RECOMB (no 19th !) was excellent in both talks, keynotes and organization. I have the book of abstracts (extended and short) for those who want to read some of the papers. I list below several that are of particular interest to our projects and some that are innovative and present notable methods.

PS. Some of the papers were only available as short abstracts as authors chose to submit full papers elsewhere. If any of you comes across the full paper please send the link around to all.

Bonnie Berger  
*Computational biology in the 21st century: Algorithms that scale*  
(KEYNOTE)

Berger and her students developed a long and impressive series of algorithms (most published already) for handling very large scale data efficiently. The idea is to perform preprocessing on the data so that later on searches can be done in sublinear time. The approach is applied to sequence comparison (cablast, cablat, psi-blast), mapping (cora), metagenomics (caBlastX), chemogenomics (ammonite) and k-mer quality score compression (quartz). The unifying theme in all works are exploiting low fractal dimension and metric entropy.

The new principles should be of interest to all. The specific techniques – especially to Roye.

Rob Patro, Stephen Mount and Carl Kingsford. *Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms*  
(Highlight)

The goal here is to quantify each isoform of a gene in an RNAseq sample. The Sailfish method uses k-mer statistics and hashing, and saves by equivalence classes of k-mers that co-occur in the reads (Nat Biotech 14). Salmon is an improved method that (1) does not need to determine the k-mers in advance, (2) exploits paired-end better, (3) replaces hashing by BWT. The paper and code are in progress in github.

Relevant to Roye.

Emily Berger, Deniz Yorukoglu and Bonnie Berger. *HapTree-X: An integrative Bayesian framework for haplotype reconstruction from transcriptome and genome sequencing data*

Cool idea for improving haplotyping (phasing): By using differential allele-specific expression (the non-equal transcription level of the two alleles of a gene), longer phased regions can be identified. Improves over the authors HapTree method. Only short abstract is available.

Not directly relevant to anyone but a cute idea that is worth knowing.
James Zou (Microsoft), Christoph Lippert, David Heckerman, Martin Aryee, Jennifer Listgarten. Epigenome-wide association studies without the need for cell-type composition (Highlight)

A very elegant generalization of the principles of searching for disease association in epigenomic data. Unlike with genomic data, different cell composition between the cases and the controls can affect the results. The authors develop a method based on latent variable mixture model to handle cell type heterogeneity. References: Lin et al NBT 13 (EWAS for rheumatoid arthritis) and Zou et al, Nat Methods 14

Not relevant to any of our particular projects, but well worth the read.

Andrew McPherson, Andrew Roth, Cedric Chauve and Cenk Sahinalp. Joint inference of genome structure and content in heterogeneous tumor samples

Present a method to unmix tumor and contaminating normal signals in tumor reads and jointly predict genome structure and contents of each tumor clone. Only a short abstract is available – worth following.

Relevant for Ron, Rami.

Mark Leiserson, Hsin-Ta Wu, Fabio Vandin and Benjamin Raphael. CoMEt: A Statistical Approach to Identify Combinations of Mutually Exclusive Alterations in Cancer

The goal here is to identify a minimal explaining set of driver mutations in cancer. The assumption is that the driver mutations that occur should be mutually exclusive – at most one per patient. The method is based on Dendrix – a previous method from the Raphael lab. The extension is giving a score to groups and not only to pairs of mutations. The method is shown to identify pathways involved in the cancer process, and to identify subtypes based on the mutations.

For Ron and Dvir (mainly for the last point)

Mark Leiserson, Fabio Vandin, Hsin-Ta Wu, Jason Dobson, Jonathan Eldridge, Jacob Thomas, Alexandra Papoutsaki, Younghun Kim, Beifang Niu, Michael McLellan, Michael Lawrence, Abel Gonzalez-Perez, David Tamborero, Yuwei Cheng, Gregory Ryslik, Nuria Lopez-Bigas, Gad Getz, Li Ding and Benjamin Raphael. Pan-Cancer Network Analysis Identifies Combinations of Rare Somatic Mutations across Pathways and Protein Complexes (Highlight)

This talk describes the authors' method HotNet2 (Nat. Genetics 15). It seeks connected subnetworks in a PPI network that contain more mutations than expected at random. It improves over the authors HotNet (RECOMB 10) which is based on heat diffusion. The application of HotNet on a larger network revealed only hubs. The improvement here is by adding directionality and asymmetry to the diffusion process (the new process is called insulated heat diffusion).
Hyunghoon Cho, Bonnie Berger and Jian Peng. **Diffusion Component Analysis: Unraveling Functional Topology in Biological Networks**

Common "guilt by association" methods for exploiting network topology to infer structure propagate information to neighbors. The authors suggest a method that uses longer term topological relations by using random walk with restart. Starting the process from each node gives a distribution of visits to all nodes in the graph. Then dimensionality reduction is used on these distribution vectors, and the vectors are compared to associate functions. The authors show impressive results on several networks. The method, called DCA (diffusion component analysis) is simple, elegant and definitely worth pursuing (short abstract only is available).

Relevant to all, mostly to Didi.

Bo Wang, Aziz Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains and Anna Goldenberg. **Similarity Network Fusion for aggregating data types on a genomic scale** (Highlight)

The goal of this paper is to integrate heterogeneous data types available in patients (genome, metagenome, diet info, epigenetics). The authors present an elegant method called SNF that builds a network with nodes=patients for each data type and then develop an iterative methods to make the different network gradually more similar to each other, while keeping the source of similarity in the summary network. In application to cancer they show strong separation in survival curves. They show improvement over PAM50 and iCluster on breast cancer data. This paper was published in Nat. Methods 14 and the SNFtool package is available in CRAN.

Relevant mostly to Dvir, also Kobi, Didi.

Mingfu Shao and Bernard Moret. **A Fast and Exact Algorithm for the Exemplar Breakpoint Distance**

The authors describe an ILP algorithm for the following problem: Given two genomes with multiple copies of genes per genome, find a single representative (exemplar) of each duplicated gene in each genome such that the resulting two (ordinary) genomes have minimum breakpoint distance. The problem is NP-hard and the authors ILP formulation improves by two orders of magnitude over a previous exact algorithm (MSOAR). This is achieved by proving several smart observations that allow stronger formulation and also leads to adding constraints to the ILP. The interest here is the smart use of ILP.

Of interest to Ron.

Hui Yuan Xiong, Babak Alipanahi, Leo Lee, Hannes Bretschneider, Daniele Merico, Ryan Yuen, Yimin Hua, Serge Gueroussov, Hamed Najafabadi, Tim
Hughes, Quaid Morris, Yoseph Barash, Adrian Krainer, Nebojsa Jojic, Stephen Scherer, Benjamin Blencowe and Brendan Frey. The human splicing code reveals new insights into the genetic determinants of disease. (Highlight)

I missed this talk but got strong recommendation from Uri Keich: A Science paper showing breakthrough results in classification and feature selection using deep learning.

For all those interested in machine learning.

David Manescu and Uri Keich. A symmetric length-aware enrichment test

A very elegant piece of work that won the best paper award. The usual hypergeometric test for enrichment is modified to take care of gene length bias (longer genes will tend to be more identified as differentially expressed). The authors propose a DP alg and an approximation to do that. However, the authors note that this modified test is not symmetric anymore like the original HG test. They provide a symmetrized version of it. R code is available.

Relevant for Adi/Tom (for Expander) and all who use HG test.