation 분석을 수행할 수 있습니다.

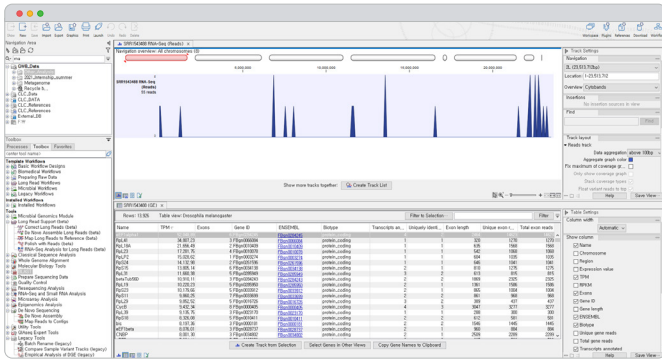
GWB CLC Genomics Workbench

다양한 NGS 데이터 분석을 내 손으로 직접!

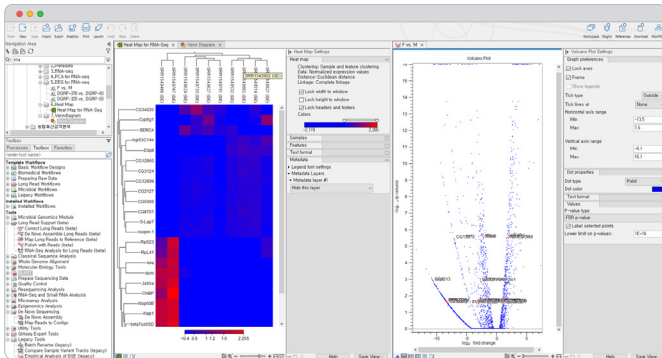


RNA-Seq analysis

CLC Genomics Workbench는 annotation 된 reference 유전체 서열과 mRNA 시퀀싱 서열들을 바탕으로 샘플 간의 유전자 발현 정도를 비교할 수 있습니다. 발현 값 계산의 기준은 read 개수, RPKM, TPM 중 선택할 수 있습니다.



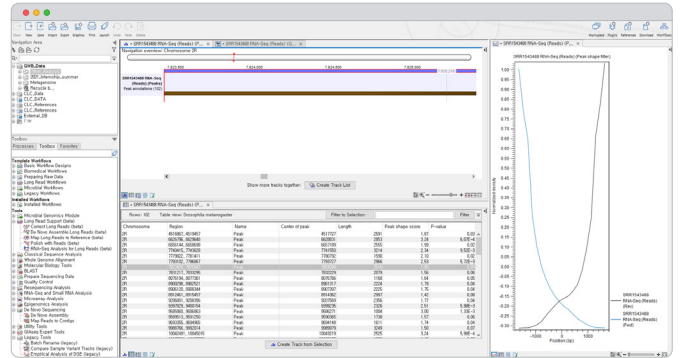
계산된 발현 값과 샘플의 정보가 담긴 메타데이터를 이용하여 그룹을 구분할 수 있습니다. PCA plot을 이용하여 샘플 간의 correlation을 확인할 수 있으며, 발현 정도가 유사한 그룹 간 클러스터링을 지원하고 heat map을 그릴 수 있습니다. 또한 negative binomial distribution을 이용한 통계분석 수행 후, volcano plot으로 확인하고 그 결과를 이용하여 그룹 간 DEG 정보를 venn diagram으로 확인할 수 있습니다.



RNA-Seq 분석 시 샘플의 정확한 발현 값 측정을 위해 Thermo Fisher Scientific의 spike in을 control로 사용하였다면 ERCC 파일 혹은 자체 시퀀스를 추가하여 분석할 수 있습니다.

ChIP-Seq analysis

CLC Genomics Workbench를 이용한 ChIP-Seq 분석은 실험적으로 짧은 DNA 절편에 binding 하는 특성 때문에 antibody만을 사용한 대조군을 설정하여 비교 및 통계적인 확률 등의 계산이 가능합니다. 또한 binding site 근처에 존재하는 유전자의 정보도 제공하며, consensus binding 서열을 도출하고, Histone ChIP-Seq, Transcription Factor ChIP-Seq도 가능합니다.



Genome informatics framework

CLC Genomics Workbench에서 NGS 데이터를 통하여 다양한 분석을 수행하고 결과들을 연계하여 확인할 때 사용할 수 있는 track tool 들은 genome 수준의 데이터들을 한눈에 볼 수 있도록 브라우저 시스템과 같은 형태로 사용할 수 있습니다.



Classical sequence analysis

CLC Genomics Workbench에서는 NGS 데이터 분석 외에 기본적인 생물정보분석 툴(BLAST, Primer design, Multiple alignment 등)들이 통합되어 있으므로 이후 2차 분석에 효율적으로 사용할 수 있습니다.