Abstracts of papers presented at the 2013 meeting on GENOME INFORMATICS

October 30–November 2, 2013

Cold Spring Harbor Laboratory
Cold Spring Harbor, New York
Medsavant
• integrated software for storage, annotation, filtration, prioritization, and inspection of genomic variants
• Open source

ELASPIC
• Method to predict stability and affinity changes induced by non-synonymous

The Genome Modeling System—An analysis engine for next generation genome sequencing
Keynote by **Michael Snyder**

Snyderome – Personal genome with 64 time points in 43 months, 6 during viral infections

- integrated analysis of Proteome, transcriptome, metabolome [george mias]

- Only 94% variants overlap from same DNA, same technology/machine using Illumina (60X covg, GQ>50)

Inosine chemical erasing (ICE)-Seq for RNA editing
- A -> I, role in stability, structure and splicing

SAILFISH
- To find gene or isoform abundance from RNA-seq
- Uses K-mers instead of reads
- 15 minutes for a set of 150 million reads vs 6 hours
- as accurate as existing read-mapping-based tools such as eXpress and Cufflinks
- [https://www.cs.cmu.edu/~ckingsf/software/sailfish/index.html](https://www.cs.cmu.edu/~ckingsf/software/sailfish/index.html)
POSTER SESSION 1

#29 Effect of SNPs on Transcription Factor Binding
- Based on regulomedb and chipseq data (SEM score)

#30 CUTTERHEAD
- Based on complete set of isoforms
- Improve the recall among low-expressed transcripts from 3.9% to 28.6%

#31 IonTorrent RNA-Seq
- Tophat-Cufflinks vs STAR-cufflinks vs TMAP-Cufflinks

#32 DGIdb – Druggable gene interaction db
- Compendium of drug-gene interactions
  - http://dgidb.genome.wustl.edu/ or dgidb.org

#33 Positive Unlabelled learning for splicing Elicitation (PULSE)
- ML approach to identify isoforms
- 54480 isoforms derived from Illumina’s Human BodyMap
POSTER SESSION 1

#42 ICGC data portal
   - 64 projects, 25000 genomes
   - w-expressed transcripts from 3.9% to 28.6%

#59 Constructing ancestral history of tumors using mutational frequencies

#60 Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms
   - Faster than BWA+GATK, equally accurate and sensitive

#61 B Allele read depth (BARD) for CNAs in tumors
   - HMM based approach infers minor and major allele copy numbers in tumor and normal genomes

#68 miRNA target interactions across diverse cancer types
   - Expression profiling of 3000 tumors
   - Network of 40 miRNAs and 72 RNAs

#72 iRegulon
   - Cytoscape plugin to predict regulatory network from set of genes
POSTER SESSION 1

#42 ICGC data portal
- 64 projects, 25000 genomes
- w-expressed transcripts from 3.9% to 28.6%

#59 Constructing ancestral history of tumors using mutational frequencies

#60 Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms
- Faster than BWA+GATK, equally accurate and sensitive

#61 B Allele read depth (BARD) for CNAs in tumors
- HMM based approach infers minor and major allele copy numbers in tumor and normal genomes

#68 miRNA target interactions across diverse cancer types
- Expression profiling of 3000 tumors
- Network of 40 miRNAs and 72 RNAs

#72 iRegulon
- Cytoscape plugin to predict regulatory network from set of genes
SESSION 4 : Sequencing pipelines and assembly

✓ GEMINI (GEnome MINIng)
  • Database framework for variant annotation and filtering
  • Uses SQLite
  • Annotates P-fam domain, clinivar, KEGG and HPRD

✓ SCALPEL
  • microassembly approach to detect INDELs
  • Scalpel vs SOAPindel vs Haplotypecaller
POSTER SESSION 2

#135 HOTSPOT
  - To infer functional importance of mutation through 3D proximity analyses

#137 SeqWare
  - Sequence analysis on cloud

#139 REFEX: Reference expression dataset for tissue transcriptome
  - http://refex.dbcls.jp/

#165 Molecular Diagnostics Service at Memorial Sloan-Kettering Cancer Center.
  - 300 genes, >1000

#177 Visualizing consequences of genetic variation in biological networks
  - Cytoscape plugin

#191 Whole transcriptome splicing analysis of subtype specific signature of prostate cancer
SESSION 8: Population and personal genomes

✓ Analysis of protein-coding variation in over 50,000 individuals
  • Reduced BAM file (collapsing the reads with consensuvariant is confidently called)
  • Annotates P-fam domain, clinivar, KEGG and HPRD

✓ Annotation of variants from 1092 humans-application to cancer genomes
  • FunSeq - Finding ultra-sensitive non-coding regions & disruptive mutations
  • Negative selection in rare non-coding variants
  • Variants in TFBS are more
  • ~99% of somatic SNVs occur in non-coding regions, including TFBSs, ncRNAs and pseudogenes