**Genome Informatics**

 **Session** **1** **VARIANT DISCOVERY AND GENOME ASSEMBLY** WEDNESDAY 11/1/2017, 7:30 PM

 Laura Clarke / Jared Simpson

 **#** **lname** **Title** **Talk Length**

 1 Clarke Variation and assembly resources at EMBL-EBI 15

 2 Hefferon tmVar 2.0—Integrating information on genomic variants from biomedical literature with dbSNP 15

 and ClinVar

 3 Sivakumar Identification and correction of problematic copy number calls in TCGA 15

 4 Giordano Sequence presence-absence detection in assembly pairwise comparison with scanPAV 15

 5 McLean Mastering variant calling of SNPs and small indels with deep neural networks 15

 6 Koren Can Nanopore sequencing finally finish the human genome? 15

 7 Garimella Graph-based discovery of complex *de novo* structural mutations in *P. falciparum* 15

 experimental crosses

 8 Kahles Large scale genomics with scalable reference graphs 15

 **Session** **2** **TRANSCRIPTOMICS, ALTERNATIVE SPLICING, GENE PREDICTIONS** THURSDAY 11/2/2017, 9:00 AM

 Mihaela Pertea / Oliver Stegle

 **#** **lname** **Title** **Talk Length**

 9 Pertea A new comprehensive human gene catalog 15

 10 Sahlin IsoCon—A novel algorithm combined with targeted transcriptome sequencing of multicopy gene 15

 families traces the origins of highly similar transcripts to individual gene copies

 11 Vitting-Seerup The landscape of isoform switches in human cancers 15

 Stegle Computational approaches for understanding single-cell expression variation 15

 12 Ma SQUID—Transcriptomic structural variation detection from RNA-seq 15

 13 Wu Transcriptome-guided genomic alignment and analysis 15

 14 Phillippy How to create a whole-genome human homology map in around a minute 15

 15 Zappia Simulation and analysis tools for single-cell RNA sequencing data 15

 **Session** **3** **POSTER SESSION I** THURSDAY 11/2/2017, 1:30 PM

 **#** **lname** **Title** **Talk Length**

 16 Adamson Genomic analysis of germline and somatic variation in high-grade serous ovarian cancer

 17 Adkins LGTSeek—A robust distributable lateral gene transfer pipeline

 18 An Systematic annotation of regulatory elements in blood cell lineage

 ***Page 1 of 10***

 19 Arora In-depth characterization of a hallmark for balancing selection—HLA heterozygote advantage

 against HIV-1

 20 Arumilli Structural variation and genome evolution in domestic dog

 21 Asghari Copy number and tumor purity estimation from targeted cell free DNA sequencing data

 22 Atwal Inferring originating cell type of cancer metastases using the spatial distribution of mutations

 23 Ballinger Modelling double strand break hotspots to interrogate structural variation in cancer

 24 Banuelos Genetic variants over generations—Sparsity-constrained optimization tools for structural variant

 detection

 25 Barreira Exploring the sequence composition, functional capacity, and regulatory role of large tandem

 repeats and their adjacent sequences in regeneration

 26 Bartholdy Genetic admixture and differentiation states shape the human methylome in stem, progenitor and

 somatic cells

 27 Berninger Yeast artificial chromosomes for biosynthetic pathway assembly

 28 Bis Rare variant burden analysis to decipher genetic architecture of Charcot-Marie-Tooth disease

 29 Blankenberg Choosing the best of all worlds—à la carte access to extant and emergent best-practice

 metagenomic pipelines

 30 Bono RefEx—A reference gene expression dataset as a web tool for the functional analysis of genes

 31 Brady MetaTRANSiT—A comprehensive toolkit for metatranscriptomics and metagenomics analysis

 32 Cain Serverless JBrowse on the cheap

 33 Chapman Machine learning strategies to identify high confidence structural variants in human genome

 reference materials

 34 Chatterjee Previously undetected genomic signatures of giant viruses are ubiquitous in metagenomes

 35 Chen Individual ancestry estimation from whole exome sequencing data in patient-derived xenograft

 samples

 36 Chen Profiling of somatic alterations in BRCA1-like breast tumors

 37 Chen De novo assembly of goldfish using pacbio long reads

 38 Cheng H3K27 tri-methyltransferases CLF and SWN redundantly buffer ABA-induced senescence in

 Arabidopsis

 39 Chervitz Harnessing a gold standard data set for immuno-oncology

 40 Chhatbar Regulation of gene expression in response to DNA base composition

 41 Choi Integrative genomic analysis of 176 Korean liver cancer reveals distinctive molecular pattern

 42 Chougule Genome annotation using the MAKER-P JetStream cloud

 43 Chung Understanding mammary stem cell state regulation through chromatin accessibility

 ***Page 2 of 10***

 44 Clarke The Human Cell Atlas data coordination platform

 45 Crouch Mapping of R-loops in *Trypanosoma bruce*i reveals conserved and novel functions

 46 Cumbo GDCWebApp—Filtering, extracting, and converting genomic and clinical data from the Genomic

 Data Commons Portal

 47 Cummins Moving towards comparative analysis of hundreds of vertebrate genomes

 48 Dale Streamlining the installation of thousands of bioinformatics software packages with Bioconda

 49 Darby Leveraging linked reads for single-sample somatic variant calling

 50 Deng Comprehensive genomic landscape of most commonly used breast cancer cell lines and patient

 derived xenograft models

 51 Dierckxsens NOVOLoci—Targeted assembly and variance detection from whole genome data

 52 Dolzhenko ExpansionHunter—A software tool to detect long repeat expansions from PCR-free whole-

 genome sequence data

 53 Dubarry Gmove—A tool for eukaryotic gene prediction using various evidence

 54 Duitama NGSEP3—Accurate, efficient and user friendly production and analysis of genomic variation

 datasets through STR-aware integrated realignment

 55 Erdos Unifying data submission at EMBL-EBI—A user-focused approach

 56 Fong DXM—An algorithm to deconvolve genomic DNA methylation data to understand epigenetic

 clonality

 57 Freed Improving CREST with the Sentieon Python API

 58 Gale Hammell Robust analysis of single cell transcriptomes using SAKE identifies markers of targeted inhibitor

 resistance in melanoma

 59 Gel Moreno Plot any data on any genome with karyoploteR

 60 Gogate Dynamic change of transcription pausing through modulating NELF protein stability regulates

 granulocytic differentiation

 61 Gonnella Random access to sequence graphs stored in large GFA files

 62 Gordon Accelerating congenital heart defect variant analysis through big data

 63 Grote *ABAEnrichment* and *GOfuncR*—Two R-packages for ontology enrichment

 analyses

 64 Guerra-Assuncao GenomeChronicler—The PGP-UK genomic report generator

 65 Gurran Splice-QTLs in the context of predisposition to colorectal cancer

 66 Guturu Improving community search on an identity-by-descent graph with millions of individuals

 67 Halldorsson Estimating RNA expression using personal genomes

 68 Hamilton The Sweetpotato Genomics Resource

 69 Hammond Creating pipelines that are repeatable, traceable, and shareable for clinical genomics and

 research

 ***Page 3 of 10***

 70 Hansen Precise detection and specification of structural variation in genomes

 71 Harris *Roslin*—A portable and reproducible workflow infrastructure for cancer genomic

 sequencing analysis

 72 Havrilla A map of highly constrained coding regions in the human genome

 73 Herrero Allele specific HLA loss is a pervasive mechanism of immune evasion and is permissive for non-

 small cell lung cancer evolution

 74 Hinrichs UCSC Variant Annotation Integrator command-line wrapper and HGVS variant nomenclature

 support

 75 Holley Highly parallel and memory efficient compacted de Bruijn graph construction

 76 Hunt A manual annotation workflow integrating signals of protein-coding conservation together with

 next-generation sequencing to identify novel protein-coding and pseudogene loci

 77 Jiao The complex sequence landscape of maize revealed by single-molecule technologies

 78 Johnson Characterizing epigenetic intratumoral heterogeneity in glioma using single-cell reduced

 representation bisulfite sequencing

 79 Kanduri Co-localization analyses of genomic elements—Essential facets and potential pitfalls

 80 Khalouei The Forge Bioinformatics Pipeline

 81 Kim A general web platform for integrating experimental proteomics data and results from GWAS and

 exome sequencing projects

 82 Kim Manually curated 16S rRNA database and associated seamless updating platform

 83 Kiran Long-noncoding RNA based prognostic signature for gliomas

 84 Kitts Diploid genome assemblies at NCBI

 85 Kolora The role of rearrangements in the divergent evolution of European green lizards

 86 Kothiyal Analysis of Mendelian inheritance errors in deep sequenced whole genomes from 1314 trios

 identifies population-specific structural variants

 87 Kovaka UNCALLED—An aligner for quickly mapping raw Nanopore signals to large references

 88 Kumar Semantic annotation and knowledge extraction using iCLiKVAL

 89 Kumar Integrating data, tools and knowledge to accelerate scientific discovery using open-source,

 biological, data-science platform of KBase

 90 Lariviere Hybrid assembly of small genomes in Galaxy

 91 Laverty Learning variable gapped sequence-structure motifs for RNA-binding proteins

 92 Lee Discovery of medium and long-sized insertion variants with accurate break points and flanking

 sequences

 93 Li Landscape of somatic mutations in inflammatory breast cancer whole-genome sequences

 94 Lin A human-specific switch of alternatively spliced *AFMID* isoforms contributes to

 *TP53* mutations and tumor recurrence in hepatocellular carcinoma

 95 Lin Evidences for the role of ZBTB33 (Kaiso) in heterochromatin priming

 ***Page 4 of 10***

 96 Ling Novel exon discovery in cellular differentiation and human disease by utilizing the Snaptron

 framework

 97 Liu Investigating the association between polygenic risk score and coronary artery calcification

 98 Lochovsky Efficient detection of highly mutated regions with Mutations Overburdening Annotations Tool

 (MOAT)

 99 Lu SciApps—A cloud-based platform for reproducible bioinformatics workflows

 100 Mahurkar Human Microbiome Project (HMP) Data Resource—A web portal for exploring and accessing

 HMP data, analysis products, and tools

 101 Marcais MUMmer4—A fast and versatile genome alignment system

 102 Martinez Barrio Resolving the full spectrum of human genetic variation using Linked-Reads

 103 Martinez Cuesta Exploring the chemistry and biology of nucleotide modifications in mammals, parasites and disease

 104 McCarter An efficient algorithm for learning a gene network underlying clinical phenotypes under SNP

 perturbations

 105 McCarthy Scaling up reference quality assembly of vertebrate genomes

 106 McCracken Pipeline for SNP discovery in RNA sequences

 107 McGee Sleuth-ALR—Improving estimation of \*Seq differential analysis using compositional data analysis

 with Sleuth

 108 Mclaren A variant by any other name... Ensembl’s Variant Recoder

 **KEYNOTE SPEAKER** THURSDAY 11/2/2017, 3:30 PM

 **Maricel Kann, “A protein-domain approach for the analysis of disease mutations”**

**Session** **4** **DATA CURATION AND VISUALIZATIION** THURSDAY 11/2/2017, 7:30 PM

 Gabor Marth / Ann Loraine

 **#** **lname** **Title** **Talk Length**

 Marth No abstract 15

 109 Kingsford Large-scale search of short-read sequencing experiments 15

 110 Layer STIX—A scalable index for mining large whole-genome sequencing cohorts for reliable 15

 structural variant population allele frequency estimates

 111 Mottarella GVCFLIB—An extensible library to analyze and accurately annotate clinically significant wild type 15

 and mutant alleles from gvcf formatted callsets

 112 Loraine Genome browsing on someone else’s computer 15

 113 Khatamifard BioArch—A reconfigurable hardware accelerator designed for bioinformatics workloads 15

 114 Nekrutenko Enhancing pre-defined workflows with ad hoc analytics using Galaxy, Docker and Jupyter 15

 115 Wagner Metaviz and the Human Microbiome Project Data Portal—Interactive statistical and visual 15

 analysis of metagenomic data from the HMP

 ***Page 5 of 10***

 **Session** **5** **COMPARATIVE AND METAGENOMICS** FRIDAY 11/3/2017, 9:00 AM

 Paul Flicek / Holly Bik

 **#** **lname** **Title** **Talk Length**

 Flicek No abstract 15

 116 Chintalapati A *de novo* assembly of the Neandertal genome provides insights into human structural 15

 variation.

 117 McLain The evolution of lifespan and the epigenome assessed by CpG frequency in conserved primate 15

 and vertebrate promoters

 118 White Global analysis of human mRNA folding disruptions in synonymous variants demonstrates 15

 significant population constraint

 119 Bik Linking molecules with morphology in the -Omics age—Computational taxonomy pipelines for 15

 microbial metazoan

 120 Lloyd-Price Strains, functions, and dynamics in the expanded Human Microbiome Project 15

 121 Molik K-mer comparison methods in metagenomics, applications at the community level 15

 122 Dikow A whole-genome phylogenetic hypothesis across the three domains of life 15

 **Session** **6** **EPIGENOMICS AND NON-CODING GENOME** FRIDAY 11/3/2017, 1:30 PM

 Elena Rivas / Adam Siepel

 **#** **lname** **Title** **Talk Length**

 123 Rivas A statistical test for structural covariations in RNA and proteins 15

 124 Hait Large-scale analysis of genome-wide enhancer and gene activity reveals a novel enhancer- 15

 promoter map

 125 Viner Modeling methyl-sensitive transcription factor motifs with an expanded epigenetic alphabet 15

 126 Dumelie Near-nucleotide mapping of R-loops shows that promoter-associated R-loops are bounded at first 15

 exon-intron junctions

 127 Siepel New methods for measuring natural selection and predicting deleterious variants in noncoding 15

 regions of the human genome

 128 Schubach Selective constraints on enhancer and promoter sequences across human cell-types 15

 129 Young The consequences of promoter birth and death in the human population 15

 130 Lihm Meta-analysis of chromatin accessibility to determine meaningful variation 15

**KEYNOTE SPEAKER** THURSDAY 11/2/2017, 4:30 PM

 **Lior Pachter, “Post-procrustean bioinformatics”**

 **Session** **7** **POSTER SESSION II** FRIDAY 11/3/2017, 5:30 PM

 **#** **lname** **Title** **Talk Length**

 131 Melsted Fast genome alignments from pseudoaligned RNA-Seq datasets using kallisto

 132 Meynert Developing a Scottish variant repository

 133 Michno Identification of candidate genes underlying nodulation-specific phenotypes in *Medicago*

 *truncatula* through integration of genome-wide association studies and co-expression networks

 134 Miller Designing cancer vaccines for trials of personalized immunotherapy

 135 Muir Pseudogenes in the mouse lineage—Transcriptional activity and strain-specific history

 136 Nandu Identification, regulation, and function of antisense transcription in the estrogen response in

 breast cancer cells

 137 Nehyba Copy number variation analysis using a targeted next-generation sequencing amplicon

 panel—NEXTflex® DMD Amplicon Panel for Duchenne muscular dystrophy

 138 Ning Scaff10x—A relational matrix based algorithm for genome scaffolding using 10x data

 139 Nothaft Highly scalable genome analysis using ADAM, Cannoli, and Avocado

 140 Oguz Network analysis of transcriptome identifies predictive biological pathways in hypertensive African

 Americans

 141 Oliver SREVED—Splicing regulatory element variant effect determination

 142 Olley The gEAR portal—Sharing and displaying gene expression now simplified and diversified

 143 Olson Gene tree guided search and visualization at Gramene

 144 Orvis Rapid, large-scale annotation using the cloud-enabled Genomic Annotation Logic and Execution

 System (GALES)

 145 Pacheco RNA sequencing and proteomics approaches reveal novel deficits in the cortex of *Mecp2*-

 deficient mice, a model for Rett syndrome

 146 Parekh zUMIs—A fast and flexible pipeline to process RNA sequencing data with UMIs

 147 Patricio Scaling up the TreeFam resource in Ensembl

 148 Petersohn Large-scale distributed genomic analysis using Lime and Gnocchi

 149 Phan dbSNP 2.0

 150 Piccolo Geney—A data ecosystem that enables biologists to efficiently subset, visualize, and analyze

 genomic data

 151 Piccolo Benchmarking 50 classification algorithms on 45 transcriptional biomarker datasets

 152 Powell Assembly optimization in both space and time of the largest genome to date

 153 Pritt FORGe—Prioritizing variants in graph genomes

 154 Pulman MicrobiomeDB—A web-based data-mining platform for interrogating microbiome experiments

 155 Qi Absence of Receptor for Hyaluranon-mediated Motility (RHAMM) alters genome-wide mutational

 landscapes associated with tumorigenesis and metastasis

 156 Qi A new method of hisone-modification guided genome assembly

 157 Quilez Oliete Managing the analysis of high-throughput sequencing data

 ***Page 7 of 10***

 158 Ramakrishnan Benchmarking RNA-Seq in plant species

 159 Ramakrishnan Identification of DNA methylation driven transcriptomic alterations and clinical outcomes in African

 American men with prostate cancer

 160 Ramani Alternative splicing of neurofibromin 1 is associated with elevated MAPK activity and poor

 prognosis in glioma

 161 Rangavittal DiscoverY—A fast and lightweight method to isolate Y chromosome-specific sequences

 162 Ronquist An algorithm for cellular reprogramming

 164 Roskosch PopDel—Population-scale detection of genomic deletions

 165 Rubanova Trackature—Reconstructing mutational signatures through time to track tumour evolution

 166 Russell An R client for The Cancer Imaging Archive REST API

 167 Sahin Machine learning and computer vision approaches for phenotypic profiling in yeast

 168 Salvatore The rainfall plot—Its motivation, characteristics and pitfalls

 169 Domanska Hierarchical GSuite HyperBrowser—Analysis across multiple dimensions of epigenomic variation

 170 Sankoff The similarity distribution of gene pairs created by recurrent alternation of whole genome

 duplication and fractionation in plants

 171 Sasse Predicting preferences of RNA binding proteins from protein sequence

 172 Saunders Strelka2—Fast and accurate small variant calling for germline and cancer sequencing

 applications

 173 Schneider Behind the veil—Using visualization tools to examine genome curation

 174 Schulze Debacter—High-resolution decontamination of genomes using deep sequencing data

 175 Sedlazeck Accurate and fast detection of complex and nested structural variations using long read

 technologies.

 176 Sedlyarov A comprehensive pipeline for analysis of complex-setup genome-scale and focused pooled

 CRISPR-Screens

 177 Shah Assembly of individual chromosomes at multi-megabase scale using Linked-Reads

 178 Sharma Nucleoprotein of influenza A virus regulates host translation machinery by targeting mTOR-eIF4E

 pathway proteins and its controlling microRNAs

 179 Shatz Environmental Health Sciences Data Commons (EDAC)—A research data management and data

 workflow automation system

 180 Shim McSplicer—A probabilistic model for alternative splicing

 181 Shin Characterization of background errors in targeted deep sequencing data specifically associated

 with plasma DNA

 182 Shirley PISCES—A package for quantitation and QC of big mRNA-seq datasets

 183 Shrestha HIT'nDRIVE—Patient-specific multi-driver gene prioritization for precision oncology

 184 Shu Rare splice sites in plant protein-coding genes

 185 Singh Increasing the lower limit of detection for mutations with limited number of reads using unique

 molecular identifiers and consensus building

 186 Skvir *RetroSuite*, an integrated pipeline for the genome-wide analysis of transposable elements

 ***Page 8 of 10***

 187 Smith RGD—Data and tools for precision models of human disease

 188 Sompallae Validation and implementation of KidneySeqTM—A comprehensive gene panel for

 genetic renal diseases

 189 Song ClassX—Scalable simultaneous transcript assembly of multiple RNA-seq data sets

 190 Standage Kevlar—Reference-free variant discovery in human genomes and beyond

 191 Tanaka Genome rearrangement triggered by thermostable restriction enzyme inducing multiple DNA

 double-strand breaks

 192 Taylor Quality assessment and large-scale integration of chromosome conformation capture datasets

 193 Tello-Ruiz Gramene—Comparative genomics, gene expression and pathway reference resources for plant

 communities

 194 Thodberg The *in vivo* transcription start site and enhancer landscape of inflammatory bowel disease

 enables disease classification and interpretation of non-coding SNPs

 195 Tutaj Comparison of rat strain-specific variants collection for all rat genome references

 196 Vasoya Rapid identification of MHC alleles and haplotypes in genetically divergent cattle populations

 Using NGS

 197 Vegesna Copy number and expression variation in ampliconic genes on the Y chromosomes of great apes

 198 Vieth powsimR—Power analysis for bulk and single cell RNA-seq experiments

 199 Wang Linked-Reads vs long reads—Balancing cost and contiguity in the *Vitis cinerea* genome

 200 Wang Computational modeling long non-coding RNAs mediated transcription regulation

 201 Wang GTD—Estimating genotype likelihood by deep neural networks

 202 Wang deFusion—A tool to improve predictions of tandemly duplicated genes created by the MAKER

 annotation pipeline

 203 Wang Metagenomics study of individuals with seasonal influenza

 204 Wang Convergence of light, stresses and circadian rhythm on nuclear-encoded chloroplast-localized

 genes (NECGs) in Arabidopsis

 205 Wasmuth Sparkle—Finding missing genes in draft genomes

 206 Wesolowski How shapes of nucleosomal DNA can regulate gene expression

 207 Wilks Snaptron—A tool and service for studying splicing in tens of thousands of individuals

 208 Williams Gene fusion information management

 209 Wintersinger Correcting discordance between six copy-number-calling methods for 2778 tumours

 210 Wyman Generating full-length, high-quality human transcriptomes from PacBio Iso-seq data

 211 Xue An iterative approach for reconstructing full-length ribosomal genes from whole

 metatranscriptomic data

 212 Yan Modularity analysis of enhancer-promoter interaction networks

 213 Yang JULIP++—Fast and ultra-sensitive identification of differential splicing events from large RNA-

 seq data collections

 214 Ye Identification of predictors for therapeutic response to immunotherapy

 ***Page 9 of 10***

 215 Ye Genomics and bioinformatics approach to investigating the role of repetitive elements in the

 chemoprevention of colorectal cancer

 216 Young Learning RNA binding protein motifs using convolutional neural networks

 217 Zhang The Cancer Genome Collaboratory

 218 Zhang Jtracker—Workflow management and execution backed on Git repository with full provenance

 219 Zhang An improved assembly identifies new features of cucumber genome

 220 Zhou The genome organization of an autotetraploid potato（*Solanum tuberosum* L.）

 221 Ziegler An integrative roadmap to PAX3 target gene networks in melanocytes and melanoma

 222 Zimin Hybrid assembly of challenging genomes with MaSuRCA mega-reads

 **Session** **8** **PERSONAL AND MEDICAL GENOMICS** SATURDAY 11/4/2017, 9:00 AM

 Konrad Karczewski / Suzanne Leal

 **#** **lname** **Title** **Talk Length**

 223 Karczewski The spectrum of loss of function tolerance in the human genome 15

 224 Gerstein Passenger mutations in 2500 cancer genomes—Overall molecular functional impact and 15

 consequences

 225 Kelly Utilization of linked-read, whole genome, whole exome and transcriptome sequencing in the 15

 comprehensive molecular profiling of pediatric brain tumors

 226 Staples Completing a human gene knockout catalog through accurate phasing of 15K rare, deleterious 15

 compound heterozygous mutations in 61K exomes

 227 Leal SEQSpark—An analysis tool for large scale sequence-based genetic epidemiological studies 15

 228 Neretti The bioinformatics of liquid biopsies—Cell-free DNA as a biomarker of disease and aging 15

 229 Kaiser Chromatin loop anchors are associated with genome instability in cancer and recombination 15

 hotspots in the germline

 230 Chesi An ultra-high resolution capture-C promoter ‘interactome’ implicates causal genes at SLE GWAS 15

 loci

 ***Page 10 of 10***