Day 1

8:30-9:00   opening remarks
9:00-10:00   keynote

Translating a trillion points of data into therapies, diagnostics, and new insights into disease **Atul Butte**

10:00-10:25   coffee break
10:25-12:05   parallel sessions (100 min)

| Molecular diagnosis with genomic data | 1. Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis  
2. The Diagnostic Application of RNA Sequencing in Patients with Thyroid Cancer: an Analysis of 851 Variants and 133 Fusions in 524 Genes  
3. Meta-analysis of sex differences in gene expression in schizophrenia  
4. Characterizing redescriptions using persistent homology to isolate genetic pathways contributing to pathogenesis |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Recent statistical advancement in Network based analysis | 1. SUMONA: A Supervised Method for Optimizing Network Alignment  
2. Characterizing mutation-expression network relationships in multiple cancers  
3. A Network based Covariance Test for Detecting Network Multivariate eQTL in Saccharomyces cerevisiae  
4. 146 Generalized logic model based on network topology able to capture the dynamical trends of cellular signaling pathways |

12:05-1:05   lunch break
1:05-2:05   keynote

The multi-facets of a data science project to answer: how are organs formed? **Bin Yu**

2:05-2:30   coffee break
2:30-4:35   parallel sessions  (125 min)

| Genetic Evolution: recent advancement | 1. Computation of Hybridization Networks on Realistic Phylogenetic Trees  
2. Algorithms for Pedigree Comparison  
4. DTL-RnB: Algorithms and Tools for Summarizing the Space of DTL Reconciliations  
5. Mapping the genomic architecture of adaptive traits with interspecific introgressive origin: A coalescent-based approach |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| miRNA analysis | 1. The Modularity and Dynamicity of miRNA-mRNA Interactions in High-Grade Serous Ovarian Carcinomas and the Prognostic Implication  
2. Identification of miRNA-mRNA regulatory modules by exploring collective group relationships  
3. c-Myc and viral cofactor Kaposin B co-operate to elicit angiogenesis through modulating miRNome traits of endothelial cells  
4. Identification of microRNA precursor based on gapped n-tuple structure status composition  
5. Missing value imputation for microRNA expression data by using a GO- |
4:35 -6:00 Poster Session I / Sponsor Exhibit (Complimentary hors d'oeuvres will be served)

**Day 2**

8:30-9:30  Keynote speech
Finding the footprints of evolutionary adaptation in the human genome. **Rasmus Nielsen**

9:30-9:50am  coffee break

9:50-12:10 parallel sessions (5, Beijing Alumni) (125 min + 15 minutes break)

| Protein Function and Statistical methodology development | 1. Protein inference: A protein quantification perspective  
2. Predicting the Absorption Potential of Chemical Compounds through Deep-Learning Approach  
3. Multi-Instance Multi-Label Distance Metric Learning for Genome-Wide Protein Function Prediction  
4. Power Estimation and Sample Size Determination for Replication Studies of Genome-Wide Association Studies  
5. Protein-protein Interface Residues Share Similar Hexagon Neighborhood Conformations |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee Break (15 minutes)</td>
<td></td>
</tr>
</tbody>
</table>
| Session PKU Alumni | 1. Panelist: Dr. Minghua Deng  
2. Panelist: Dr. Jing Huang  
3. Panelist: Dr. Jie Peng  
   Coffee Break (15 minutes)  
4. Panelist: Dr. Xinmin Zhang  
5. Panelist: Dr. Fengzhu Sun |

12:10-1:30  lunch break

1:30-3:10 parallel sessions (100 min)

| Optimize Genomic information in sequencing data | 1. Identifying micro-inversions using high-throughput sequencing reads  
2. Locating rearrangement events in a phylogeny based on highly fragmented assemblies  
3. Codon Context Optimization in Synthetic Gene Design  
4. A maximum-likelihood approach for building cell-type trees by lifting |
|---|---|
| Protein Function | 1. SOHSite: incorporating evolutionary information and physicochemical properties to identify protein S-sulfenylation sites  
2. PredRSA: A Gradient Boosted Regression Trees Approach for Predicting Protein Solvent Accessibility  
3. A new scheme to discover functional associations and regulatory networks of protein ubiquitination  
4. Incorporating two-layered machine learning method with substrate |
motifs to identify lysine ubiquitination sites

3:10-3:30 Coffee Break

3:30 – 5:10 parallel sessions (100 min)

| Algorithm development and machine learning in Drug Discovery and precision medicine | 1. Drug Repositioning Discovery for Non-Small Cell Lung Cancer by Using Machine Learning Algorithms and Topological Graph Theory  
2. PDOD: Prediction of drugs having opposite effects on disease genes in a directed network  
3. Algorithmic Mapping and Characterization of the Drug-Induced Phenotypic-Response Space of Parasites Causing Schistosomiasis  
4. Inference of domain-disease associations from domain-protein, protein-disease and disease-disease relationships |
2. Computational prediction of CRISPR cassettes in gut metagenome samples from Chinese type-2 diabetic patients and healthy controls  
3. Learning a Hierarchical Representation of the Yeast Transcriptomic Machinery using an Autoencoder Model  
4. Transcriptome Sequencing Based Annotation and Homologous Evidence Based Scaffolding of Anguilla Japonica Draft Genome |

5:10 -6:30: Poster Session II/ Exhibition

6:30-8:30 on-site Banquet with dinner speaker (open to all registered attendees)

**Day 3**

9:00-10:00 Keynote Speech by Dr. Robert Gentleman (60 min)

10:00-10:20 coffee break

10:20 – 12:00 parallel sessions (100 min)

| Epigenome and RNA-seq Data Analysis | 1. genome reconstruction with ShRec3D+ and Hi-C data  
2. A Full Bayesian Partition Model for Identifying Hypo- and Hyper-methylated Loci from Single Nucleotide Resolution Sequencing Data  
3. A tale of two gene sets: low and high variability in single cell RNA-seq data  
4. Copy Number Variants Calling for Single Cell Sequencing Data by Multi-constrained |
| Data mining and feature extraction in biomedical applications | 1. A semi-parametric statistical model for integrating gene expression profiles across different platforms  
2. Weakly Supervised Learning of Biomedical Information Extraction from Curated Data  
3. hc-OTU: A Fast Homopolymer Compaction-Based Operation Taxonomic Unit Clustering Algorithm  
4. Medoidshift Clustering Applied to Genomic Bulk Tumor Data |

12:00 – 1:00 Lunch break
1:00 – 2:15  two parallel sessions (75 min)

| Epigenome                                      | 1. Predicting transcription factor site occupancy using DNA sequence intrinsic and cell-type specific chromatin features  
|                                               | 2. Epigenome Overlap Measure (EpOM) for comparing tissue/cell types based on chromatin states  
|                                               | 3. MOCCS: clarifying DNA-binding motif ambiguity using ChIP-Seq data |
| RNA-seq Data Analysis                         | 1. Bayesian method for estimate Allele-specific expression based RNA-seq  
|                                               | 2. A non-negative matrix factorization based preselection procedure for more accurate isoform discovery from RNA-seq data  
|                                               | 3. RDDpred: A condition-specific RNA-editing prediction model from RNA-seq data |

2:15- 2:30 Coffee Break

2:30 – 3:30 Keynote Speech by Dr. Wing Hung Wong

3:30 – 4:00 closing remarks, award ceremony and lottery drawing (see you in a year!)