ILANIT conference report – February 2017

Report by Tom Hait, March. 16 2017, reviewed by Ron

This meeting was held on 20-23 February 2017 at Eilat, Israel. FISEB, the Federation of Israel societies of experimental biology (whose Hebrew acronym is ILANIT), contains 31 Israeli biology societies and is held every three years in Eilat. The purpose of FISEB is to bring together researchers and students from different biology fields for sharing and proposing new research directions. The conference is divided into parallel sessions in different research fields and plenary lectures. Apart from the regular conference, a special session was dedicated to precision medicine (also known as personalized medicine), which is an area that combines big data analysis to propose specific individual treatments. Below I tried to cover lectures that included some computational analysis from different sessions.

Session: Frontiers in computational genomics and proteomics

Collective Action of MicroRNAs: Probabilistic and Dynamic Approaches

Michal Linial – HUJI biological chemistry

Michal developed a probabilistic tool, COMICS (competition on miRNA in cell simulation), to mimic the dynamics of the miRNA-mRNA network by capturing the effectiveness of the miRNA binding sites (MBSs) recognition at the 3'-UTRs. miRNAs are small ncRNA composed with ~22 nucleotides. There are ~1900 genes that code to ~2300 mature miRNAs, which regulate >60% of the mRNAs in stabilization, degradation, and translational repression. A single miRNA may target many sites in a gene’s 3'-UTR and there is a competition between miRNAs on MBSs.

COMICS is an iterative tool where in each step emulates the interaction of miRNAs to their MBSs on their target mRNAs. COMICS calculates the binding strength, accuracy and specificity of the miRNA’s target mRNA in each step and removes bounded miRNAs for the next step. The learning is based on experimental data from cell perturbations, e.g., by actinomycin D (ActD) treatment that inhibits transcription leading to a drop in mRNA expression levels but does not change the miRNA expression levels. This drop in mRNA levels may be due to miRNA regulation. Removal of bounded mRNAs will result with a subset of miRNAs that directly interact with the bounded mRNAs and they are over-expressed.

Deciphering Modes of Action of Long Noncoding RNAs

Igor Ulitsky – Weizmann institute of science

Igor presented different experimental methods for tackling long non-coding RNAs (lncRNAs) functionality. Recent estimates show that there may be 50,000 distinct lncRNA loci in the human genome, many of which are tissue specific and deregulated in human diseases.
Fundamental questions regarding lncRNAs still need to be addressed. Which lncRNAs are functional? How is the functional information encoded within the lncRNA sequence? What are the codes that keep lncRNAs to the nucleus and are responsible for lncRNA retention/degradation?

GFP RNA libraries were used to show that lncRNAs are localized in nucleus. Later Igor explained that nucleus lncRNAs possess SIRLOIN motifs shown to be correlated with nuclear localization by using eCLIP-Seq, which identifies protein-RNA binding sites. eCLIP-seq targeting RNA-protein complex called HnPNPK were used to identify lncRNAs with C-rich content, which is correlated with nuclear localization.

**Identifying Promoter-Enhancer Interactions from Hi-C Data**

**Tommy Kaplan –HUJI - paper**

Tommy presented a computational method, named PSYCHIC, to find the optimal segmentation of each chromosome into topological domains using Hi-C, by combining a probabilistic model and dynamic programming (Viterbi). PSYCHIC iteratively merges neighboring domains into hierarchical structures and fits a bi-linear power law model to create a background model for each merged domain of the Hi-C map, allowing the identification of over-represented DNA-DNA pairs, including putative promoter-enhancer interactions.

The segmentation into domains is affected by functional and non-functional Hi-C interactions and noise. The noise is estimated and removed by fitting a background model. The remaining DNA-DNA interactions are divided into windows of 2Mb surrounding each promoter. Functional interactions are defined as highly enriched interactions with the promoter. The statistical enrichment score is calculated by modeling the DNA-DNA interactions within each 2Mb window using the Normal distribution, and a P-value is derived for all regions interacting with a promoter, followed by FDR correction for multiple hypotheses.

Validation of the method was performed on 12,278 putative enhancer-promoter interactions (FDR<1e-2) from the mouse cortex. It was shown that the identified interactions manifest higher coverage of epigenetic marks of active enhancers and promoters (obtained using ChIP-Seq) compared to a random set of genomic loci within a window of 2Mb around each promoter (see Figure bellow taken from their paper):
Relevant to: Tom, Idan and Nimrod

**Calling Mutation from Bulk and Single-Cell RNA-seq Data for Studying Cancer Initiation, Progression and Heterogeneity**

**Keren Yizhak**, Cancer program, Broad institute of Harvard and MIT, USA

Keren developed multiple computational tools for identifying mutation events in a clonal evolution by expanding mutation detection to multiple tissues and individuals. Somatic mutations that occur in only a subset of the sequenced cells, either from tumor heterogeneity or contamination by normal cells, may be missed due to their low-allelic fraction. Gadi Getz's group (in which Keren works) developed the [MuTect tool](#) that applies a Bayesian classifier to detect somatic mutations with very low allele frequencies supported by low RNA-Seq read count. Using MuTect they were able to filter out 99% of the FPs variants detected in RNA, while correctly identifying mutations in major driver genes.
Keren utilized MuTect for studying genetic alterations during cancer initiation and progression via two different datasets: (1) The GTEx database, which includes RNA-Seq collected from 30 normal tissues across ~600 individuals. Analysis of GTEx DB revealed multiple cancerous mutations in normal tissues indicating initial cancer events. (2) Single cell (SC) RNA-Seq from different cancer types to identify genetic variations across different cells of a single tumor. Keren compared DNA-seq (normal tissues) vs. the single cell RNA-seq (cancer) and showed that more variants detected in RNA-seq may be due to noisy sites, and that there is a limited coverage in SC experiments. Using MuTect, Keren showed that the removal of 99% of the noisy sites had a small effect on detecting the true positive mutations.

Keren plans to further expand MuTect to include somatic mutation identification using available CNV data by counting the number of alternative alleles versus the normal allele.

Relevant to: Ron Zeira and Nimrod

Session: Protein interaction networks

Network-Based Approaches toward the Enigma of Tissue-Specific Hereditary Diseases

Esti Yeger-Lotem, Ben-Gurion University - paper

Esti asked what makes hereditary diseases tissue-specific. For example, mutations in BRCA1 lead to breast and ovarian cancers while BRCA1 is expressed ubiquitously throughout the body without harming other tissues. Esti integrated RNA-Seq with PPI datasets to generate 16 tissue specific networks covering over 300 hereditary diseases. Esti mentioned that only 6% of the diseases have known causal tissue-specific genes.

To form a tissue-specific PPI network, she uses the tissue expression and includes in the network only PPIs between genes that are highly expressed in that specific tissue. It was mentioned that causal genes with many paralogs are less involved in tissue-specific diseases than those with few paralogs. Esti showed that there is high fold change (FC) in RPKM between causal genes and paralogs in muscle related diseases. Causal genes are up-regulated in muscle while their paralogs from other tissues are down regulated. In addition, 80% of the chaperones are highly expressed in all tissues but their interacting protein partners, according to expression values, are different and tissue-specific. A chaperon network can reveal disease specific associations. Expression data was taken from known DBs such as GTEx and The human protein atlas.

Examples for such tissue-specific tools developed in Esti’s lab include: MyProteinNet and TissueNet.

Relevant to: Gal and Nimrod
Session: Precision Medicine

Translating a trillion points of data therapies and new insights in health and disease

Atul Butte, University of California, San Francisco

Atul slides (at least most of them) are available at: http://www.aacp.org/meetingsandevents/AM/2014/Documents/AM14SciencePlenaryAtulButteSlides.pdf

Atul is the inaugural Director of the Institute of Computational Health Sciences at the University of California, San Francisco, and a Distinguished Professor of Pediatrics.

Atul started by explaining that currently there are ~2M microarrays publicly available for use in ArrayExpress database and there are 184 ‘breast cancer’ GEO datasets covering 76,256 samples and 49 platforms. All of these public big data can be now used to develop algorithms on the basis of retroactive crowd-sourcing. Atul’s definition for big data in biomedicine is:

Big Data in Biomedicine is...

Predicting the disease before it strikes
Explaining the rare disease that defies experts
Finding drugs for diseases lacking attention
Making sure we do the right thing for patients
An amazing platform for biomedical innovation

Atul sketched out meta-analysis projects done in his lab. The first project was on identifying serum/urine protein biomarkers whose genes are differentially expressed in biopsy-based microarrays from >16,000 biological conditions covering 41 diseases available in GEO. Deregulated genes with known detectable proteins in serum or urine marked those proteins as candidate biomarkers for disease diagnostic (paper). The second project (paper) talked about the use of established drug compounds for new therapeutic indications (also known as drug repositioning). The idea was to correlate molecular signatures in drug-disease pairs. They integrated gene expression microarray measurements from 100 diseases and gene expression measurements on 164 drug compounds. The idea was to compute a fold change (FC)-like gene signature across the samples between cases and controls in disease and between drug treated and non-treated samples (see figure A below). A connectivity map (paper) is used to create a reference database of drug gene expression (see figure B below). They computed a similarity score for each gene disease-drug pair ranging from +1 indicating perfect correlation of signatures and -1 indicating the opposite signatures. The hypothesis was negative similarity score should indicate a drug that has a potential therapeutic effect on the paired disease.
Another project discussed was on Preeclampsia (PE), which is a pregnancy-related vascular disorder and the leading cause of maternal and fetal death. The project aimed to identify serological protein markers to diagnose PE by performing multi-omics comparisons. 266 expression arrays (n=111 PE and n=152 control placenta samples) collected from GEO were used to identify significant fold changes in genes whose protein products are detectable in the blood serum. Two-dimensional gel profiling (pooled n = 5 PE and pooled n = 5 control serum proteomes) was used to identify candidate serum proteins for PE. ELISA test was performed on independent PE (n=32) and control (n=32) cohorts to identify both proteomic and genomic serum-protein candidates. Demographic information on the cohorts was combined with the ELISA results to perform feature selection on PE candidate serum-proteins.

Other projects that were mentioned are: A work of Euan Ashley, Russ Altman, and Steve Quake to identify patient pedigree by using nano-pore technique, a work of Rong Chen focusing on disease-SNP association studies from 2,113 publications, which showed that non-synonymous and synonymous coding SNPs shared similar likelihood and effect size for disease associations. This led to the construction of VARIMED (VARIants Informing MEDicine) database.

Relevant to: Didi, Gal and Nimrod
Precision cancer medicine down of the genomics and immunotherapy era

Razelle Kurzrock, UCSD Moores Cancer

Razelle is the Chief of the Division of Hematology and Oncology in the UC San Diego School of Medicine.

Note – it was not easy to follow Razelle, so some things below may be misleading.

Razelle slides (at least most of them) are available at:

Razelle first started by explaining what defines Precision Medicine in these days: (1) The study of the connection between genomics and immunotherapy, (2) Finding genomic characteristics that make different tumors distinct and complex, (3) the study of new cutting edge technologies including Tumor Mutational Burden (TMB) and liquid biopsy, and (4) super-responders in immunotherapy.

Super-responders are set of immune biomarkers that treatment against them will achieve high respond rate (i.e., the patient will get better). PD-L1 is a known immune protein that suppresses the immune system during particular events such as pregnancy, autoimmune diseases and cancer. Samples with positive indication of PD-L1 occurrence tend to achieve higher response rates to immunotherapy (36-100% chance of curing by immunotherapy). However, there is a high variability in PD-L1 protein expression levels tested via immuno-histochemistry staining, which makes it difficult to decide to take immunotherapy actions based on a single sample. TMB is a more robust method that measures the frequency of mutations in a tumor genome. High TMB values are believed to express more neoantigens (a type of cancer-specific antigen) that may increase response rate to immunotherapy (e.g., 8 samples out of 12 with TMB>50 had respond rate of 67%).

The above strategies led to creating a protocol called PREDICT (Profile-related Evidence Determining individualized Cancer Therapy) that decides a treatment for individual based on super-responders, TMB, and liquid biopsy. Liquid biopsy was performed on 1500 patients to extract circulating-tumor DNA (ctDNA) from the blood and hyper-mutated genes were identified from the ctDNAs. Hyper-mutated ctDNA genes were associated with high response rate to immunotherapy.

Relevant to: to anyone who seek general knowledge on immunotherapy

Playing chess against cancer and resistance to drugs

Yosef Yarden, Weizmann Institute of Science

Yosef was recently announced as the winner of the Israel Prize (2017).

Yosef focused on cancer treatment via drug perturbation of networks responsible for cancer initiation, progression and development. Cancer networks are robust to single drug perturbation (mono therapy) but not for two or more drug perturbations. Hence, the common strategies are to use a combination of treatments using kinase inhibitors and antibodies to perturb many edges in a cancer network.

Homo or hetero combinations of antibodies and drugs targeting kinases may be the best practice to treat cancer by perturbing its network. As an example, Yarden described the epidermal growth factor
receptor (EGFR) and the co-receptor denoted HER2/ERBB2 as receptors that are frequently overexpressed or mutated in solid tumors (paper). Drugs targeting EGFR are highly effective in cancer treatment, however if the EGFR is mutated then the cancer cell remains resistant to EGFR kinase inhibitors and antibodies (paper). In addition, down-regulating EGFR can cause to HER2/3 to be up-regulated and therefore the cancer signal transduction network is not entirely blocked. Yarden’s strategy is to develop therapies by discovering mechanisms of acquired resistance to kinase inhibitors. For example, treatment with Imatinib drug that targets the BCR-ABL complex related to cancer is conferred by high amplification of the BCR-ABL altered genes in chronic myeloid leukemia (CML). Second generations of BCR-ABL inhibitors were developed based on identifying the genes’ mutations that amplified the BCR-ABL complex. In the bottom line, a combination of kinase inhibitors and drugs targeting these acquired mechanisms may treat cancer diseases (paper).

Relevant to: Gal
**Session: COMPUTATIONAL BIOLOGY AND IMMUNOTHERAPY: FROM DRUG DESIGN TO THERAPY**

**A Roadmap for Cancer Immunotherapy**

*Gal Markel, Sheba Medical Center & TAU*

One of the most promising approaches for treating various types of cancer is based on manipulating the patient's own immune system to improve its ability to identify and attack cancerous cells. Different tumors exhibit varying level of immunogenicity, which is defined as the tendency of the tumor to provoke an immune response by the patient's body. As tumors develop under continuous pressure of immune cells, they gradually may decrease their level of immunogenicity, thus reaching a certain status of equilibrium with the immune system.

Melanoma is an example for a cancer that invokes an immune response in the patient's body, but the response is usually inefficient and thus the immune system by itself cannot eliminate the tumor.

Anti-CTLA-A and anti-PD-1 are two immune modulators that work on distinct pathways to stimulate the immune response in metastatic melanoma patients. They may however trigger also autoimmune response. A newer immune modulator named anti-CEACAM1 is expected to increase a response specific to the tumor cells while evoking low autoimmunity.

A newer approach uses genetically modified T Cells whose specificity against the target tumor cells is increased by creating a chimeric antigen receptors (CARs). This approach relies on technological advancements in genetical modification methods, and is much faster than isolating existing T-cells, proliferating them in-vitro and reintroducing them into the patient's body.

**Session: PROTEOMIC PROFILING OF CANCER**

**Proteomic Landscapes of Breast Cancer Subtypes**

*Tamar Geiger, TAU*

Modern mass spectrometry methods can produce a high-resolution and highly-accurate proteomic profiling of a given tissue. A typical proteomic profile is composed of more than 10,000 features describing the abundance levels of various proteins. These data can be used similarly to other high-throughput genomic technologies to identify disease subtypes and also to shed light on key proteins that characterize each subtype. Since proteins are the end point of the transcription/translation process, and since their dynamic range is much higher than mRNAs for instance, protein profiling may prove even more valuable to these goals than more traditional data types like gene expression. Analysis at the protein level may more directly reflect cellular functions.
Three breast cancer patient cohorts [40 (Tyanova)+65 (Hadassa)+88 (Pozniak – Sheba)] were proteome profiled by mass spectrometry and analyzed. Clustering analysis partitioned the samples of the first 40 sample cohort into distinct subgroups. The HER2 protein was highly expressed on the Her2 subtype. When comparing genomic, transcriptomic and proteomic profiles for these samples – some changes between the subtypes did not appear on the genomic level but did appear on both transcriptome and proteome level. SVM classifier was trained to predict the subtype assignment based on the proteomic profile.

Joint analysis of the three sample cohorts enabled the partitioning of the samples into four clusters – a clear basal cluster and three luminal clusters. The Her2 samples were not clustered into a distinct cluster. Supervised analysis identified differentially expressed proteins, enriched for functions such as energy metabolism and cell growth.
Session: Evolution and Systems Microbiology

Sugar synthesis from CO2 in E. coli

Ron Milo - Plant & Environmental Sciences, Weizmann institute

Most bacteria cannot fixate CO2 - that is, create organic compounds from CO2, like plants can. Engineering bacteria to fixate carbon can have future applications in reducing the CO2 level in the atmosphere.

In the work Ron presented, they engineered E. Coli to synthesize sugar from CO2. Most of the enzymes required for carbon fixation were already present in E. Coli and missing enzymes (such as RuBisCO) were added.

In addition, a gene was deleted to separate the carbon flow into different modules: the fixation module and the energy module, which should create energy for the fixation module from organic compounds.

E Coli produced by this approach of cellular pathways design could not fixate carbon, and the group decided to adopt an evolutionary approach to the problem. Instead of further engineering the strain, it was grown in a medium with gradually decreasing amount of xylose, which is a sugar carbon source. Within this medium, mutants capable of CO2 carbon fixation will be under strong positive selection. This evolutionary approach worked, and a few mutants emerged. By growing the strain in the presence of isotopically labeled CO2, mass spectrometry confirmed that the E. Coli used carbon from CO2. A single gene, PRS, was mutated in all mutants able to synthesize sugar. A flux based model they developed suggests PRS fine tuning is required for metabolic stability.

Session: Model Organisms and their Contribution to Human Genetics

A Platform for Rapid Exploration of Aging and Diseases in a Naturally Short-Lived Vertebrate

Itamar Harel - Postdoc in Anne Brunet's lab in Stanford

Aging research is mostly done on model organisms such as C. Elegans, Drosophila, Mouse and Zebrafish. The former two models lack many characteristics of higher eukaryots, such as blood and bones. The latter two have more complex systems, but a much increased lifespan - 2 years for the mouse and 5 for zebrafish, making aging research more costly and difficult.

A new emerging model for aging is the killifish (picture ->), which is the shortest lived vertebrate that can be raised in captivity. Its lifespan is 4-6 months, and it exhibits all
phenotypes of human aging. In addition, they are cheaper than mice. Itamar worked on establishing methods required for model animals: mainly, creating a method to perform genome engineering using CRISPR-Cas9. He also sequenced the fish and created a map of its genome, identifying potential aging and disease related genes. Using genome engineering methods, a telomerase deficient fish was created that indeed displayed phenotypes resembling human phenotypes for telomerase deficiency.

The future plan for the killifish model is to perform vertebrate screens for genes involved in aging for the first time, and use the model to investigate aspects of aging not present in models such as C Elegans and Drosophila: aging of systems (such as the immune system) and organs.

Note: just yesterday (as of writing this), a killifish experiment was published in which microbiome from young killifish was transplanted into older killifish and extended their lifespan showed a causative role for microbiome in vertebrates. The experiment was performed by Valenzano's lab in Cologne, who was a postdoc at the same lab as Itamar: see http://biorxiv.org/content/early/2017/03/27/120980.