This workshop was part of the special semester-long program that I am organizing at the Simons Institute for Theoretical Computer Science at UC Berkeley on Algorithmic Challenges in Genomics. It brought together over 100 scientists for five days of presentations and discussions. The program is available [here](#) along with links to video presentations of all talks. Many are worth watching!

The workshop was excellent and there were many outstanding talks. I will focus on a few only. A lot of the talks discussed and analyzed data from the big cancer projects, particularly TCGA, ICGC and PCAWG - a TCGA+ICGC consortium for "pan-cancer analysis of whole genomes".

**Niko Beerenwinkel** (Zurich) described two works on cancer evolution. Both are forthcoming in RECOMB 2016. In the first, the input is mutation information from single cancer cell reads. The goal is reconstruction of the phylogeny tree of cancer development under the perfect phylogeny model. The model (called SCITE) allows errors and missing data, and computes (approximate) maximum likelihood solution using MCMC on the space of trees. In the second, the goal is to find a partial order on the mutations using the principle of mutual exclusivity. A probabilistic model for the waiting times to mutations is assumed. The problem is solved using Hidden Conjunctive Bayesian Network. Both works are technically impressive and worth reading.

**Vineet Bafna** (UCSD) talked about amplifications in cancer. His talk was a mixture of review of his own computational work on the topic and experimental findings from the literature. He mentioned his elegant linear algorithm for BFB (Breakage-Fusion-Bridge) in PNAS '13, but admitted that in real data they found only three convincing examples of BFB. Regarding chromothripsis, he criticized the simulations done in the original paper of Stevens that described the phenomenon (the critique was published in NAR '14 and we discussed it in the group). He mentioned an interesting paper from Pullman's group in Nature '12, which described a phenomenon of creation of micronuclei due to "lagging chromosomes" in mitosis. The result is "pulverization of chromosomes in micronuclei", which may explain chromothripsis. Finally, he mentioned the episome / double minute model, where a genomic fragment is circularized into a independent mini-chromosome and amplified multiple times. He mentioned a fascinating paper from Nathanson et al. from Science '14, where application of a drug to glioblatoma samples leads to elimination of the double minute mini-chromosomes that contain the mutant EGFR gene. When the drug is withdrawn, the double minutes reappear. A "hide and seek" game between the cancer and the drug!

**Laura van 't Veer** (UCSF), the leading author of the famous paper on 70-gene Mammaprint fingerprint for breast cancer prognosis, talked about clinical studies of breast cancer. She described in detail the I-SPY TRIAL, a ground-breaking type of breast cancer clinical trial (phase 2) where rather than deciding on a fixed drug to be tested in advance and running the whole trial with it, a set of seven drugs are used, and results are continuously reevaluated over time using Bayesian statistic. At some point during the
trial a drug can be declared successful and moved out of the trial (and into a phase 3 trial), based on the evidence accumulated so far, and another drug can be introduced into the trial instead. She described a success story of a drug (veliparib/carbo, a PARP inhibitor) that passed this pipeline rapidly based on 72 patients, and is now in phase 3 trial. In this study the decision about the drug(s) to be used on each patient is based on on biomarkers of several types, including mutations, transcription levels, copy number, protein phosphorylation and tissue imaging.

Fabio Vandin (Padova) described methods for finding mutated subnetworks associated with survival in cancer. The input is a binary patient x genes matrix of mutations, and two survival vectors: one of continuous values, showing for each patient i its survival time \( t_i \), and the other binary, indicating if patient i died at time \( t_i \) or is surviving. They developed a log-rank type test for seeking a gene whose mutation is correlated with survival. Exact computation of the probabilities is needed, and they developed an improvement over extant tools. However, in tests on real data they did not find a significantly correlated gene. For that reason they moved to gene groups: Given a PPI network, they are looking for a connected subnetwork of size at most \( k \) that best correlates with survival. The problem is NP-hard (even under further restrictions) and they developed a color-coding algorithm. A problem is that their score is not additive over the genes. They show that under assumption that mutations are independent, with sufficiently many samples their method finds the optimal solution with high probability. The paper describing the algorithm, called NOMAS, will be presented in RECOMB 2016. This work is very elegant but biological relevance is unclear yet.

Josh Stuart (UCSC) described the recent TCGA efforts for tumor interpretation. He described their recent Tumor Map paper in *Cell '14*, where they analyzed multiple types of omic data from \(~3500\) samples from 12 cancer types. The main result was an integrated cross-tumor-type classification, with prognostic value. He described forthcoming unpublished work where 33 (!) cancer types are analyzed using a double number of samples (!). One of their conclusions is that one of five patients is classified to an unexpected cancer type, which does not match the tissue of origin. Among the tools that they use are the Paradigm algorithm from Haussler’s group and methods developed by Califano’s group to identify the active transcription factors and kinases in each tissue. They build a so-called scaffold network, which integrates mutations, gene expression and protein expression (and phosphoproteomics), and that network can be applied on the data in a single patient. This looks like an impressive effort that thinks of direct hospital application down the road.

Trey Ideker (UCSD) described the Cancer Cell Map Initiative. He argued that sequencing will not identify all the genes relevant to cancer in a particular patients, and in fact most known cancer genes to date were discovered in mice knock-outs and not by human sequencing. His idea is to use a collection of reference cancer cell maps to help infer from the patient’s genomic sequence its phenotype. His work (with Matan Hofree) in *Nat. Methods ’13* exemplifies how networks can improve the robustness of cancer patients clustering and prognosis. We’ll see if the initiative proves successful, unlike his previous attempts to create a map resource of the community.
Other impressive talks for those who wish to watch were given by Saurabh Shah (BC Cancer Research), David Haussler (UCSC), Dana Silverbush (TAU, Roded's group), Eytan Ruppin (TAU and Maryland), and Christina Curtis (Stanford).