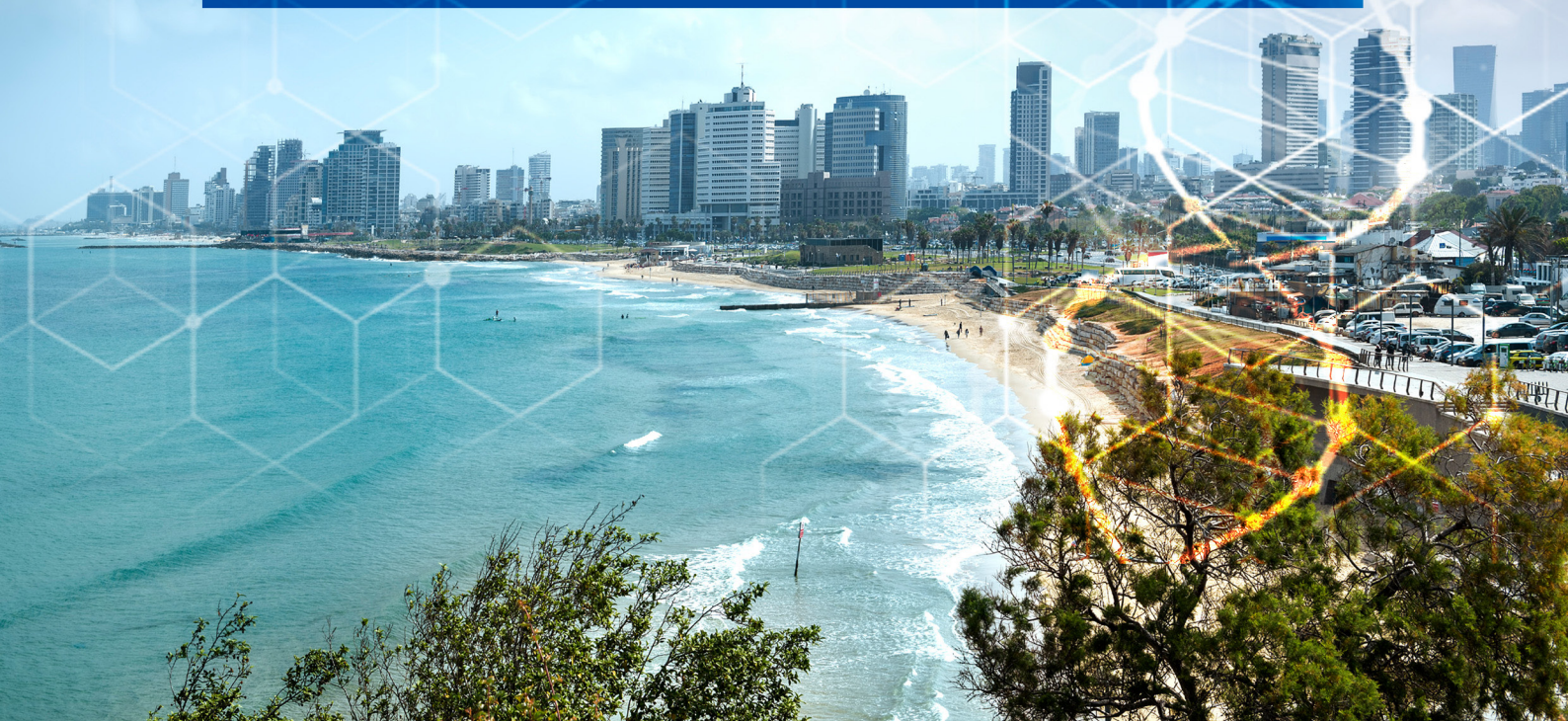
hgm 2022 HUMAN GENOME MEETING Tel Aviv, Israel

The 25th Human Genome Meeting

The Next Frontiers in Genomics

May 23-25, 2022| Dan Tel Aviv Hotel, Israel

Book of Abstracts



Thank you to our sponsors and partners

Our Sponsors:

DanyelBIOTECH at the heart of action

Syntezza

hylabs

teva

almog diagnostic

EISENBERG

mission bio

אלדן קבוצת ניאופרם

BIO-RAD

MGI

The Jackson Laboratory
Leading the search for tomorrow's cures

RHENIUM

Our Partners:

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

BioMed@TAU

The Chaim Sheba Medical Center at Tel HaShomer
ISRAEL'S CITY OF HEALTH

המרכז הרפואי שערי צדק
SHAARE ZEDEK MEDICAL CENTER

Yad Hanadiv יד הנדיב יָד هَنَدِيف

אוניברסיטת בר-אילן
Bar-Ilan University

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

ScienceAbroad
The Organization of Israeli Scientists Abroad

Organized by

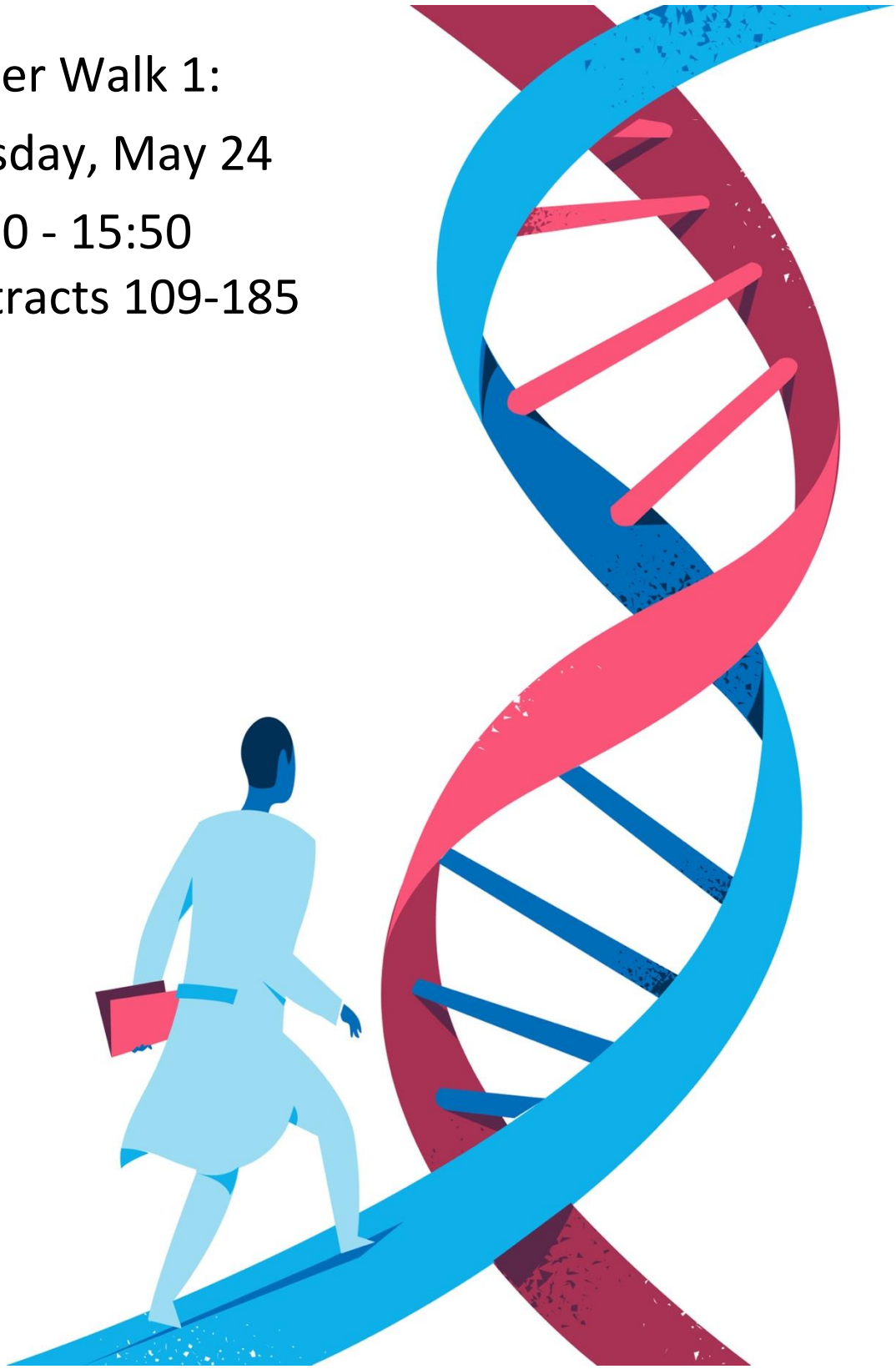
info@ortra.com ORTRA MEETING POINT 03-6384444

Poster Walk 1:

Tuesday, May 24

14:50 - 15:50

Abstracts 109-185



109

Elevated plasma levels of CXCL16 in severe COVID19 patients - significant association with mortality

Sandra Smieszek, Christos Polymeropoulos, Mihael Polymeropoulos

Genetics, Vanda Pharmaceuticals, USA

Genome-wide association studies have recently identified 3p21.31, with lead variant pointing to the CXCR6 gene, as the strongest thus far reported susceptibility risk locus for severe manifestation of COVID-19. In order the determine its role, we measured plasma levels of Chemokine (CXC motif) ligand 16 (CXCL16) in the plasma of COVID-19 hospitalized patients. CXCL16 interacts with CXCR6 promoting chemotaxis or cell adhesion. The CXCR6/CXCL16 axis mediates homing of T cells to the lungs in disease and hyper-expression is associated with localized cellular injury.

To characterize the CXCR6/CXCL16 axis in the pathogenesis of severe COVID-19, plasma concentrations of CXCL16 collected at baseline from 115 hospitalized COVID-19 patients participating in ODYSSEY COVID-19 clinical trial were assessed together with a set of controls. We report elevated levels of CXCL16 in a cohort of COVID-19 hospitalized patients. Specifically, we report significant elevation of CXCL16 plasma levels in association with severity of COVID-19 (as defined by WHO scale) (P-value0.02). We replicate this finding in an independent replication set CALYPSO (P-value0.0012). We also observe a highly significant effect on mortality (P-value0.0004) in association with higher CXCL16 plasma levels. We are further characterizing the expression of CXCR6 in CD8+ T cells.

Our current study is the largest thus far study reporting CXCL16 levels in COVID-19 hospitalized patients (with whole-genome sequencing data available). The latest results support the significant role of the CXCR6/CXCL16 axis in the immunopathogenesis of severe COVID-19 and warrant further studies to understand which patients would benefit most from targeted treatments.

114

An international and multi-ancestral genome-wide association study meta-analysis of cannabis use disorders

Daniel Levey[1,2], Marco Galimberti[1], Frank Wendt[1], Emma Johnson[3], Megan Cooke[4], Veera Rajagopal[5], Joseph Deak[1], Arjun Bhattacharya[6], Hang Zhou[1], Max Lam[7], Yaira Nunez[1], Howard Edenberg[8], Todd Lencz[7], Jordan Smoller[4,12], Arpana Agrawal[3], Ditte Demontis[5], Michael Gandal[6], J Michael Gaziano[11], Renato Polimanti[1], Murray Stein[9,10], Joel Gelernter[1]

[1]*Psychiatry, Yale University, USA*; [2]*Research, VA CONNECTICUT HEALTHCARE SYSTEM, USA*; [3]*Psychiatry, Washington University School of Medicine, USA*; [4]*Psychiatry, Massachusetts General Hospital, USA*; [5]*IPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Denmark*; [6]*Psychiatry, University of California Los Angeles, USA*; [7]*Institute for Behavioral Science, The Feinstein Institutes for Medical Research, USA*; [8]*Medical and Molecular Genetics, Indiana University, USA*; [9]*Psychiatry and Public Health, University of California San Diego, USA*; [10]*Psychiatry, VA San Diego Healthcare System, USA*; [11]*MAVERIC, VA Boston Healthcare System, USA*; [12]*Psychiatry, Harvard Medical School, USA*

Background Cannabis use is becoming more pervasive as access has increased due to changes in legal status of the drug worldwide. It is crucially important to understand genetic risk factors underlying cannabis use disorder (CUD) that may have an impact on public health in light of the changing legal status of cannabis and treatment needs.

Methods We performed a genome-wide association study in MVP and meta-analysis of CUD in more than 1 million individuals from four genetically-determined ancestries (European, EUR; African, AFR; Admixed-American, AMR; East Asian, EAS). Covariate-adjusted LD score regression (covLDSC) was applied to calculate heritability in each ancestry. Downstream in-silico analyses were performed to make causal inferences implied by genetic association, genetic correlations with related traits, and functional gene enrichments for cannabis use disorders. MultiXcan was used to integrate genetic association with information regarding expression quantitative trait loci in brain and blood tissues.

Results We discovered genomewide-significant loci in all 4 ancestries: 22 in EUR (lead SNP rs56372821 p=7.24e-14, eQTL for CHRNA2), 2 each in AFR (lead SNP rs574008891 p=2.68e-8, in MCCC2) and EAS (lead SNP rs78561048 p=6.71e-9, in SEMA6D), and 1 in AMR (lead SNP rs9815757 p=4.36e-8). We show significant heritability: EUR h2=8.5% (SE=0.0045), AFR h2=12.7% (SE=0.0276), AMR h2=16.8% (SE=0.0653). MultiXcan implicated several genes including upregulation of bassoon presynaptic cytomatrix protein (BSN) expression with strongest evidence in cortex (p=9.10e-10).

Discussion We extended previous CUD studies, greatly increasing the number of discovered risk loci in populations of European descent and uncovering the first genome-wide significant associations in other ancestry groups.

115

A single amino acid in RTEL1 determines telomere length in mice

Riham Smoom[1], Catherine Lee May[2], Mark Tigue[2], Hannah Kolev[2], Ashleigh Morgan[2], Nachshon Egyes[1], Dan Lichtental[1], Klaus H. Kaestner[2], Yehuda Tzfati[1]

[1]*Genetics, Hebrew University, Israel*; [2]*Genetics and Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania, USA*

The large difference in telomere length between *Mus musculus* and *M. spretus* was previously exploited to identify the helicase RTEL1 as a dominant regulator of telomere length1. However, the underlying difference between the two RTEL1 proteins has not been found2. We identified germline mutations in human RTEL1, which cause Hoyeraal-Hreidarsson syndrome3. One of the mutated amino acids, methionine 492, is conserved across vertebrates except for a lysine in *M. spretus*, suggesting that this change is responsible for the shorter spretus telomeres3. To test this hypothesis, we generated a M. musculus strain in which M492 was changed to lysine using CRISPR-Cas9 genome editing. Interestingly, the Rtel1M492K mice telomere length set point was gradually shortened 2-3 times, reaching generation thirteen by now, and the mice are looking fertile and healthy. Additionally, we established Rtel1M492K mouse embryonic fibroblasts and followed their growth and telomere phenotypes in culture. Telomeres shortened gradually over 250 population doublings to an average length of ~15kb (WT MEFs ~40kb). The finding that a single amino acid change is responsible for the telomere length difference between the two species provides an important insight into the mechanism of telomere length regulation. Furthermore, the healthy Rtel1M492K mouse with short telomeres represents a novel and invaluable model for studying the implications of short telomeres in aging and cancer.

116

CTCF and active transcription support short lamina-associated domains

Nehora Levi[1], Mali Salmon-Divon[1,2], Gabi Gerlitz[1]

[1]*Department of Molecular Biology, Ariel University, Israel*; [2]*Adelson School of Medicine, Ariel University, Israel*

Nuclear architecture affects all aspects of genome maintenance and functions. The nuclear lamina associates with large parts of the genome that are termed lamina-associated domains (LADs). Although LADs have been associated with genome organization and gene repression, it is not clear how LADs are established, and their borders maintained. In mammals, sharp borders of LADs are enriched with active promoters and CTCF-binding sites. CTCF is a multifunctional protein involved in genome organization, transcription, and insulation. Here, to test if CTCF supports the LADs landscape we generated melanoma cells with a partial loss of function (pLoF) of CTCF by the CRISPR-Cas9 system and determined the LADs landscape by Lamin B ChIP-seq analysis. We found that CTCF pLoF led to moderate changes in the LADs landscape that included increased signals in 2% of the LADs decreased signals in 8% of them. However, CTCF importance for the LADs landscape enhanced upon induction of chromatin stress by inhibiting RNA polymerase II – an intervention that is known to alter chromatin compaction and supercoiling. Notably, only in CTCF pLoF cells chromatin stress led to the dissociation of 7% of the LADs from the lamina. The CTCF-dependent LADs had almost three times shorter median length than the non-affected LADs, were enriched in CTCF binding at their borders and were higher in their facultative-status (cell-type specific). Thus, it appears that CTCF is a key factor in facilitating the association of short facultative LADs with the nuclear lamina upon chromatin stress.

118

Identification of two heterozygous mutations in a cryptozoospermic patient and a de-novo mutation in a boy with testicular failure born by IVF-ICSI

Florence Abou[1,2], Eliahu Levitas[3,4], Atif Zeadna[3,4], Iris Har-Vardi[3,4], David Shaki[5], Neta Leventhal[5], Eitan Lunenfeld[6], Ruti Parvari[1,2]

[1]*The Shraga Segal Department of Microbiology, Immunology & Genetics Faculty of Health Sciences (A.B.), Ben-Gurion University of the Negev, Israel*; [2]*National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel*; [3]*Fertility and IVF Unit, Department of Obstetrics and Gynecology, Soroka University Medical Center, Israel*; [4]*Faculty of Health Sciences, Ben-Gurion University of the Negev, Israel*; [5]*Pediatric Endocrinology Unit, Soroka University Medical Center, Israel*; [6]*School of Medicine, Ariel University, Israel*

Introduction: About half of the unexplained Cryptozoospermia and underdeveloped male traits are caused by genetic defects.

Objective: Identify the genetic causes of an adult male`s Cryptozoospermia and a child with abnormal reproductive system, born following Intra Cytoplasmic Sperm Injection (ICSI) due to his father`s severe oligozoospermia.

Methods: DNA was extracted from the patient`s blood and analyzed by Whole Exome Sequencing using bioinformatic tools. Sanger sequencing validated the WES findings.

Results: In the adult patient, we identified a novel heterozygote frameshift mutation in the SYCP2 gene, which is essential role in meiosis. Previous studies showed that heterozygote mutations in this gene could cause low sperm count. Additionally, we detected two heterozygote frameshift mutations in ADAM20. ADAM20 encodes a membrane protein involved in sperm-oocyte fusion. Since the couple underwent an ICSI procedure, the mutations in ADAM20 had no effect. Pregnancy was achieved, and a healthy baby girl was born.

After birth, the boy presented as micropenis and undescended testis. Bilateral orchiopexy preformed at the age of 2 revealed a small undeveloped testis. After testosterone replacement therapy, at 15 years old, he has micropenis and descendent testis. We identified a known heterozygous missense mutation in the DHX37 gene, encoding a DEAH-box RNA helicase previously reported to result in this phenotype. The parent`s DNA availability enabled us to determine that the mutation appeared de-novo since it was absent in either parent.

Conclusion: As probably in many idiopathic male infertility cases, genetic testing could identify the causes of the disruption in patients and lead to better fertility counseling.

123

Immune system cells from COVID-19 patients display compromised mitochondrial-nuclear expression co-regulation and rewiring toward glycolysis

Hadar Medini, Amit Zirman, Dan Mishmar

Life Science, Ben Gurion University of the Negev, Israel

Mitochondria are pivotal for bioenergetics, as well as in cellular response to viral infections. Nevertheless, their role in COVID-19 was largely overlooked. Here, we analyzed available bulk RNA-seq datasets from COVID-19 patients and corresponding healthy controls (three blood datasets, N = 48 healthy, 119 patients; two respiratory tract datasets, N = 157 healthy, 524 patients). We found significantly reduced mtDNA gene expression in blood, but not in respiratory tract samples from patients. Next, analysis of eight single-cells RNA-seq datasets from peripheral blood mononuclear cells, nasopharyngeal samples, and Bronchoalveolar lavage fluid (N = 1,192,243 cells), revealed significantly reduced mtDNA gene expression especially in immune system cells from patients. This is associated with elevated expression of nuclear DNA-encoded OXPHOS subunits, suggesting compromised mitochondrial-nuclear co-regulation. This, together with elevated expression of ROS-response genes and glycolysis enzymes in patients, suggest rewiring toward glycolysis, thus generating beneficial conditions for SARS-CoV-2 replication. Our findings underline the centrality of mitochondrial dysfunction in COVID-19.

126

Clustering of clinical-echocardiographic phenotypes of COVID-19 disease using machine-learning techniques

Eran Shpigelman*[1], Aviram Hochstadt*[2], Dan Coster[1], Ilan Merdler[2], Yan Topilsky[2], Ron Shamir[1]

[1]*Blavatnik School of Computer Science, Tel Aviv University, Israel*; [2]*Department of Cardiology, Tel Aviv Sourasky Medical Center, Israel*

Background. Covid-19 patients who deteriorate usually suffer from impaired lung function, but there is a spectrum of cardiac manifestations in patients and cardiac involvement in the infection is heterogeneous. Machine learning based discovery of subtypes of a clinical entity has been previously used in cardiology to define sub-phenotypes of common diseases. We sought to use an unsupervised machine learning technique to divide hospitalized COVID-19 patients into identifiable phenotypical clusters in order to shed light on the pathogenesis of the disease.

Methods. 506 patients hospitalized with COVID-19 infection underwent complete evaluation including echocardiography and lung ultrasound at admission. All patients' clinical and imaging data at admission was used to detect clusters. The data was entered into a k-prototype clustering algorithm in order to divide the patients into an optimal number of clusters.

Results. Patients were partitioned into four phenotypical clusters. Clusters 1 and 2 were younger with less past diseases, 3 and 4 were older with worse cardiac indexes. Clusters 2 and 4 had a more robust inflammatory response - a more severe Covid-19, with a higher propensity for respiratory and hemodynamic support. Survival was best for cluster 1, worst for cluster 4 and intermediate for clusters 2 and 3, which implies that the course of the disease is not entirely dominated by age. The echocardiography measurements played a considerable role in cluster formation.

Conclusion: COVID-19 manifests differently for distinctive clusters of patients. These clusters have an influence of disease manifestation and prognosis.

129

Genomic analysis of the spatial organization of the genome and its effect on cell-type specific p53 transcriptional responses

Hadar Amira-Haham[1], Gony Shanel[2], Tsung-Han S. Hsieh[3], Xavier Darzacq[3], Ron Shamir[1], Ran Elkon[2]

[1]*The Blavatnik School of Computer Science, Tel Aviv University, Israel*; [2]*Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Israel*; [3]*Department of Molecular and Cell Biology, University of California, Berkeley, USA*

Many studies have observed that transcriptional responses to multiple stresses are highly cell-type specific. However, the mechanisms that underlie this tissue specificity remain largely elusive. In our study we focus, as a model system, on the transcriptional network activated by p53, which serves as a pivotal defence mechanism against cancer transformation.

p53 activation in different cell types results in induction of very different transcriptional networks, alongside an activation of a universal p53 core response. Our main goal is to elucidate factors that determine cell-type specific responses.

We measured and analyzed three layers of data: RNA-Seq, ChIP-Seq and Micro-C (a high-resolution variant of HiC), in ten different cell lines. Measurements were taken both in control cells and in cells treated by Nutlin-3, a potent p53 activator.

Our analysis identified dozens of cell-type specific p53 chromatin-binding events that correlated with cell-type specific p53-induced gene expression. We are currently testing to what extent the cell-type specific p53-induced responses that we observed in our datasets are associated with cell-type specific features of the spatial organization of the genome.

131

CTCF supports melanoma cell migration by repressing cholesterol biosynthesis

Nehora Levi[1], Lukasz S. Kaczmarczyk[1], Mali Salmon-Divon[1,2], **Gabi Gerlitz**[1]

[1]*Molecular Biology, Ariel University, Israel*; [2]*Adelson School of Medicine, Ariel University, Israel*

CCCTC-binding factor (CTCF) is a zinc finger protein that regulates correct three-dimensional chromatin folding in addition to transcriptional activation and repression and splicing. Cellular CTCF levels were found to affect tumor cell migration, however the molecular mechanism is still obscure. To identify the molecular mechanism by which CTCF supports melanoma cell migration we used mouse melanoma cells with a partial loss of function (pLoF) of CTCF generated by the CRISPR-Cas9 system. Here we show that CTCF pLoF inhibited cell migration rate. Using RNA sequencing we identified an increase in the expression of multiple enzymes in the cholesterol biosynthesis pathway along with an increase in the cellular cholesterol levels in CTCF pLoF cells. Inhibition of cholesterol synthesis in CTCF pLoF cells by Fatostatin restored the cellular migration rate, suggesting that CTCF supports cell migration by inhibition of cholesterol synthesis. Detailed analysis of the promoter of the cholesterol synthesis enzyme Hmgcs1 revealed that CTCF pLoF led to reduced H3K27me3 levels in parallel to increased binding of the master activator of sterol synthesis SREBP2. Moreover, inhibition of H3K27 methylation in wildtype cells led to increased SREBP2 binding to this promoter.

Taken together, our results suggest that CTCF represses transcription of cholesterol biosynthesis enzymes by promoting H3K27 methylation at their promoters to suppress SREBP2 binding. By that, CTCF fine tune the cellular cholesterol levels to support cell migration.

134

Big data to therapy: Precision medicine for the deaf

Zippora Brownstein[1], Tal Koffler-Brill[1], Lara Kamal[1], Yazeed Zoabi[2], Rachel Katz[3], Noam Shomron[2], Tal Patalon[3], Karen B. Avraham[1]

[1]*Department of Human Molecular Genetics & Biochemistry, Tel Aviv University, Israel*; [2]*Department of Cell & Developmental Biology, Tel Aviv University, Israel*; [3]*KSM-Kahn-Sagol-Maccabi Research and Innovation Center, Maccabi Healthcare Services, Israel*

Precision medicine has become the optimal direction for treating human disease, as well as ensuring health for individuals of all ages. The field of deafness is optimal for implementing precision medicine, with over 150 responsible genes known worldwide. Phenotype-genotype correlations are reflected among the diverse Israeli Jewish ethnic groups. While next-generation sequencing (NGS) has rapidly advanced gene discovery, about half of inherited deafness still remains unsolved. With the availability of electronic medical records (EMR) for deaf individuals, data of large numbers of patients is available for improving personalized treatment. We are conducting a large-scale study of the hearing-impaired population in Israel, utilizing KSM (Kahn-Sagol-Maccabi) TipaBiobank of 60,635 anonymized samples. NGS is being performed on 1189 adult deaf from the Biobank, with audiograms documenting their phenotype, and the hearing loss diagnosed before the age of 60. Following variant detection by bioinformatics mega-analysis, pathogenic variants are being evaluated. New variants will be applied in audiology and genetics clinics, setting a guideline for precision genetic counseling and personalized audiological treatment and rehabilitation for deafness in Israel. In parallel, novel variants are being functionally investigated by advanced experiments of epigenomics, CRISPR/Cas9 gene editing, and gene therapy. Solving the etiology of deafness in the diverse Israeli Jewish population in a large-scale study and finding genotype-phenotype correlations are the key for precision medicine for hearing loss, including diagnosis, prevention and treatment. This work can also be applied to the rest of the world as a model study.

Supported by the Israel Precision Medicine Partnership Program.

136

The long and the short: Non-coding RNAs in the auditory system

Tal Koffler-Brill[1], Yael Noy[1], Mor Bordeynik-Cohen[1], Reut Yacobi[1], Michal Sperber[2], Lena Ebbers[3], Hans Gerd-Nothwang[3], Ran Elkon[2], Karen B. Avraham[1]

[1]*Laboratory of Neural and Sensory Genomics, Department of Human Molecular Genetics & Biochemistry, Faculty of Medicine, Tel Aviv University, Israel*; [2]*Department of Human Molecular Genetics & Biochemistry, Faculty of Medicine, Tel Aviv University, Israel*; [3]*Center for Neuroscience, Carl von Ossietzky University, Germany*

Non-coding RNAs (ncRNAs) are a group of RNA molecules that include a wide variety of sub-classes, mainly divided into two major groups: the long and the short ncRNAs. Long non-coding RNAs (lncRNAs) and microRNAs (miRNA), which belong to these groups, respectively, are essential regulators in many model systems and biological processes. Furthermore, their involvement in the development of pathogenic conditions, including hearing impairment, has been demonstrated. Aimed at exploring the regulatory role of the different ncRNAs in the mouse auditory system, we performed high-throughput RNA sequencing at critical time points during development of the inner ear sensory epithelium. Thousands of lncRNAs and miRNAs are expressed at embryonic day 16 and at birth, with hundreds presenting with differential expression between the two time points. Several lncRNAs and miRNAs were chosen for further analysis, among them miR-34c and miR-449a. We demonstrate that these miRNAs target *Atoh1*, a critical transcription factor for the development of hair cells. Transgenic mice carrying knockouts of specific lncRNAs or miRNAs are being generated using CRISPR-Cas9 genome editing. The relative expression of cochlear-abundant miRNAs in the sensory epithelium of the peripheral auditory system and the auditory brainstem is being evaluated. Establishing the regulatory roles of ncRNAs in the development and maturation of the inner ear is critical and will promote our ability to define the pathways essential for establishing a functional auditory system. Furthermore, these defined pathways may serve as new therapeutic and customized treatment targets for deafness.

138

TRACEvar: Prioritizing and interpreting pathogenic variants that underlie hereditary diseases by using machine learning and tissue contexts

Chanan Argov[1], Eric Sabag[1], Avigdor Mansbach[1], Yair Sepunaru[1], Emmi Filtzer[1], Yuval Yogev[2], Ohad Birk[2,3], Vered Chalifa-Caspi[4], Lior Rokach[5], Esti Yeger-Lotem[1,3]

[1]*Department of Clinical Biochemistry and Pharmacology, Ben-Gurion University of the Negev, Israel*; [2]*Faculty of Health Sciences, Morris Kahn Laboratory of Human Genetics and the Genetics Institute at Soroka Medical Center, Israel*; [3]*The National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel*; [4]*Ilse Katz Institute for Nanoscale Science & Technology, Ben-Gurion University of the Negev, Israel*; [5]*Department of Software & Information Systems Engineering, Ben-Gurion University of the Negev, Israel*

Background/Objectives: The genetic diagnosis of hereditary diseases is a longstanding challenge in basic and clinical human genetics. Although many hereditary diseases manifest clinically in specific tissues, variant interpretation schemes are mostly oblivious to tissue contexts.

Methods: TRACEvar is a machine-learning method that prioritizes pathogenic variants in tissue contexts. Its novelty is in predicting pathogenic variants that exert their effects in distinct human tissues. TRACEvar random-forest models employed 1,313 features, including 1,196 tissue-specific multi-omics features of 53 tissues, and was trained and tested on ~68,000 benign and pathogenic mutations from ClinVar database and on genetic data 52 from patients.

Results: TRACEvar machine-learning models predict variant pathogenicity in 17 tissues. With mean auROC of 0.97 and mean auPRC of 0.57 (expected 0.013), TRACEvar outperformed well-established variant prioritization methods, including CADD, Sift, PolyPhen and CAPICE. The models' interpretation revealed that features of disease-affected tissues were among the top contributing features across models. When tested on genetic data from 52 patients, TRACEvar ranked the pathogenic variant at the top 10% of the patient's variants in 60% of the cases. Hence, TRACEvar is an accurate and powerful tool for variant prioritization and interpretation in tissue contexts. A TRACEvar online webserver for variant prioritization and interpretation in 17 tissue contexts is available at https://netbio.bgu.ac.il/TRACEvar/.

Conclusion: Tissue-specific analysis of genetic variants enhances disease understanding and clinical diagnosis. Since the information of affected tissues is part of the clinical assessment of patients, we call for its incorporation into additional variant prioritization schemes.

Grant: Israel Science Foundation [317/19 to E.Y.-L.].

140

Two comprehensive databases for whole genome sequence disease decipherment

Simon Fishilevich, Ruth Barshir, Chana Rosenblum, Tsippi Iny-Stein, Shmuel Pietrokovski, Marilyn Safran, Doron Lancet

Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot, Israel

Interpreting whole genome sequences (WGS) for disease decipherment is a major challenge, since 98% of potentially significant variants reside in non-coding “dark matter”. Over the last five years, GeneCards (PMID:27322403) has been significantly expanded to address this challenge. For this, we have incorporated within GeneCards two data systems that provide all-inclusive coverage of two major types of non-coding genomic entities.

The first system is GeneHancer (PMID:28605766, 500 citations), a regulatory element database encompassing ~400,000 enhancers and promoters. Using information amalgamated from key epigenetic resources, GeneHancer creates a unique non-redundant comprehensive view of the human genes regulatory network. Most genes in GeneCards now have detailed links to GeneHancer elements, and these elements are linked to target genes, harbored transcription factor binding sites, phenotype/disease information and tissue activity patterns, all crucial for clinical interpretation of non-coding variants. GeneHancer is now used worldwide to interpret non-coding variants, constituting a native regulation track at the UCSC genome browser.

The second data system is GeneCaRNA (PMID:33676929), comprising ~275,000 non-coding RNAs (ncRNAs) of 17 functionally diverse types, such as lncRNAs, miRNAs and srpRNAs. Its unique gene-centric buildup is crucial for genomic variant interpretation. GeneCaRNA entries are seamlessly integrated into GeneCards, inheriting its comprehensive classifications and annotations infrastructure.

These novel non-coding compendia now provide an indispensable augmentation for our VarElect tool (PMID:27357693), allowing non-coding variant disease prioritization in enhancers, promoters and ncRNAs, via direct and target-gene mediated inferences. This helps facing the growing popularity of WGS, and often leads to the elucidation of unsolved clinical cases.

142

Variant re-classification at the Shamir Medical Center: Lessons from 425 analyses

Yuval Weigl[1], Avichai Reich[1], Adi Szmulewicz[1], Orit Reish[1,2]

[1]*Genetics Institute, Shamir Medical Center (Assaf Harofe), Israel*; [2]*The Sackler School of Medicine, Tel Aviv University, Israel*

Aim: To evaluate the impact of revision processing by re-calling of already postulated variants and re-analyses from raw FASTQ.

Methods: Our study applied a permissive, comprehensive and updated in-house pipeline on (A) 327 variants with a pre-determined classification acquired from 63 exomes and 105 panels; and (B) 45 WES-FASTQ files.

Results: Analysis A resulted in re-classification of 45/327 (13.8%) of the variants. Transition to an informative pathogenicity scale appeared in 31/45 re-classified variants (68.9%). A novel ACMG annotation applied for another 13/327 variants (4%) assigning an informative pathogenicity in 10 of them (3.1%). Analysis B resulted in re-classification of 11/87 major variants (12.6%). A switch to an informative pathogenicity scale appeared in 5/11 re-classified variants (45.5%). Moreover, 11 novel candidates were discovered with an informative scale in 9 of them. Altogether, the revision processing enabled to define a novel phenotype-genotype correlation in 11/205 patients (5.4%) and offer a proactive preventive treatment in 6 cases (2.9%). Overall, the re-classification increased the severity of 30/78 (38.5%) variables, whereas, decreased in 26/78 (33.3%). The decision of re-classification was based on bioinformatics-related data mostly regarding: (a) Computational and predictive data (70.9%); (b) Functional data (34.5%) and (c) Population data (29.1%).

Conclusions: The overall reclassification rate was 12.6-13.8% of raw FASTQ-driven/already postulated variants. Such values are in accordance with medical literature range (4.7-71.8%). As variant detection based on type of data mining and type of filtering on same capture, revision processing should be encouraged prior to embarking on another technology when variant is not detected.

144

A methodology for classifying tissue-specific receptor functions applied to subcutaneous and visceral adipose

Gur Arye Yehuda, Judith Somekh

Information Systems, University of Haifa, Israel

To achieve homeostasis, the human biological system relies on the interaction between organs through the binding of ligands secreted from source organs to receptors located on destination organs. Currently, the changing roles that receptors perform in tissues are only partially understood. Recently, a methodology based on receptor co-expression patterns to classify their tissue-specific metabolic functions was suggested. Here we present an advanced framework to predict an additional class of inflammatory receptors that use a feature space of biological pathway enrichment analysis scores of co-expression networks and their eigengene correlations. These are fed into three machine learning classifiers – eXtreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), and K-Nearest Neighbors (k-NN). We applied our methodology to subcutaneous and visceral adipose gene expression datasets derived from the GTEx (Genotype-Tissue Expression) project and compared the predictions. The XGBoost model demonstrated the best performance in predicting the pre-labeled receptors, with an accuracy of 0.89/0.8 in subcutaneous/visceral adipose. We analyzed ~700 receptors to predict eight new metabolic and 15 new inflammatory functions of receptors and four new metabolic functions for known inflammatory receptors in both adipose tissues. We validated multiple predictions using the published literature. Our results establish a picture of the changing functions of receptors for two adipose tissues that can be beneficial for drug development.

149

3CAC: improving the classification of phages and plasmids in metagenomic assemblies using assembly graphs

Lianrong Pu, Ron Shamir

The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel

Bacteriophages and plasmids usually coexist with their host bacteria in microbial communities and play important roles in microbial evolution. Accurately identifying sequence contigs as phages, plasmids, and bacterial chromosomes in mixed metagenomic assemblies is critical for further unraveling their functions. Many classification tools have been developed for identifying either phages or plasmids in metagenomic assemblies. However, only two classifiers, PPR-Meta and viralVerify, were proposed to simultaneously identify phages and plasmids in mixed metagenomic assemblies. Due to the very high fraction of chromosome contigs in the assemblies, both tools achieve high precision in the classification of chromosomes but perform poorly in classifying phages and plasmids. Short contigs in these assemblies are often wrongly classified or classified as uncertain.

Here we present 3CAC, a new three-class classifier that improves the precision of phage and plasmid classification. 3CAC starts with an initial three-class classification generated by existing classifiers and improves the classification of short contigs and contigs with low confidence classification by using proximity in the assembly graph. Evaluation on simulated metagenomes and on real human gut microbiome samples showed that 3CAC outperformed PPR-Meta and viralVerify in both precision and recall, and increased F1-score by 10-60 percentage points.

153

Time-dependent iterative imputation for multivariate longitudinal clinical data

Omer Noy and Ron Shamir

Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel

Missing data is a major challenge in various domains. In clinical research, electronic medical records often have a large amount of missing values in laboratory tests and vital signs. The missingness can lead to biased estimates and limit our ability to draw conclusions from the data. Additionally, commonly used machine learning algorithms can be only applied to complete datasets. A common practice to deal with this problem is data imputation, the process of filling in the missing values with substituted values. However, some of the popular imputation approaches are limited when applied to clinical data.

We developed a new approach, Time-Dependent Iterative (TDI) imputation, that offers a practical solution for imputing individualized time-series data. It addresses both multivariate and longitudinal data, by integrating forward-filling and Iterative Imputer, a version of the MICE algorithm. The integration employs a patient, variable, and observation-specific dynamic weighting strategy, based on the clinical patterns of the data, including missing rates and measurement frequency. We evaluated its performance by randomly masking values in clinical datasets, imputing them, and comparing the imputed values to the ground truth values. When applied to a cohort of 45,000 patients from MIMIC III, our approach outperformed state-of-the-art imputation methods for 14 out of 16 clinical variables, with an overall root-mean-squared-error of 13.83, compared to 16.18 for MissForest, the second best method. Similar results were achieved for three Israeli datasets of COVID-19 inpatients. Importantly, tests on these datasets also demonstrated that TDI imputation can lead to improved risk prediction.

155

GeniePool: genomic database with corresponding annotated samples based on a cloud data lake architecture

Noam Hadar[1], Grisha Weintraub[2], Ehud Gudes[2], Shlomi Dolev[2], Ohad Birk[1,3]

[1]*The Morris Kahn Laboratory of Human Genetics, Faculty of Health Sciences and NIBN, Ben Gurion University of the Negev, Israel*; [2]*Department of Computer Science, Faculty of Natural Sciences, Ben Gurion University of the Negev, Israel*; [3]*Genetics Institute, Soroka Medical Center, Israel*

In recent years there is huge influx of genomic data and a growing need for its phenotypic correlations, yet existing genomic databases do not allow easy storage and accessibility to the combined phenotypic-genotypic information. Allele-Frequency (AF) databases are crucial tools for evaluating variants in research and in the clinic. While there are multiple AF databases such as gnomAD and ALFA, none provide genotype-phenotype data. Thus, genomic variants in such databases cannot be easily associated with or traced back to their phenotypic traits. The Sequence Read Archive (SRA) accumulates hundreds of thousands of next-generation-sequencings (NGS) samples tagged by their submitters and various attributes. However, samples are stored in large raw format files, inaccessible for the common user. GeniePool (geniepool.bgu.ac.il) is a simple and intuitive web service and API for querying NGS data from SRA with direct access to phenotypic information related to each sample and related studies. Freely available online, implemented using cloud data lake architecture to achieve perfect scaling and modest overhead cost, GeniePool has significant advantages absent from other existing databases. At practically unlimited scalability, it enables easy access to all the genetic data, yet also practically unlimited storage of attached easily-accessible associated phenotypic clinical information.

157

Deciphering the epigenetic mechanisms that govern tissue-specific expression of imprinted genes using a WGBS Human Cell Atlas

Jonathan Rosenski[1], Ayelet Peretz[2], Judith Magenheim[2], Yuval Dor[2], Benjamin Glaser[3], Ruthi Shemer[2], Tommy Kaplan[1,2]

[1]*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel*;
[2]*Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Hadassah Medical Center and Faculty of Medicine, The Hebrew University of Jerusalem, Israel*;
[3]*Endocrinology and Metabolism, Hadassah Medical Center and Faculty of Medicine, The Hebrew University of Jerusalem, Israel*

Genomic imprinting, the expression of the maternally or paternally inherited alleles, plays a critical role in embryonic development, and misregulation of imprinted genes is involved in various genetic conditions. While imprinting often involves differential methylation at imprinting control regions (ICRs), the exact regulatory mechanisms are yet to be discovered, for most genes.

In this study, we generated a whole-genome human cell type-specific DNA methylation atlas, and developed computational algorithms for the joint analysis of genetic and epigenetic patterns, across 200 samples. We identified 28,000 genomic regions with allele-specific DNA methylation patterns. Many of these show sequence-dependent methylation, elucidating the relationship between genotype and phenotype.

We also identified 275 “imprinted” regions (with an allele-specific parent-of-origin differential methylation), including most known ICRs and 100 novel regions, many of which show cell type-specific “escape” from imprinting in a limited number of cell types.

These regions shed light on the molecular mechanisms underlying uniparental expression of imprinted genes, and include differential methylation of regulatory regions, allele-specific methylation of CTCF sites which alter the 3D folding of the genome, and recruitment of distal, cell type-specific enhancers. In addition, this human cell type-specific imprinting atlas and associated methods offer a rich resource for the study of genomic imprinting biology and provide insight in the pathogenesis of imprinting-related genetic diseases.

158

Enhanced batch correction on the GTEx data when comparing two populations

Ehud Sussman, Judith Somekh

Information Systems, University of Haifa, Israel

The GTEx dataset is a highly heterogenous gene expression dataset of over 17,000 human RNA-seq samples derived from approximately 1000 donors across more than 50 human tissues. The dataset also includes data pertaining donor phenotype and batch attributes.

The correction of gene expression can target known confounders or hidden ones. One method for correcting for known confounding factors is the multivariate linear regression based (MLR). MLR has been demonstrated to be successful in improving gene expression level estimates with minimal harm to the genuine biological signal for the GTEx data. Among the known artefacts that affect the GTEx gene expression levels and need to be corrected are batch effects, ischemic time (IT) which is the time interval between actual death and sample extraction and type of death (TD) defined on the Hardy Scale. We found a strong multicollinearity between IT and the TD when applied in a MLR-based correction. This multicollinearity yields anomalies in correlation coefficients. These correlations stem from the fact that certain death classification types could potentially have shorter IT, due to proximity to a hospital at the time of death or for other reasons.

We developed a new two-step regression approach that resolves the coefficient anomaly and also preserves the batch effect data. This approach can be adopted for yielding a more accurate gene expression levels, in comparison to other methods that were attempted that were prone to overfitting.

159

Fasting hormones synergistically induce amino acid catabolism genes to promote gluconeogenesis

Noga Korenfeld[1], Maya Finkel[1], Nufar Buchshtab[1], Meirav Bar-Shimon[1], Meital Charni-Natan[1], Ido Goldstein[1]

Institute of Biochemistry, Food Science and Nutrition, Hebrew University of Jerusalem, Israel

BACKGROUND & AIMS: Gluconeogenesis from amino acids (AAs) maintains glucose homeostasis during fasting. Although glucagon is known to regulate AA catabolism, the contribution of other hormones to it and the scope of transcriptional regulation dictating AA catabolism are unknown. We explored the role of the fasting hormones glucagon and glucocorticoids in transcriptional regulation of AA catabolism genes and AA-dependent gluconeogenesis.

METHODS: We tested the RNA expression of AA catabolism genes and glucose production in primary mouse hepatocytes treated with fasting hormones (glucagon, corticosterone). We analyzed genomic data of chromatin accessibility and chromatin immunoprecipitation in mice and primary mouse hepatocytes. We performed chromatin immunoprecipitation in livers of fasted mice to show binding of cAMP responsive element binding protein (CREB) and the glucocorticoid receptor (GR).

RESULTS: Fasting induced the expression of 31 genes with various roles in AA catabolism. Of them, 15 were synergistically induced by co-treatment of glucagon and corticosterone. Synergistic gene expression relied on the activity of both CREB and GR. Enhancers adjacent to synergistically induced genes became more accessible and were bound by CREB and GR on fasting. Akin to the gene expression pattern, gluconeogenesis from AAs was synergistically induced by glucagon and corticosterone in a CREB- and GR-dependent manner.

CONCLUSIONS: Transcriptional regulation of AA catabolism genes during fasting is widespread and is driven by glucagon (via CREB) and corticosterone (via GR). Glucose production in hepatocytes is also synergistically augmented, showing that glucagon alone is insufficient in fully activating gluconeogenesis.

162

RNA editing in pancreatic beta-cells prevents aberrant innate immune activation and diabetes

Udi Ehud Knebel[1], Shani Peleg[1], Roni Cohen-Fultheim[2], Erez Levanon[2], Agnes Klochendler[1], Yuval Dor[1]

[1]*Developmental Biology and Cancer Research, The Hebrew University - Hadassah Medical School, Israel*; [2]*Faculty of Life Sciences, Bar Ilan University, Israel*

RNA editing involves the deamination of adenosines to inosines by adenosine deaminases acting on RNA (ADAR). This process is crucial for dismantling endogenously produced double stranded RNA (dsRNA), which can trigger an anti-viral inflammatory program. RNA editing has been studied in several organs and in cancer, yet little is known about its importance in pancreatic beta-cell biology. Furthermore, since islet inflammation is a feature of type 1 diabetes (T1D), and since GWAS revealed protective variants of the IFIH1 dsRNA sensor on T1D risk, alterations in RNA editing may contribute to islet inflammation in T1D. We deleted the Adar1 gene in beta-cells of 1-month-old mice. Reduced editing resulted in the induction of interferon-stimulated genes in mutant beta cells, triggering a massive immune response, resembling the anti-viral-like response seen in early stages of T1D. Moreover, editing elimination impaired the insulin synthesis and beta-cell expression program only in editing-deficient cells, sparing wild type neighbors. Surprisingly, we observed an inverse correlation between the age at Adar inactivation and the severity of insulitis and diabetes. Both activation of the IFN pathway and its paracrine effects are weaker upon Adar1 deletion in older mice. Altogether, we present a new model for the interferon response and the recruitment of innate immune cells to pancreatic islets, mimicking processes taking place in early stages of T1D. Ongoing experiments assess the role of adaptive immunity in RNA editing-related beta cell destruction and the effects of editing elimination on other islet cell types.

163

Genotype imputation strategy for improving disease risk prediction on diverse ethnic groups

Hagai Levi[1], Ran Elkon[2], Ron Shamir[1]

[1]*Blavatnik School of Computer Science, Tel-Aviv University, Israel*; [2]*Sackler Faculty of Medicine, Tel-Aviv University, Israel*

Recent studies have established that individuals can be stratified according to their risk to develop a disease using polygenic risk score (PRS). PRS accrues the effect of genomic variants discovered by genome-wide association studies (GWASs). However, to date, sufficiently large GWASs exist mostly for European population. PRS performance declines with the genetic distance between the discovery and target populations. A key factor of PRS is imputation, the inference of the genotype of un-typed SNPs, using a set of sequenced individuals (called the reference panel). The resulting imputed SNPs can differ, depending on the ethnic composition of the reference panel. Several studies have shown that imputing a genotype dataset using an ethnic-matched imputation panel improves the completion of missing SNPs. Nevertheless, accurate SNP imputation does not necessarily imply better risk prediction with PRS.

Here, we explore the effect of imputation reference panel PRS on prediction accuracy. We use a PRS obtained on the European population and examine its performance on non-European target populations, including Ashkenazi Jewish, South Asian and African. Our analysis points to an underappreciated complexity, and, unexpectedly, demonstrates that using an imputation reference panel that from the target population does not necessarily yield better PRS performance for that population. Our results show how adaptation of imputation panels may increase the accuracy of disease prediction on non-European populations based on currently available genomic data.

164

Insights from multigene panel testing in non-Ashkenazi breast cancer patients – carrier rates and variant classification in a genetically diverse population

Sari Lieberman[1], Omer Murik[1,2], Fouad Zahdeh[1,2], Rachel Beeri[1], Pinhas Renbaum[1], Michal Barzily[1], Tehila Klopstock[1], Ariela Tomer[1], David A Zeevi[1,2], Tzvia Mann[1,2], Oded Olsha[4], Tal Hadar[3], Moshe Carmon[4], Hadar Goldvaser[5,7], Shani Paluch-Shimon[6], Eitan Friedman[8,9], Amnon Lahad[10,11], Rinat Bernstein-Molho[8,9], Ephrat Levy-Lahad[1,7]

[1]*Medical Genetics Institute, Shaare Zedek Medical Center, Israel*; [2]*Translational Genomics Lab, Medical Genetics Institute,, Shaare Zedek Medical Center, Israel*; [3]*Department of Breast Surgery, Hadassah Hebrew University Medical Center, Israel*; [4]*Surgical Department, Shaare Zedek Medical Center, Israel*; [5]*Institute of Oncology, Shaare Zedek Medical Center, Israel*; [6]*Sharett Institute of Oncology, Hadassah Hebrew University Medical Center, Israel*; [7]*Faculty of Medicine, The Hebrew University of Jerusalem, Israel*; [8]*The Susanne Levy Gertner Oncogenetics Unit, Institute of Genetics, Sheba Medical Center, Israel*; [9]*Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, Israel*; [10]*Department of Family Medicine, The Hebrew University of Jerusalem, Israel*; [11]*Jerusalem District, Clalit Health Services, Israel*

Background: There is limited information on the landscape of inherited breast cancer (BC) in non-Ashkenazi Jews. This heterogeneous, little tested population,can offer insights into novel variants in known BC-predisposition genes.

Methods. In 2015-2021 consecutively diagnosed non-Ashkenazi BC patients underwent multigene panel testing (MGPT). MGPT was also offered to a control group of unaffected non-Ashkenazi participants.

Results: Genetic testing was performed in 751 affected and 810 unaffected women. BRCA1/2 pathogenic variants (PV) were identified in 23 (3.3%) affected vs. 5 (0.6%) unaffected women (P .0002). Only 3 BRCA1 PVs were recurrent. PVs in other genes were found in 30 (4.8%) affected vs. 8 (1%) unaffected women. This included PVs in ATM, CHEK2, BRIP1, TP53, MRE11, FANCM, NBN, PALB2 and PTEN.

Rare variants classification: 9490 rare variants (.01 MAF) were observed in 28 genes: 695 (7.3%) exonic, 8789 (92.7%) non-coding. 541 exonic variants were previously in ClinVar. Using frequencies from this cohort enabled downgrading of 15 ClinVar VUSs to Benign, based on MAF 0.01 among unaffected low-FH women. MAF by sub-ethnicity downgraded another 28 VUSs. Among non-coding variants, 68 were significantly more common in affected with significant FH, suggesting possible pathogenicity. Conversely, 52 non-coding variants were classified as benign based on higher frequency in controls.

Conclusion: Non-Ashkenazi women exhibit low rates of PVs in known breast cancer genes, even in affected with substantial FH. Genetic analysis in diverse populations can contribute to variant classification, especially of non-coding variants whose interpretation by standard tools is limited.

165

Obesity and cancer: A microRNA story

Ari Meerson[1,2], Rahaf Salalha[1,2], Hila Yehuda[1,2], Yaniv Eliraz[1,2]

[1]*Genomics Unit / Health and Nutrition, MIGAL Galilee Research Institute, Israel*
[2]*Faculty of Sciences, Tel Hai Academic College, Israel*

Obesity is a risk factor for several cancer types, suggesting shared molecular mechanisms. We identified cancer-relevant microRNAs that respond to metabolic hormone signaling in cultured cells and/or to metabolic changes in human subjects. In collaboration with Lorna Harries (Exeter University, UK) we reported that miR-10b is more strongly downregulated in the primary breast tumors of obese patients, demonstrating that the metabolic state of the organism can lead to a significant difference in the molecular pathology of cancer. In ductal but not lobular tumors, significant inverse correlations were observed between the tumor levels of miR-10b and miR-30c and the mRNA levels of cancer-relevant target genes SRSF1, PIEZO1, MAPRE1, CDKN2A, TP-53 and TRA2B, as well as tumor grade.Suppression of miR-10b levels in BT-549 primary BC–derived cells increased cell proliferation and invasive capacity, while exogenous miR-10b mimic decreased invasion. Manipulation of miR-10b levels also inversely affected the mRNA levels of miR-10b targets BCL2L11, PIEZO1 and NCOR2. Previously, we had reported that in cultured colon cancer cells, miR-4443 was upregulated by leptin and insulin in a MEK1/2-dependent manner. MiR-4443 overexpression decreased invasion and proliferation, and directly downregulated NCOA1 and TRAF4, genes involved in metastasis. Insulin and/or leptin resistance (e.g. in obesity) may suppress this tumor-suppressive pathway and increase cancer risk. Supporting this notion, the miR-4443 locus is frequently deleted in cancers. We recently showed that miR-4443 is a non-canonical microRNA with a yet unknown biogenesis pathway.

Our current research explores the endocrine and epigenetic mechanisms which regulate these cancer-relevant microRNAs under conditions of obesity.

166

Effects on mammalian cell transcriptome reveal a plant extract`s medicinal properties: Potential anti-COVID-19 and anti-cancer properties of Sarcopoterium spinosum

Valid Gahramanov[1], Moria Oz[2], Tzemach Aouizerat[3], Mali Salmon-Divon[1], Tovit Rosenzweig[1], Jonathan Gorelick[3], Elyashiv Drori[2], Michael Y.Sherman[1], Bat Chen R. Lubin[2]

[1]*Molecular Biology, Ariel University, Israel*; [2]*Agriculture and Oenology Department, Eastern Regional R&D Center, Israel*; [3]*Chemical Engineering, Biotechnology and Materials Department, Hebrew University, Israel*

Plants with medicinal properties are usually identified based on traditional medicine knowledge or using low-throughput screens for specific pharmacological activities. Here, we suggest a different approach to uncover a range of pharmacological activities of a chosen plant extract without the need for functional screening. This tactic predicts biological activities of a plant extract based on pathway analysis of transcriptome changes caused by the extract in mammalian cell culture. In this work, we identified transcriptome changes after exposure of cultured cells to an extract of the medicinal plant Sarcopoterium spinosum. Gene Set Enrichment Analysis (GSEA) confirmed known anti-inflammatory and anti-cancer activities of the extract and predicted novel biological effects on oxidative phosphorylation and interferon pathways. Experimental validation of these pathways uncovered strong activation of autophagy, including mitophagy, and astounding protection from SARS-CoV-2 infection. Our study shows that gene expression analysis alone is insufficient for predicting biological effects since some of the changes reflect compensatory effects, and additional biochemical tests provide necessary corrections. In conclusion, this study defines the advantages and limitations of an approach towards predicting the biological and medicinal effects of plant extracts based on transcriptome changes caused by these extracts in mammalian cells.

167

Double mutation in ARID1A and PIK3CA in luminal breast cancer is associated with inflammation-driven tumorigenesis

Liat Sinberger[1], Tamar Zahavi[1], Amir Sonnenblick[2,3], Mali Salmon-Divon[1]

[1]*Department of Molecular Biology, Adelson School of Medicine, Ariel University, Israel*; [2]*Institute of Oncology, Tel Aviv Sourasky Medical Center, Israel*; [3]*Sackler Faculty of Medicine, Tel Aviv University, Israel*

Breast cancer is the most frequent cancer among women worldwide. Mutations in ARID1A, are the most common alterations of the SWI/SNF complex in estrogen-receptor-positive (ER+) breast cancer and tumor relapse is suggested to be associated with mutations in ARID1A. Co-occurring ARID1A and PIK3CA mutations have been reported in patients with other cancers, suggesting the possibility of cooperativity of these genes. The main goal of this project is to reveal the effect of ARID1A loss separately and in combination with PIK3CA mutations on important cellular and immune signaling pathways in luminal breast tumors. Our cohort study contains 2609 breast cancer samples from TCGA and METABRIC databases, of which expression data as well as mutational data were available. In silico data analysis revealed that ARID1A mutation exhibit statistically significant co-occurrence with PIK3CA mutation in luminal breast cancer. We found that ARID1A mutation is associated with downregulation of immunological pathways. However, when this mutation co-occurs with PIK3CA mutation (double mutations) the immune signaling pathways are highly upregulated. Weighted correlation network analysis detected seven hub genes with high connectivity in the double mutation-related gene module, most of which have a role in T cell receptor signaling pathway. In addition, a unique immune cell infiltration profile has been detected in double mutated tumors. Our results suggest that ARID1A and PIK3CA mutations cooperate to promote inflammation-driven tumorigenesis, and these tumors may be good candidates for immunotherapy treatments.

168

A missense variation in a gene regulating actin leads to impaired actin dynamics and is associated with dilated cardiomyopathy

Pierre Majdalani[1,2], Aviva Levitas[3], Hanna Krymko[3], Leonel Slanovic[3], Alex Braiman[1], Uzi Hadad[4], Salam Dabsan[1,2], Amir Horev[5], Raz Zarivach[2,6], Ruti Parvari[1,2]

[1]*The Shraga Segal Department of Microbiology, Immunology & Genetics, Faculty of Health Sciences, Ben-Gurion University of the Negev, Israel*; [2]*The National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel*; [3]*Department of Pediatric Cardiology, Soroka University Medical Center and Faculty of Health Sciences, Ben-Gurion University of the Negev, Israel*; [4]*The Ilse Katz Institute for Nanoscale Science and Technology, Ben-Gurion University of the Negev, Israel* [5]*Pediatric Dermatology Service, Soroka University Medical Center, Israel*; [6]*Department of Life Sciences, Faculty of Natural Sciences, Ben-Gurion University of the Negev, Israel*

Background: Dilated cardiomyopathy (DCM) is a primary myocardial disease leading to contractile dysfunction, progressive heart failure, and excessive risk of sudden cardiac death. Around half of DCM cases are idiopathic, and genetic factors seem to play an important role.

Aim: We investigated a possible genetic cause of DCM in a consanguineous child from a Bedouin family.

Methods and results: Using whole-exome sequencing and searching for rare homozygous variations, we identified a missense variant in a gene, involved in actin regulation, not previously associated with human disease as the causative mutation. The missense change may affect the binding of the protein to actin by eliminating hydrogen bonds between them. The protein localization to the cytoplasm is not affected by the change in amino acid. Using patient's fibroblasts, we show by Western blotting a decrease in the ratio of globular to fibrillary actin in comparison to control fibroblasts. Re-polymerization of fibrillary actin after treatment with cytochalasin D, which disrupts the actin filaments is slower in patients' fibroblasts. Finally, the patient's fibroblasts bridged a scar gap slower than control fibroblasts both because of slower and indirect movement.

Conclusions: This is the first report of a mutation in humans in the relevant gene. A recessive deleterious mutation in the gene is associated with DCM in humans. Our data underscore the importance of this gene in regulating the monomeric actin pool, the kinetics of actin polymerization, and cell movement, emphasizing the importance of actin regulation for the normal function of the human heart.

169

The effects of non-oncology drugs on the risk of distant recurrence in luminal early breast cancer patients

Assaf Goldberg[1], Tamar Zahavi[1], Barliz Waissengrin[2,3], Ido Wolf[2,3], Uri Ben-David[3], Amir Sonnenblick[2,3], Mali Salmon-Divon[1]

[1]*Department of Molecular Biology, Ariel University, Israel*; [2]*Institute of Oncology, Tel Aviv Sourasky Medical Center, Israel*; [3]*Sackler Faculty of Medicine, Tel Aviv University, Israel*

Luminal breast cancer is characterized by the expression of estrogen receptor (ER) and lack of expression of the HER2 receptor, constitutes at least half of all new breast cancer diagnoses, with best prognosis. The risk of disease recurrence and death in these patients can persist for many years. Prognostic factors for distant recurrence include classical clinicopathological parameters such as tumor size, tumor grade, and nodal status, tumor molecular testing, and non-oncology medication use. Here we systematically examined the impact of non-oncology drugs on the risk of distant recurrence in 1385 luminal early breast cancer patients as measured by the OncotypeDx Recurrent Score (RS). We found that Levothyroxine was significantly associated with high OncotypeDx RS (mean=24.78 ;p0.0001) and Metformin was associated with low OncotypeDx RS (mean=14.87; p0.01) in comparison to patients not receiving these medications (mean=18.7). No difference was detected in the levels of the classic proliferation marker Ki67 between patients who used the drugs and those who did not. On the other hand, the PR and ER expression levels were different between patients who received Levothyroxine or Metformin and patients who did not. Finally, by using contemporary guidelines to recommend adjuvant chemotherapy based on clinical risk and genomic risk (OncotypeDx), we show that patients (Age 50) who received Metformin treatment had 14.5% chance to be recommended adjuvant chemotherapy while patients who received Levothyroxine had 49% (p= 0.0001). To summarize, our results indicate significant impact of Metformin and Levothyroxine on clinical decisions with potential impact on early BC patients.

170

Characterization of human adipose tissues at the single cell level using single nuclei RNA-sequencing

Or Lazarescu[1], Yulia Haim[1], Idan Hekselman[1], Juman Jubran[1], Danny Kitsberg[2], Antje Korner[3], Rinki Murphy[4], Matthias Bluher[5], Naomi Habib[2], Assaf Rudich[1,6], Esti Yeger-Lotem[1,6]

[1]*Department of Clinical Biochemistry & Pharmacology, Ben-Gurion University of the Negev, Israel*; [2]*Edmond & Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Israel*; [3]*Medical Faculty, LIFE Child (Leipzig Research Centre for Civilization Diseases), University of Leipzig, Germany* [4]*Department of Medicine, Faculty of Medical and Health Sciences, University of Auckland, New Zealand*; [5]*Department of Medicine, University of Leipzig, Germany*; [6]*National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Israel*

Adipose tissue (AT) is a complex tissue that apart from storing extra energy is a major endocrine organ. For example, it produces the adiponectin hormone that has autocrine effect on adipocytes, paracrine effects on AT immune system and endocrine effects on brain, liver and pancreas. The complexity of AT and its impact on human health, particularly obesity, call for its characterization at the single-cell resolution. However, previous efforts to characterize AT were mostly limited to its non-adipocyte fraction, since adipocytes are not amenable to single cell RNA-sequencing due to their size and fragility. These efforts led to a partial view of AT that ignores adipocytes, the hallmark of adipose tissue. To overcome this hurdle we applied single-nuclei RNA sequencing (snRNA-seq) that is amenable to adipocytes as well as other AT cell types. Here we describe the application of snRNA-seq to visceral AT from 11 donors of both sexes and a range of BMI. We sequenced and analyzed 88,919 nuclei, resulting in the most comprehensive atlas of visceral AT to date. In addition to stem cells, adipose progenitor cells, immune cells and endothelial cells, 19% of the atlas is composed of adipocytes. Analysis of adipocyte sub-types revealed classical adipocytes as well as specialized adipocytes related to inflammation and immune system interaction. We also observed differences in AT and adipocyte compositions between sexes that were not observed before. This atlas presents a resource for further clinical investigations of AT effect on human health.

171

Determining the “Position – Effect” in 3D aspect

Garima Chaturvedi[1], Ronen Krispil[1], Ofir Hakim[1]

The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Israel

The division of the genome into topologically related regions can play a significant role in gene expression control. But what happens when there is no separation into topologically associating domains (TADs). In plants, understanding high-resolution three-dimensional genome structure and its impact on transcription is substantially lacking. Transgenes randomly integrated into plant genome show position dependent expression level, thus can be a powerful tool to reveal the influence of its position on gene expression. The cause of transgene expression variability remains unclear. However, certain T-DNA integration characteristics often relate to the extent of transgene expression. Although transgene silencing is often associated with multiple T-DNA insertions at the same locus, this need not to be always the case, and single-T-DNA insertions may also silence. Obviously, if the transgene will be in heterochromatin, it is likely to be silenced, but when it is active, what determines the expression level? What is the context that matters? How far it is positioned relatively to the transgene? How it influences quantitatively on the expression level? To understand how position in the genome may influence the expression level of gene, we transform Arabidopsis plants with the library of barcoded reporter plasmids via Agrobacterium-mediated transformation. The RNA level as well as the genomic integration site of individual transgenes will be measured by RNA-seq and 4C-seq, respectively. Furthermore, 4C-seq allows measuring the spatial genomic organization of each transgenic insert. We will enable characterizing the 3D genomic environment of the transgenes and understanding its possible regulatory impact.

173

The story of Phoenician populations across the Mediterranean told through ancient DNA

Harald Ringbauer[1,3], Ayelet Salman-Minkov[2], Iñigo Olalde[3,4], Alissa Mitnick[3], Tomer Peled[2], Arie Shaus[3], Maria Bofill[3], Rebecca Bernardos[3], Nasreen Broomandkhoshbacht[3], Kim Callan[3], Elizabeth Curtis[3], Aisling Kearns[3], Ann Marie Lawson[3], Matthew Mah[3], Swapan Mallick[3], Adam Micco[3], Jonas Oppenheimer[3], Liju Qiu[3], Kristin Stewardson[3], Noah Workman[3], Fatma Zalzala[3], Nicholas Márquez-Grant[3,5], Antonio M. Sáez Romero[3,6], María Luisa Lavado Florido[3,6], Juan Manuel Jimenez Arenas[3,7], Isidro Jorge Toro Moyano[3,8], Enrique Viguera[3,9], Jose Suarez Padilla[3,10], Alicia Rodero Riaza[3,11], Patricia Smith[3,12], Marina Faerman[3,12], Luca Sineo[3,13], Gioacchino Falsone[3,14], Davide Pettener[3,15], Peter Van Dommelen[3,16], Francesca Oliveri[3,17], Pamela Toti[3,18], Valentina Giuliana[3,19], Alon Barash[3,20], Liran Carmel[3,21], Elisabetta Cilli[3,22], Anna Chiara Fariselli[3,22], Donata Luiselli[3,22], Brendan Culleton[3,23], Elisabetta Boaretto[3,24], Nadin Rohland[3], Alfredo Coppa[3,25], David Caramelli[3,26], Ron Pinhasi[3,27], Carles Lalueza-Fox[3,29], Dalit Regev[3,28], **Ilan Gronau**[2], David Reich[3]

[1]*Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Germany*; [2]*Efi Arazi School of Computer Science, Reichman University, Israel*; [3]*Department of Human Genetics, Harvard Medical School, USA*; [4]*Biomics Research Group, University of the Basque Country UPV/EHU, Spain*; [5]*Defence Academy of the United Kingdom, Cranfield Forensic Institute, Cranfield University, UK*; [6]*Department of Prehistory and Archaeology, University of Seville, Spain*; [7]*Department of Prehistory and Archaeology, University of Granada, Spain*; [8]*Archaeological Museum of Granada, Spain*; [9]*Department of Cellular Biology, Genetics, and Physiology, University of Malaga, Spain*; [10]*Department of Historical Sciences, University of Malaga, Spain*; [11]*Department of Protohistory and Colonizations, Nat, Spain*; [12]*Hadassah Faculty of Dental Medicine, Hebrew University, Israel*; [13]*Department of Biological Chemical and Pharmaceutical Sciences and Technologies, University of Palermo, Italy*; [14]*Department of Culture and Society, University of Palermo, Italy*; [15]*Department of Biological, Geological, and Environmental Sciences, University of Bologna, Italy*; [16]*Joukowsky Institute of Archaeology, Brown University, USA*; [17]*Sicily Region, Ministry of Cultural Heritage, Italy*; [18]*The Giuseppe Whitaker Foundation, --, Italy*; [19]*Department of Archaeology, University of Palermo, Italy*; [20]*The Azrieli Faculty of Medicine, Bar Ilan University, Israel*; [21]*The Alex Silberman Institute of Life Science, Hebrew University, Israel*; [22]*Department of Cultural Heritage, University of Bologna, Italy*; [23]*Institute of Energy and the Environment, Pennsylvania State University, USA*; [24]*Scientific Archaeology Unit, D-REAMS Radiocarbon Dating Laboratory, Weizmann Institute of Science, Israel*; [25]*Department of Human and Animal Biology, University of Rome, Italy*; [26]*Department of Biology, University of Florence, Italy*; [27]*Department of Evolutionary Anthropology, University of Vienna, Australia*; [28]*Israel Antiquities Authority, --, Israel*; [29]*Department of Evolutionary Biology, Pompeu Fabra University, Spain*

Phoenicians played a central role in establishing trade routes throughout the Mediterranean during the second and first millennia BCE, with settlements spread from their homeland in the Levant all the way west to the Iberian Peninsula. However, due to the lack of primary written records, our knowledge about Phoenician people is quite limited. Ancient DNA can finally help us tell their story. We sequenced 150 genomes from 13 different Phoenician sites: four sites from the Iberian Peninsula and nearby islands, six from Sicily and Sardinia, two from North Africa, and one site from the Levant. Our data set spans a time range from the 8th century BCE until the Roman imperial period. We find that during this time period, Phoenicians from the Central and Western Mediterranean did not share significant ancestry with Phoenicians from the Levant. Populations

from all 12 sites sampled outside of the Levant appear to attribute most of their ancestry to Bronze Age populations in the Central Mediterranean (Sicily or Greece). We also find evidence of North African ancestry in individuals from various sites, likely facilitated by the Phoenician presence in North Africa. The proportion of North African ancestry appears to significantly increase during the height of Carthage in the 4th and 3rd centuries BCE. We used long shared genomic segments to reconstruct a network of familial relationships within some of these sites. Interestingly, we also find family relationships between individuals from North Africa and individuals from Sicily, demonstrating the high mobility of Phoenicians across the Mediterranean.

174

The lncRNA Cytoskeleton regulator RNA (CYTOR) as a master regulator of epigenetics in cancer and a novel therapeutic target.

Susie Altalef-Mishaan[1], Nisrine Lahoud-Jeries[1], **Assaf Bester**[1]

Faculty of Biology, Technion-Israel Institute of Technology, Israel

Toxicity and drug resistance are two major barriers to achieving complete remission for cancer patients. Therefore, identifying cancer-specific therapeutic targets is an imminent need. While we are close to achieving a comprehensive understanding of the role of protein-coding genes in cancer, the role of the vast non-coding genome is elusive and controversial. Long non-coding RNAs (lncRNAs) are the largest group of non-protein-coding genes and the most diverse in their function. While more than 500 cancer-associated lncRNAs have been identified, their therapeutic potential has not yet been exploited. Cytoplasmic enriched lncRNAs are druggable targets, but the function of many of them is elusive. To study the function of cytoplasmic enriched lncRNAs, we used CRISPR interference (CRISPRi)-based gene perturbation followed by RNA sequencing. Our analysis identified that the lncRNA Cytoskeleton regulator RNA (CYTOR) affects the expression of hundreds of genes. These differentially expressed genes were strongly enriched to epigenetic pathways. Specifically, CYTOR regulates the expression of TET1/2/3 and IDH1/2. Mechanistically, our transcriptomic data show that CYTOR regulates gene expression outside of its locus, suggesting trans regulation activity. Clinically, CYTOR is overexpressed in a number of cancer types and is associated with a poor prognosis. Consistent with its oncogenic role, CYTOR knockdown led to a dramatic reduction in the fitness of cancer cells. Overall, our study identified a new function for the oncogenic lncRNA CYTOR as a master regulator of epigenetics in cancer, and as a potential therapeutic target.

176

Elucidating the differential role of tRNAs in cellular physiology

Orna Dahan[1], Yitzhak Pilpel[1], Noam Stren- Ginossar[1]

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

The translation machinery is under extensive selection pressure to optimize its efficiency. Translation optimization is facilitated in part by the adjustment of the tRNA pool to the transcriptome's codon usage, which constitutes an essential factor of gene expression programs in various species, cell types, and conditions. We aim to reveal the role of tRNAs in various physiological states, particularly proliferative state and viral infection. Using a custom-made tRNA deep-sequencing method, we followed the changes in the cellular tRNA pool during viral infection. We found a dynamic change of the tRNA pool in response to hCMV infection that was recapitulated following interferon treatment, which suggests that the change is part of the cellular immune response. We also found that tRNA modifications that are associated with cell proliferation are regulated during hCMV infection. An exciting hypothesis is that these changes in modification level may have an active role in entering the cell cycle hence contributing to viral reproduction.

We are currently examining the essentiality of various cellular tRNAs for viral infection. For that, we designed a tRNA-CRISPR library that systematically targets all human tRNA genes. We have already successfully applied our innovative tRNA-CRISPR library approach in the context of cellular proliferation, in which we target 20 tRNA families in diverse human cell lines, and found that tRNA essentiality depends on proliferation rate and cell origin. We are currently applying the tRNA-CRISPR system on hCMV infection to reveal the role of each human tRNA in facilitation or abrogation of viral infectivity.

181

A human DNA methylation atlas reveals principles of cell type-specific methylation and identifies thousands of cell type-specific regulatory elements

Netanel Loyfer[1], Judith Magenheim[2], Ayelet Peretz[2], Gordon Cann[3], Joerg Bredno[3], Agnes Klochendler[2], Ilana Fox-Fisher[2], Sapir Shabi-Porat[1], Merav Hecht[2], Tsuria Pelet[2], Joshua Moss[2], Zeina Drawshy[2], Hamed Amini[3], Patriss Moradi[3], Sudharani Nagaraju[3], Dvora Bauman[4], David Shveiky[4], Shay Porat[4], Gurion Rivkin[5], Omer Or[5], Nir Hirshoren[6], Einat Carmon[7], Alon Pikarsky[8], Abed Khalaileh[7], Gideon Zamir[7], Ronit Grinboim[7], Machmud Abu Gazala[7], Ido Mizrahi[7], Noam Shussman[7], Amit Korach[9], Ori Wald[9], Uzi Izhar[9], Eldad Erez[9], Vladimir Yutkin[10], Yaacov Samet[11], Devorah Rotnemer Golinkin[12], Kirsty L. Spalding[13], Henrik Druid[14,15], Peter Arner[16], A.M. James Shapiro[17], Markus Grompe[18], Alex Aravanis[3], Oliver Venn[3], Arash Jamshidi[3], Ruth Shemer[2], Yuval Dor[2], Benjamin Glaser[12], Tommy Kaplan[1,2]

[1]*School of Computer Science and Engineering, Hebrew University, Israel*; [2]*Dept. of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [3]*GRAIL, GRAIL, Inc., USA*; [4]*Dept. of Obstetrics and Gynecology, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [5]*Dept. of Orthopedics, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [6]*Dept. of Otolaryngology, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [7]*Dept. of General Surgery, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [8]*Surgery Division, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [9]*Dept. of Cardiothoracic Surgery, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [10]*Dept. of Urology, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [11]*Dept. of Vascular Surgery, Shaare Zedek Medical Center, Shaare Zedek, Israel*; [12]*Dept. of Endocrinology and Metabolism, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*; [13]*Dept. of Cell and Molecular Biology, Karolinska Institutet, Sweden*; [14]*Dept. of Oncology-Pathology, Karolinska Institutet, Sweden*; [15]*Dept. of Forensic Medicine, The National Board of Forensic Medicine, Sweden*; [16]*Dept. of Medicine (H7) and Karolinska University Hospital, Karolinska Institutet, Sweden*; [17]*Dept. of Surgery and the Clinical Islet Transplant Program, University of Alberta, Canada*; [18]*Pape Family Pediatric Research Institute, Oregon Health & Science University, USA*

DNA methylation is a fundamental epigenetic mark that governs chromatin organization, cell identity, and gene expression. Here we describe a human methylome atlas, based on deep whole-genome bisulfite sequencing allowing fragment-level analysis across thousands of unique markers for 39 cell types sorted from 207 healthy tissue samples.

Replicates of the same cell-type are 99.5% identical, demonstrating robustness of cell identity programs to genetic variation and environmental perturbation. Unsupervised clustering of the atlas recapitulates key developmental elements, whereas loci uniquely unmethylated in an individual cell type often reside in transcriptional enhancers and contain DNA binding sites for tissue-specific transcriptional regulators. Uniquely hyper-methylated loci are rare and are enriched for CpG islands, polycomb targets, and CTCF binding sites, suggesting a role in shaping cell type-specific chromatin looping.

The atlas provides an essential resource for interpretation of disease-associated genetic variants, and a wealth of potential tissue-specific biomarkers for use in liquid biopsies.

182

AI-based predictions of gene essentiality from gene expression data in 800 cancer cell lines

Jonathan Rosenski[1], Sagiv Shifman[2], Tommy Kaplan[1,3]

[1]*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel*;
[2]*Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Israel*;
[3]*Department of Developmental Biology and Cancer Research, Faculty of Medicine, The Hebrew University of Jerusalem, Israel*

The study of gene essentiality, which measures the importance of a gene for cell division and survival, is important for the identification of potential drug targets and understanding of tissue-specific manifestation of genetic conditions.

Using gene essentiality and expression data from the DepMap project, covering 800 cancer cell lines, we developed machine learning algorithms to identify ~3,000 genes whose essentiality levels are explained by the expression of a small set of "modifier genes". For some genes, essentiality depends on the expression levels of their paralogue genes, suggesting a redundancy (backup) mechanism. In others, essentiality is more complex and can only be deciphered using deep learning or other computational models.

For this, we developed a computational framework for identifying modifier gene sets for each target gene, using an ensemble of statistical tests capturing linear and non-linear dependencies. We then train several regression models for the essentiality of the target gene, and use an automated model selection procedure to identify the optimal one. Overall, we examined linear models, random bootstrapping forest, gradient boosted trees (XGBoost), Gaussian process regression models, and deep (fully connected) networks.

Our modeling framework outperforms current state-of-the-art methods, by matching the optimal model complexity for each gene, thus avoiding overfitting. Overall, we present an accurate computational approach for interpretable modeling of essentiality in a wide-range of cellular conditions, thus contributing to a better understanding of the molecular mechanisms that govern tissue specific effects of genetic diseases.

183

Computational suites for fragment-level analysis of human DNA methylation sequencing data

Netanel Loyfer[1], Jonathan Rosenski[1], Tommy Kaplan[1,2]

[1]*School of Computer Science and Engineering, Hebrew University, Israel*; [2]*Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Hadassah Medical Center and Faculty of Medicine, Hebrew University, Israel*

Next-generation sequencing of DNA methylation sheds light on the fundamental role of methylation in cellular function and packaging, in health and disease. These data are commonly represented and analyzed at CpG site resolution, while the fragment-level representation and analysis are often overlooked.

Here, we present wgbstools, an extensive general-purpose computational suite tailored for bisulfite-sequencing data. By stripping away genetic information while retaining methylation data, wgbstools represents fragment-level data at a 100-fold compression, allowing for fast access and anonymization of human genetic information. We also introduce machine learning algorithms for genomic segmentation of homogenous methylation, computational tools for merging samples from the same cell types, or mixing samples from multiple cell types at various proportions, statistical tools for identification of differentially methylation regions, and joint fragment-level analysis of genetic and epigenetic information. These tools allow for efficient state-of-the-art analysis of DNA methylation data, while considering multiple samples at a fragment-level resolution.

We also introduce UXM, a computational suite for fragment-level analysis and deconvolution of next-generation sequencing data. UXM allows fragment-level classification of a reference atlas or cell-free DNA fragments as (U)nmethylated, mi(X)ed, or (M)ethylated, depending on the single-molecule methylation patterns at multiple neighboring CpGs. We then use non-negative least-square regression across multiple markers to infer the cellular composition of circulating cell-free DNA, outperforming state-of-the-art results from DNA methylation arrays.

These computational suites facilitate the accurate analysis of circulating cell-free DNA fragments, with implications for clinical diagnosis of liquid biopsy data.

184

Finding biomarkers for immunotherapy in lung cancer patients by cell-type specific expression

Yossef Glantzspiegel[1], Jair Bar[2], Irit Gat-Viks[1], Iris Kamer[2]

[1]*School of Molecular Cell Biology and Biotechnology, Tel Aviv University, Israel*; [2]*Institute of Oncology, Sheba Medical Center, Israel*

Lung cancer is the number one cause of cancer related death worldwide. Most cases are diagnosed at an advanced stage of disease, when treatment goal is essentially palliative. The most common type of lung cancer is non-small cell lung cancer (NSCLC). The major advances in the treatment of advanced NSCLC are the advent of targeted agents for tumors with driver mutations and immunotherapy. Immune checkpoint inhibitors (ICI) have been proven to be of benefit for advanced NSCLC and are currently part of the standard of care for these patients. However, it isn`t clear if ICI works better as a prior treatment to surgery (neoadjuvant treatment) or a consolidative treatment. In order to give better treatment we need to find the best way to administer it, identify biological markers that could predict treatment success, and understand the biological mechanisms of immunotherapy. By treating patients with immune checkpoint blockade prior to surgery (neoadjuvant treatment) with different intervals until the surgery, using Nanostring`s GeoMX digital spatial profiler we obtained gene and protein expression of cancer cell or immune cell regions from the biopsy of each patient. Using bioinformatics analysis, we identified potential biological markers for immunotherapy and assessed the effectiveness of different schedules of neoadjuvant treatment.

185

Stabilizing gene symbols for the clinical genome

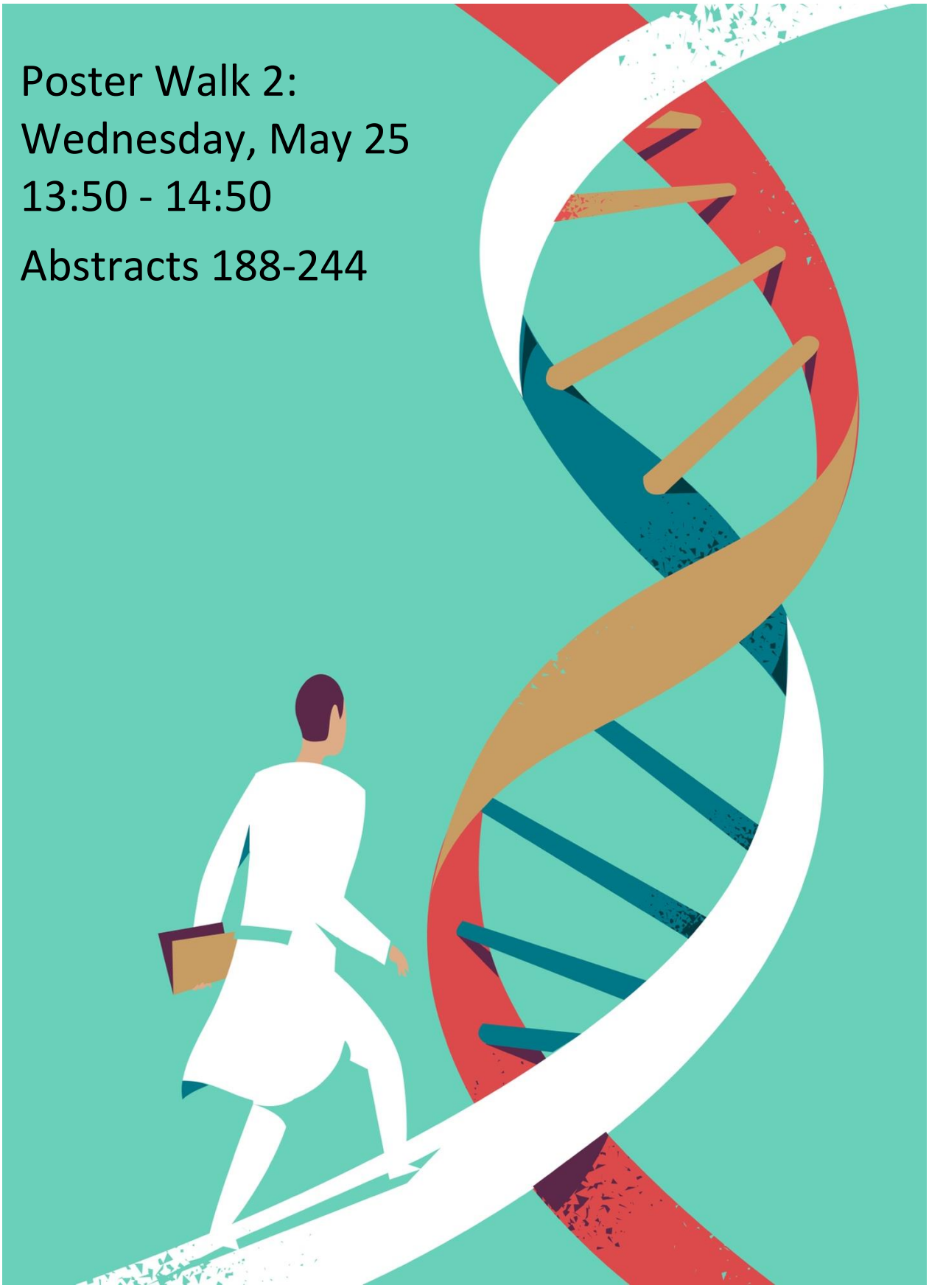
Liora Vilmovsky[1], Bryony Braschi[1], Kristian Gray[1,2], Tamsin Jones[1], Ruth Seal[1,2], Elspeth Bruford[1,2]

[1]*HGNC, EMBL-EBI, UK*; [2]*Department of Haematology, University of Cambridge, UK*

The HUGO Gene Nomenclature Committee (HGNC) has been responsible for approving human gene symbols and names for over 40 years; all of our approved nomenclature, stable HGNC gene IDs and catalogued gene aliases, are available at www.genenames.org. The rise of genomic medicine means that gene symbols are no longer confined to the scientific literature but now feature regularly in clinical practice, patient information, home test kits and the media, meaning it is more important than ever that genes are identified unambiguously. Confusion can arise from the use of alternative unapproved gene symbols but can also arise when gene symbols change. Clinical geneticists have stressed the need for symbol stability for effective clinical reporting. Our current Nomenclature Guidelines, discussed in Nature Genetics (Bruford et al., 2020), emphasise our commitment to stable gene symbols and describe the limited scenarios where changes could happen – including changing misleading symbols and updating uninformative placeholders. We are currently reviewing our gene nomenclature to address any problems and ensure long-term symbol stability. This includes checking the language used in gene names and symbols is appropriate for discussion with patients and their families. Symbols that pass our rigorous review are labelled with a 'stable' tag on our website. In this presentation we describe our gene symbol stabilisation process, the criteria for symbol updates and report our progress stabilising symbols for genes in our priority targets sets: genes associated with Covid-19 and genes with disease associations verified by the Gene Curation Coalition (thegencc.org).

Poster Walk 2:
Wednesday, May 25
13:50 - 14:50

Abstracts 188-244



188

Landscape of adenosine-to-inosine RNA recoding across human tissues

Erez Levanon[1], Orshay Gabay[1], Eli Eisenberg[2]

[1]*Life Sciences, Bar-Ilan University, Israel*; [2]*School of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel*

RNA editing by adenosine deaminases changes the information encoded in the mRNA from its genomic blueprint. Editing of protein-coding sequences can introduce novel, functionally distinct, protein isoforms and diversify the proteome. The functional importance of a few recoding sites has been appreciated for decades. However, systematic methods to uncover these sites perform poorly, and the full repertoire of recoding in human and other mammals is unknown. Here we present a new detection approach, and analyze 9125 GTEx RNA-seq samples, to produce a highly-accurate atlas of 1517 editing sites within the coding region and their editing levels across human tissues. Single-cell RNA-seq data shows protein recoding contributes to the variability across cell subpopulations. Most highly edited sites are evolutionary conserved in non-primate mammals, attesting for adaptation. This comprehensive set can facilitate understanding of the role of recoding in human physiology and diseases.

189

Accurate age prediction from blood using a small set of DNA methylation sites with a non-parametric cohort-based GPR model

Miri Varshavsky[1], Gil Harari[1,2], Daniel Neiman[2], Benjamin Glaser[3], Yuval Dor[2], Ruth Shemer[2], Tommy Kaplan[1,2]

[1]*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel*; [2]*Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada,; Hadassah Medical Center and Faculty of Medicine, The Hebrew University of Jerusalem, Israel*; [3]*Depatment of Endocrinology and Metabolism, Hadassah Medical Center and Faculty of Medicine, The Hebrew University of Jerusalem, Israel*

Chronological age prediction from DNA methylation sheds light on human aging, with indications of poor health and predicted lifespan, using a simple blood test. Previous studies developed DNA methylation clocks based on linear regression models of DNA methylation array data. While accurate, these models are limited to fixed-rate changes in methylation levels across age. Moreover, the high cost of Illumina BeadChip arrays, compared to targeted PCR sequencing, hinders widespread utility of such predictors.

Here, we present an AI-based alternative termed GP-age, that uses a non-parametric Gaussian process regression approach with a large cohort of blood-derived methylomes. Given a new blood sample, methylation levels are computed and compared to ≥11,000 samples, across a minimal pre-defined set of CpG sites. The cohort samples are then weighted and integrated to predict a chronological age. Using only 30 CpG sites, our approach outperforms state-of-the-art methylation clocks, including both parametric and cohort-based models, with an average accuracy of less than 4 years and a median error of 2 years.

Finally, we applied our model to next-generation sequencing data, by averaging neighboring CpG sites using a Laplace kernel, yielding highly accurate predictions. Overall, our new method provides an accessible alternative to current array-based methylation clocks, with future applications in aging research, forensic profiling, and monitoring epigenetic processes in transplantation medicine and cancer.

191

Genomic alterations panel of Israeli pediatric cancer in the era of personalized medicine

Ruty Shai[1], Hadar Cohen[1,2], Michal Yalon[1], Iris Kventsel[1], Michal Lurye-Hazut[1], Shani Caspi[1], Amos Toren[1,2]

[1]*Pediatric Hemato-Oncology, Sheba Medical Center, Israel*; [2]*Sackler School of Medicine, Tel Aviv University, Israel*

Background: Personalized medicine is moving towards a first-line strategy for cancer patients. The genetic mapping enables the treating physician to make an accurate treatment decision. Growing evidence shows that cancer types affecting children have unique biological characteristics that can affect prognosis and therapy. In this study we assembled the first Israeli pediatric tumor genomic alteration panel. Methods: Tumors genetically profiled in 97 pediatric patients with solid, hematologic and CNS malignancies were reviewed. We compiled a database with genetic variations for a meta-analysis of the most common mutated genes, positions overall and in each tumor type and classification. Results: The genetic testing detected 386 different variants.133 variants were detected in our cohort. In 10 genes, mutations repeated 8 or more times. The most common mutated gene in the Israeli cohort was P53. Nine of the most common variants in the cohort are known to be pathogenic. Four of them may be targeted by drug therapy. 7 unique gene variants were detected in ALL. On the other hand, the most common mutated genes in osteosarcoma were not found in our cohort. Conclusions: Genetic data is becoming an inherent part of multidisciplinary care of children. However, not enough is known about the genetics of pediatric cancers and even less so about the Israeli pediatric population. This study shows that indeed genomic variation which underwrites susceptibility to disease, and response to therapy not only differs in different tumor types, but also in different populations including the Israeli pediatric cancer population.

192

The role of modifier genes in embryo and tumor development - BRCA1 as a model

Hala Kassis[1], Dvora Kidron[3], Galia Tsarfaty[2], Judith Horev[1], Ilan Tsarfaty[1]

[1]*Department of Clinical Microbiology and Immunology, Sackler School of Medicine, Tel Aviv University, Israel*; [2]*Department of Diagnostic Imaging, The Chaim Sheba Medical Center, Tel Hashomer, Israel*; [3]*Department of Pathology, Meir Medical Center, Kfar Saba, Israel*

The Breast Cancer 1 protein (BRCA1) is a tumor suppressor involved in essential cellular functions necessary for genome integrity and embryonic development. New evidence demonstrates that the crosstalk between BRCA1, MET, and specific inherited driver modifier genes (IDMGs) plays a significant role in embryo and tumor development and response to therapy.

We studied the effect of the genetic background on homozygous BRCA1 inducing embryonic lethality. Embryonic ultrasound follow-up demonstrates early embryonic death in 01XC4 homozygous BRCA1 line. Pathological analysis indicates that all the empty sacs contain placenta and no evidence of erythrocyte. Dead newborn mice developed severe hypoxia and expanded facial space. In 01XC9, most empty sacks have erythrocytes signifying that embryonic lethality occurred at an advanced pregnancy stage. The dead newborn mouse showed enlargement in the neural tube.

To study the role of BRCA1, we established a CRISPR/Cas9 KO. The cells were subjected to single-cell morphokinetic analysis TASC developed in our lab. The results demonstrate that BRCA1 plays a significant role in cell motility and morphology. To our surprise when the cells express BRCA1 gene at a low level of 10-30% the motility of the cells increases, and when the expression levels of BRCA1 are eliminated the motility of the cells decreases. These results can explain what drives BRCA1 mutation carriers to differentially develop metastasis

Meta-analyses of human BRCA1-candidate modifier genes yield 508 published candidate genes. Several knowledge-based bioinformatics analyses and survival analyses were utilized to rank these genes` importance in BC progression. Ten IDMGs serve as an excellent prognostic factor together with BRCA1.

193

Multiple copies of microRNA binding sites in long 3'UTR variants regulate axonal gene expression

Ariel Ionescu[1], Luba Faeberov[1], Yazeed Zoabi[1], Guy Shapira[1], Eran Perlson[1,2], Noam Shomron[1,2,3]

[1]*Sackler Faculty of Medicine, Tel Aviv University, Israel*; [2]*Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel*; [3]*Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel*

Rapid responses to changes over space and time within highly polarized cells, such as motor neurons, depend on local translation and post-transcriptional regulation in distinct subcellular compartments such as axons. MicroRNAs (miRNAs) can regulate this process in a mechanism that is not yet fully understood. Here, using live cell imaging and RNA sequencing analysis, we demonstrated how miRNAs can differentially control hundreds of transcripts at the subcellular level. We confirmed that the length of the miRNA target-sequence regulates both mRNA stability and protein translation rates; longer seed regions have an increased inhibitory effect. Interestingly, transcriptome analysis of the motor neuron subcellular compartments did not reveal any differences in seed length between axonal and somata mRNAs. However, we recognized that long 3'UTR mRNA variants are enriched in axons and contain more sequence repeats, particularly for binding a few specific miRNA. Finally, as proof of concept, we demonstrated that the long 3'UTR mRNA variant of the motor protein kif5b is enriched explicitly in motor neuron axons and contains multiple sequence repeats for binding miR-129-5p. This subsequently results in the differential post-transcriptional regulation of kif5b and its synthesis in axons. Thus, we suggest that the number of miRNA binding sites at the 3'UTR of the mRNA, rather than the miRNA seed length, regulates the axonal transcriptome. These new insights clarify how globally expressed miRNA can adaptively interact with their targets at the subcellular level, while possibly driving multiple cellular functions.

194

Characterization of miR-126 5p role in ALS pathology

Lior Ankol Redlich[1], Roy Maimon[1], Ariel Ionescu[1], Nimrod Rotem[1], Tal Gradus Pery[1], Eran Perlson[1]

Department of Physiology and Pharmacology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Amyotrophic lateral sclerosis (ALS) is an adult-onset neurological disease characterized by muscle atrophy and motor neuron degeneration. Despite some progress, there is currently no effective treatment available for ALS. Although the disease`s etiology is not fully understood, it involves a non-cell-autonomous mechanism and alterations in RNA metabolism. In an unbiased screen for miRNAs, we identified the downregulation of miR-126 5p levels in ALS models` motor neuron axons and muscles. Overexpression of miR126-5p in primary MNs cultures from SOD1G93A and TDP-43 ALS mice models increases axon growth. Then, using a unique compartmental co-culture system that model the motor unit, we showed that miR1265p rescue NMJ activity in SOD1G93A and TDP-43 ALS mice models.

Moreover, using in vivo models, we demonstrated that overexpression of miR126-5p in the spinal cord of SOD1G93A mice reduced MN loss by inhibition of the activated Caspase 3. Further, we identified and verified some miR-126-5p targets that encode vital regulatory proteins. Finally, preliminary data of overexpressing miR-126 5p in SOD1G93A mice suggest improved behavioral motor activity. These findings support the idea that miR126 can regulate key processes impaired in ALS pathology, including axon growth and muscle innervation, and suggest that miR126 can be targeted for therapeutic development.

195

The effect of a GRIN2D variant in developmental and epileptic encephalopathy in a CRISPR/Cas9 mouse model and iPS cells

Danielle Galber[1], Jolan Nassir[2], Eden Peled[1], Mor Yam[2], Moore Lev[1], Tatiana Rabinski[3], Sivan Sagiv[3], Gad Vatine[3], Karen B. Avraham[1], Moran Rubinstein[2]

[1]*Department of Human Molecular Genetics and Biochemistry, Tel Aviv University, Israel*; [2]*Department of Human Molecular Genetics and Biochemistry & Goldschleger Eye Research Institute, Tel Aviv University, Israel*; [3]*Department of Physiology and Cell Biology, Ben-Gurion University of the Negev, Israel*

Developmental and epileptic encephalopathies (DEEs) are devastating disorders that involve global developmental delay and intractable epilepsy, with patients suffering from a wide range of cognitive, motor and electrophysiological impairments. Recent advances in technology lead to the development of next generation sequencing, and with it, the discovery of many genes associated with DEEs. Specifically, a de-novo variant (c.1999GA) was identified in GRIN2D that has been found to cause early onset epilepsy, intellectual disabilities and hypotonia. GRIN2D encodes GluN2D, a subunit of the NMDA receptor that plays a key role in excitatory synaptic transmission and long term plasticity. Cellular studies have shown that this variant results in increased glutamate and glycine potency, increased channel open probability, reduced receptor sensitivity to negative modulation, prolonged channel deactivation time course, dendritic swelling and neuronal cell death.

In order to understand the implications of this variant, we developed a mouse model using CRISPR/Cas9 technology and measured disease progression. This model mimicked the human phenotype, exhibiting motor deficits, severe epileptiform abnormalities and premature death. Additionally, we created two iPS cell lines from fibroblasts taken from a patient and a healthy parent. Therefore, we were able to create two independent models that represent this rare genetic disorder, enabling us the ability to further examine the effect of this variant.

We are currently working on a range of approaches to enhance our understanding of the complex neuronal and network mechanism underlying GRIN2D pathology, including high-throughput drug screening, molecular dynamics simulations, base editing technologies, mRNA manipulation therapies and electrophysiology.

197

Interpretation of human complex diseases by proteome-wide association study (PWAS)

Roei Zucker[1], Nadav Brandes[1], Nathan Linial[1], **Michal Linial**[2]

[1]*School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel*;
[2]*Department of Biological Chemistry, The Hebrew University of Jerusalem, Israel*

Over the last two decades, GWAS has become a canonical tool for exploratory gene-phenotype associations. The main limitation of GWAS that hinders its success is the poor interpretability and the low statistical power of individual genetic variants. PWAS (Proteome-Wide Association Study) is a new method for detecting protein-coding genes associated with phenotypes through the alteration of their protein function. It is a gene-based approach that aggregates all variants of a gene to assess the impact of function through machine-learning and probabilistic models. PWAS also captures complex modes of heritability, including recessive inheritance. Using genotyping and imputed data from 330K individuals from the UK Biobank cohort, over 100 prominent phenotypes were studied and revealed about 22% of significant PWAS gene associations (2,743 of 12,444) that were missed by GWAS. Moreover, the discovered PWAS genes for human traits and complex diseases are supported by a concrete molecular model. Interpretation of common diseases such as autoimmunity, cancer, and hypertension highlight new associations with plausible biological mechanisms. PWAS results emphasize the importance of recessive inheritance in complex human diseases and complement the variant-based method of GWAS. We conclude that PWAS recovers many overlooked causal protein-coding genes in a wide range of complex diseases and human trials.

200

A data-driven approach for predicting the impact of drugs on the human microbiome

Yadid Algavi[1], Elhanan Borenstein[1,2]

[1]*Sackler Faculty of Medicine, Tel Aviv University, Israel*; [2]*The Blavatnik School of Computer Science, Tel Aviv University, Israel*

Our gastrointestinal tract harbors a flourishing and diverse community of microorganisms, collectively known as the human gut microbiome. Over the past decade, we came to appreciate a complex and bidirectional interaction between the microbiome and numerous pharmaceuticals. Specifically, it has been shown that the microbiome can modulate the efficacy of various drugs, but also that many medications can negatively impact its composition, depleting beneficial strains and thereby causing gastrointestinal side effects. While recent studies have shed light on several such interactions, a comprehensive and complete understanding of this microbiome-drug interplay is still mostly lacking, and the incorporation of such interactions into clinical practice is currently still out of reach.

Towards this end, we developed a data-driven approach that integrates chemical and bacterial genomic information that systematically predicts the interaction between every drug and microbiome member. This framework can accurately predict the results of in-vitro pairwise drug-microbe experiments and drug-induced dysbiosis in complex ecologies, both in animal models and clinical trials. Applying this methodology, we comprehensively map the interactions between pharmaceuticals and bacteria and demonstrate that the anti-microbial properties of medications are tightly related to adverse effects. This computational approach will unlock the development of precision medicine approaches and microbiome-based interventions to improve the therapeutic outcome and minimize side effects.

201

A genomic duplication of 83 Kbp is associated with the Mammary-Digital-Nail (MDN) syndrome

Golan Nadav[1], Ayalla Fedida[1], Meir Shamay[2], Leonid Kogan[3], Tameema Attallah Moadi[1], Mary Noffi Bahoum Noffi Bahoum[4], Hector I Cohen[5], Alexander Kaplun[6], Rana Nasara Nasara[5], Limor Kalfon[1], Malte Spielman[7], Tzipora C. Falik-Zaccai Falik-Zaccai[1,2]

[1]*Institute of Human Genetics, Galilee Medical Center, Israel*; [2]*Azrieli Faculty of Medicine, Bar-Ilan University, Israel*; [3]*Department of Plastic Surgery, Galilee Medical Center, Israel*; [4]*Pediatric Endocrinology Service, Clalit, Health Service, Israel*; [5]*Department of Pathology, Galilee Medical Center, Naharia, Israel, Israel*; [6]*Variantyx, Variantyx Inc, USA*; [7]*Institute of Human Genetics, University of Lbeck, Germany*

Background

Mammary Digital Nail syndrome (MDN) is linked to a 4.3 Mb interval on chromosome 22q12.3-13.1. We aimed to reveal the causative genetic variant, and molecular mechanism underlying the MDN phenotype. An Irish family with onychodysplasia and a Druze family with MDN were recruited to the study.

Results

Whole genome sequencing (WGS) revealed a novel heterozygous genomic duplication of 83 Kbp, within the linked interval on chromosome 22 in affected individuals from both families. This duplication contains two ORFs: one encoding the potassium channel KCNJ4 and the other encoding an inositol lipid phosphatase pseudogene (TPTEP2). Relative qPCR confirmed an autosomal dominant segregation pattern within the MDN Druze family. An overlapping duplication including KCNJ4 gene and only the first out of five exons of TPTEP2, was detected in the Irish family presenting ODP. RT-qPCR and WB analysis revealed a significant increase in the transcript as well as protein level of KCNJ4 in skin and breast biopsies derived from affected females. Transcriptome analysis identified dysregulated pathways specific in breast from affected females

Conclusions

A genomic duplication on chromosome 22q12.3-13.1, is linked to the MDN and ODP phenotypes in two unrelated families, strongly suggesting that it is the MDN causative variant. The significant higher abundance of KCNJ4, encoded in the duplication, in the MDN tissues may suggests it`s involvement in the molecular mechanism causing the diseases phenotype. Elucidation of the mechanism of pathogenicity may reveal novel insights on the embryonic development of digits and nails, and pubertal breast development.

202

High-throughput multi-dimensional tracking of chromosome reconfiguration during double-strand break repair

Yael Shalev Ezra[1], Alon Saguy[1], Lucien Weiss[1,2], Nadav Opatovski[1], Reut Orange Kedem[1], Onit Alalouf[1], Yoav Shechtman[1]

[1]*Department of Biomedical Engineering and Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion - Israel Institute of Technology, Israel*; [2]*Department of Engineering Physics, Polytechnique, Canada*

Nuclear functions are influenced by DNA dynamics. An example is the repair process of the extremely genotoxic DNA double-strand-break (DSB) by homologous recombination (HR) that maintains chromatin integrity. During HR the DSB site and the intended template undergo spatial reconfiguration allowing close proximity for repair.

We utilized the mating-type switching process in Saccharomyces Cerevisiae as a model system to visualize DSB repair, and constructed strains that are stably labeled with LacO-LacI-eGFP and TetO-TetR-mCherry next to the MAT and HMLα loci, respectively. We integrated a mating-type switching induction system that includes the expression of a fluorescent reporter upon completion of repair. Loci dynamics visualization was done with fluorescence microscopy and multicolor point-spread function (PSF) engineering, where we added optical elements in either a widefield microscope or in an image flow cytometry (IFC), that modulated the emitted light as a function of the emitter depth and color to give 3D multicolor information in gray scale camera. We developed an end-to-end analysis pipeline to track the 3D distance between the differently labeled loci per cell over time.

Our results include single-cell 3D trajectories as well as high throughput imaging (~10^3 cells/minute), revealing shorter distances of ~2-fold between the loci than previously reported. Specifically, the high throughput aspect of IFC was found to be particularly useful, since only ~10% of the population undergoes switching. Our work demonstrates the utility of optical wavefront shaping, together with complementary image analysis, as a tool to study the DSB repair process in live cells.

203

Whole-genome sequencing reveals that variants in the Interleukin 18 Receptor Accessory Protein 3`UTR protect against ALS

Chen Eitan[1,2], Aviad Siany[1,2], Elad Barkan[3], Tsviya Olender[1], Kristel R van Eijk[4], Matthieu Moisse[5,6], Sali M K Farhan[7,8], Yehuda M Danino[1,2], Eran Yanowski[1,2], Hagai Marmor-Kollet[1,2], Natalia Rivkin[1,2], Nancy Sarah Yacovzada[1,2,3], Shu-Ting Hung[9,10,11], Johnathan Cooper-Knock[12], Chien-Hsiung Yu[13,14], Cynthia Louis[13,14], Seth L Masters[13,14], Kevin P Kenna[4], Rick A A van der Spek[4], William Sproviero[15], Ahmad Al Khleifat[15], Alfredo Iacoangeli[15], Aleksey Shatunov[15], Ashley R Jones[15], Yael Elbaz-Alon[1], Yahel Cohen[1,2], Elik Chapnik[1], Daphna Rothschild[3,16,17], Omer Weissbrod[3], Gilad Beck[18], Elena Ainbinder[18], Shifra Ben-Dor[18], Sebastian Werneburg[19], Dorothy P Schafer[19], Robert H rown Jr[20], Pamela J Shaw[12], Philip Van Damme[5,6,21], Leonard H van den Berg[4], Hemali Phatnani[22], Eran Segal[3], Justin K Ichida[9,10,11], Ammar Al-Chalabi[15,23], Jan H Veldink[4], Eran Hornstein[1,2]

[1]*Department of Molecular Genetics, Weizmann Institute of Science, Israel*; [2]*Department of Molecular Neuroscience, Weizmann Institute of Science, Israel*; [3]*Department of Computer Science And Applied Math, Weizmann Institute of Science, Israel*; [4]*Department of Neurology, Utrecht University, Netherlands*; [5]*Department of Neurosciences, KU Leuven - University of Leuven, Belgium*; [6]*Laboratory of Neurobiology, VIB, Center for Brain & Disease Research, Belgium*; [7]*Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital and Harvard Medical School, USA*; [8]*Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, USA*; [9]*Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine, University of Southern California, USA*; [10]*Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, USC, USA*; [11]*Zilkha Neurogenetic Institute, Keck School of Medicine of the University of Southern California, USA*; [12]*Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, UK*; [13]*Inflammation Division, The Walter and Eliza Hall Institute of Medical Research, Australia*; [14]*Department of Medical Biology, University of Melbourne, Australia*; [15]*King's College London, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology & Neuroscience, UK*; [16]*Department of Developmental Biology, Stanford University, USA*; [17]*Department of Genetics, Stanford University, USA*; [18]*Department of Life Sciences Core Facilities, Weizmann Institute of Science, Israel*; [19]*Department of Neurobiology, Brudnick Neuropsychiatric Research Institute, University of Massachusetts Chan Medical School, USA*; [20]*Department of Neurology, University of Massachusetts Medical School, USA*; [21]*Department of Neurology, University Hospitals Leuven, Belgium*; [22]*Center for Genomics of Neurodegenerative Disease, New York Genome Center, USA*; [23]*Denmark Hill, King's College Hospital, UK*

The noncoding genome is substantially larger than the protein-coding genome but has been largely unexplored by genetic association studies. Here, we performed region-based rare variant association analysis of 25,000 variants in untranslated regions of 6,139 amyotrophic lateral sclerosis (ALS) whole genomes and the whole genomes of 70,403 non-ALS controls. We identified interleukin-18 receptor accessory protein (IL18RAP) 3` untranslated region (3`UTR) variants as significantly enriched in non-ALS genomes and associated with a fivefold reduced risk of developing ALS, and this was replicated in an independent cohort. These variants in the IL18RAP 3`UTR reduce mRNA stability and the binding of double-stranded RNA (dsRNA)-binding proteins. Finally, the variants of the IL18RAP 3`UTR confer a survival advantage for motor neurons because they dampen neurotoxicity of human induced pluripotent stem cell (iPSC)-derived microglia bearing an ALS-associated expansion in C9orf72, and this depends on NF-κB signaling. This study reveals genetic variants that protect against ALS by reducing neuroinflammation and emphasizes the importance of noncoding genetic association studies.

204

Transcriptional reprogramming by TP53 oncogenic mutations in breast carcinogenesis

Leslie Cohen[1], Libi Marchenkov Flam[1], Reut Perez[1], Bracha Zukerman Attia[1], Sivan Shlisselberg[1], Miriam Tannenbaum[1], Avital Sarusi Portuguez[3], Olga Loza[2], Ofir Hakim[1]

[1]*Life Sciences, Bar-Ilan University, Israel*; [2]*Life Sciences, Tel Aviv University, Israel*; [3]*Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Sciences, Israel*

TP53 tumor suppressor gene is mutated in more than 30% of breast cancer cases. Of these, hotspot mutations (R273H, R248W, R175H, R249S) affect the p53 DNA binding domain and are thought to provide additional cancer-related capabilities. Most efforts to understand the direct regulatory functions of mutant p53 were performed in transformed cells with different genetic backgrounds. Therefore, the precise contribution of specific p53 mutations to the transcriptional program driving the malignant cellular phenotypes and breast cancer progression is still not determined. Using an isogenic cell line to study carcinogenesis may reveal transcriptional changes caused by the p53 hotspot mutations, which may contribute to cancerous phenotypes. We analyzed MCF10A, immortalized breast epithelial cells, deficient for p53, and expressing p53 mutants R273H, R248W, R175H, or R249S. RNA-seq measurements showed that core p53 programs are impaired in all cell lines, but more so in mutant cells than in p53-null cells. These genes were related to different cancer-associated functions, but mutant cells showed “anti-cancer” phenotypes including lower proliferation compared to wild-type and p53 null cells. Moreover, ATAC-seq and p53 ChIP-seq analysis show that ~80% of WT p53 binding loci have low accessibility that remained stable across the p53 mutants and KO cell lines. Interestingly, R273H p53 bound chromatin 2-fold more sites than WT, while R249S p53 lost 80% of it, and R175H p53 lost it entirely. p53 binding loci were co-localized with differentially expressed genes in the same topologically associated domain, suggesting that these genes are targeted by p53.

206

Geometric and quantitative characterization of the transcriptional heterogeneity in Wilms' tumors

Yaron Trink[1], Achia Urbach[2], Benjamin Dekel[3], Peter Hohenstein[4], Jacob Goldberger[1], Tomer Kalisky[1]

[1]*Faculty of Engineering and Bar-Ilan Insititute of Nanotechnology and Advanced Materials (BINA), Bar Ilan University, Israel*; [2]*The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, Israel*; [3]*Pediatric Stem Cell Research Institute and Division of Pediatric Nephrology, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Israel*; [4]*Department of Human Genetics, Leiden University Medical Center, Netherlands*

Wilms tumors are cancers of the kidney that occur primarily in young children. They are histologically diverse, with variable contributions of blastemal, stromal, and epithelial elements. This heterogeneity is poorly understood both in terms of its origin and characterization. Computational methods to better classify tumors into subtypes and assess the progression of tumors in individual patients is a promising avenue in which to better understand and characterize this heterogeneity. We therefore obtained gene expression measurements from (Wegert et al., 2015), and applied statistical techniques to uncover patterns in the data.

We observe that the tumors fill a triangle shaped continuum along the first two principal components. We compute the vertices (cellular archetypes) of the best fitting triangle using Pareto Task Inference (Hart et al., 2015), and use gene enrichment to identify the archetypes as having stromal, blastemal and epithelial characteristics. Next we use single cell deconvolution algorithms (Frishberg et al., 2019), to model the bulk tumors in terms of their underlying single cell composition. Finally, we use Latent Dirichlet Allocation (Blei et al., 2003; Peter Carbonetto, Kevin Luo, Kushal Dey, 2021), to infer the parameters of a multinomial distribution over topics for each individual tumor, and use Bayes' theorem to infer the biological identity of the topics. We observe that the global geometry of the tumors, relative to PCA, is preserved, and that these models confer additional interpretability by providing a probabilistic distribution for each tumor over each of the underlying cellular elements.

207

Blood transcriptional response to treatment-resistant depression during electroconvulsive therapy

Ifat Israel-Elgali[1,2], Libi Hertzberg[2,3,4], Guy Shapira[2], Aviv Segev[2,3], Israel Krieger[2,3], Uri Nitzan[2,3], Yuval Bloch[2,3], Nir Pillar[2], Ori Mayer[2], Abraham Weizman[2,5], David Gurwitz[1,2], Noam Shomron[1,2]

[1]*Sagol School of Neuroscience, Tel Aviv University, Israel*; [2]*Sackler Faculty of Medicine, Tel Aviv University, Israel*; [3]*Shalvata, Shalvata Mental Health Center, Israel*; [4]*Department of Physics of Complex Systems, Weizmann Institute of Science, Israel*; [5]*Geha, Geha Mental Health Center, Israel*

Background: Selective serotonin reuptake inhibitors (SSRIs) are currently the first-line antidepressant drug treatment for major depressive disorder (MDD). Treatment-resistant depression (TRD), defined as failure to achieve remission despite adequate treatment, affects ~30% of persons with MDD. The current recommended treatment for TRD is electroconvulsive therapy (ECT), while ketamine is an experimentally suggested treatment.

Results: This study aimed to elucidate the transcriptional differences in peripheral blood mononuclear cells (PBMC) between individuals with TRD and a control group without a psychiatric illness; and between patients with TRD, treated with either standard antidepressant drugs alone, or in combination with ECT or ketamine. Additionally, PBMC transcriptomics were compared between treatment responders, following completion of their treatment protocols. Total RNA was extracted from PBMC of the TRD group at two time points, and RNA and miRNA expression were profiled. Multiple mRNAs and miRNAs were found to be modified, with two protein coding genes, FKBP5 and ITGA2B, which are up- and downregulated, respectively; and several miRNAs have shown changes following successful ECT treatment. Further analysis demonstrated the direct functional regulation of ITGA2B by miR-24-3p.

Conclusions: Our findings suggest that PBMC expression levels of FKBP5, ITGA2B, and miR-24-3p should be further explored as tentative ECT response biomarkers.

208

The beneficial effect of ultra-low dose delta 9 – THC on neurodegeneration

Meitar Grad[1], Ifat Israel-Elgali[1,2], Guy Shapira[1,3], Lital Rachmany[1], Yosef Sarne[1], Noam Shomron[1,2,3]

[1]*Sackler Faculty of Medicine, Tel Aviv University, Israel*; [2]*Sagol School of Neuroscience, Tel Aviv University, Israel*; [3]*Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Israel*

Δ9-tetrahydrocannabinol (THC), the main psychoactive component in marijuana, has long been criticized for its harmful effects on brain development and function. However, in recent years evidence has emerged regarding the neuroprotective potential of this exogenous cannabinoid when applied under controlled measures. Regardless, a definite and efficient treatment protocol has not yet been formed, and the underlying mechanism for the protective effect of THC is uncertain. Sarne *et al.* revealed that an acute *i.p.* injection of ultra-low dose (ULD) THC (0.002mg/kg; 3-4 orders of magnitude lower than the conventional doses) improved cognitive performance among old (2-years-old) mice relative to vehicle-treated old mice, and brought their cognitive performance to a level equivocal to young (2-months-old) control mice. Interestingly, the same treatment with ULD-THC caused cognitive impairments in healthy young mice. These results suggest that the beneficial effects of THC are context- and dose-dependent.

To explore the effect of acute ULD-THC treatment on gene expression, we extracted total RNA from old and young mice hippocampi, and utilized next generation sequencing and rt-qPCR to delineate the changes in mRNA levels between THC- and vehicle- treated mice. These analyses revealed significant alterations in mRNA expression levels of glucocorticoid-related genes in the hippocampus of THC treated mice relative to controls.

Thus, we suggest that acute ULD-THC treatment may ameliorate neurodegenerative processes that are induced by dysregulation of the glucocorticoid pathway. Further experiments will aim to uncover molecular changes occurring in the brains of THC-treated mice, and broaden its applicability to major depression-related neurodegeneration and behavior.

210

Uncovering the role of palladin in the MET/HGF axis in breast cancers

Ori Mayer[1], Ohad Doron[2], Joshua Bugis[1], Or Megides[2], Ori Moskowitz[2], Ilan Tsarfaty[2], Noam Shomron[1]

[1]*Department of Cell and Developmental Biology, Tel Aviv University, Tel Aviv, Israel*; [2]*Department of Clinical Microbiology and Immunology, Tel-Aviv University, Israel*

Palladin (PALLD gene) is a structural protein widely expressed in mammalian tissues and has a pivotal role in cytoskeletal dynamics and motility in health and disease. Palladin has been linked to the progression of breast, pancreatic and renal cancers. The MET/HGF axis is a signaling pathway with a central part in embryonic development and cancer progression. In this study, we investigate in depth palladin`s contribution to cellular motility and characterize its part in the MET/HGF axis in breast cancer.

First, by analyzing gene coexpression, we determined that palladin is significantly correlated with MET/HGF signaling. Next, palladin deficient breast cancer cell line was developed using CRISPR/Cas9. Using live-cell imaging and tracking, we conclude that HGF induced motility is hindered in the absence of palladin. Further analysis of migration patterns implicates palladin as a major contributor to collective cell motility. Transcriptome and proteome analysis of $PALLD^{KO}$ cells revealed 11 differentially expressed genes which reinforced palladin`s centrality to cellular motility. Finally, analysis of the TCGA-BRCA cohort validated that HGF expression is in correlation with decreased survival of patients with tumors expressing high levels of palladin.

Our work strengthens the link between palladin and cellular motility and metastasis in cancer. More specifically, we present evidence for the dependence of migration induced by the MET/HGF axis, on an adequate level of functional palladin. Additionally, we offer an analysis of palladin`s precise contribution to cellular motility. Overall, these results support our ongoing hypothesis: interference with palladin`s expression, translation or function will reduce the metastatic spread.

211

Willingness to donate and to receive results: Should we return to donors with research findings?

Yehudit Cohen[1], Shimon Reisner[1], Adelina Ovcharenko[1]

Management, MIDGAM - Israel National Biobank for Research, Israel

Israel, like many countries, enters an era of large longitudinal studies, often genetic, in personalized medicine, conducted mainly as part of the National-Digital-Health-Projects. These studies are based on biosamples and annotated data (demographic, clinical, genetic, behavioral, etc.).

MIDGAM, The Israeli National Biobank for Research, promotes biomedical research in Israel, and as such is the best platform for such projects.

To assess population willingness to contribute to research studies, including willingness to donate and to receive results, 1,607 men and women (general population: N=922; ultra-orthodox sector: N=384; Arab sector: N=301) were asked to fill out a quantitative survey questionnaire. Participants were asked about their willingness to donate biosamples and data, and their opinions regarding receiving results including incidental findings. Specific focus was given to genetic data.

We found that within the general population 52% are willing to donate; most willing subjects are interested in receiving results (40% of all, 77% of those willing-to-donate). We also found that the main barriers to sample donation are related to concerns about data-leakage, privacy-violations and lack of understanding the purpose of donation. The small part of the population not interested in receiving findings, cites mainly reluctance to know.

To meet the needs inferred from these results, we adjust and update our consent forms and continually examine significance of research findings for clinical purposes. Findings with medical significance and therapeutic potential should be regularly updated and shared according to clinical guidelines. Furthermore, a professional committee will discuss the implications of incidental-findings and examination of donors` expectations.

212

Chromothripsis drives the evolution of gene amplification in cancer

Ofer Shoshani

Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel

Focal chromosomal amplification contributes to the initiation of cancer by mediating overexpression of oncogenes and to the development of cancer therapy resistance by increasing the expression of genes whose action diminishes the efficacy of anti-cancer drugs. We used whole-genome sequencing of clonal cell isolates that developed chemotherapeutic resistance to show that chromothripsis, the catastrophic shattering of a chromosome and random religation of its pieces, is a major driver of circular extrachromosomal DNA (ecDNA) amplification (also known as double minutes) through mechanisms that depend on poly(ADP-ribose) polymerases (PARP) and the catalytic subunit of DNA-dependent protein kinase (DNA-PKcs). Longitudinal analyses revealed that a further increase in drug tolerance is achieved by structural evolution of ecDNAs through additional rounds of chromothripsis. In situ Hi-C sequencing showed that ecDNAs preferentially tether near chromosome ends, where they re-integrate when DNA damage is present. Intrachromosomal amplifications that formed initially under low-level drug selection underwent continuing breakage–fusion–bridge cycles, generating amplicons more than 100 megabases in length that became trapped within interphase bridges and then shattered, thereby producing micronuclei whose encapsulated ecDNAs are substrates for chromothripsis. We identified similar genome rearrangement profiles linked to localized gene amplification in human cancers with acquired drug resistance or oncogene amplifications. We propose that chromothripsis is a primary mechanism that accelerates genomic DNA rearrangement and amplification into ecDNA and enables rapid acquisition of tolerance to altered growth conditions.

213

Assessment of methylation of 10 miRNA genes in clear cell renal cell carcinoma in early non-invasive diagnosis

Irina Pronina[1], Aleksey Burdyonny[1], Natalya Ivanova[1], Elena Filippova[1], Svetlana Lukina[1], Tatiana Kazubskaya[2], Vitaly Loginov[1]

[1]*Laboratory of Pathogenomics and Transcriptomics, Institute of General Pathology and Pathophysiology, Russia*; [2]*Department of Pathomorphology, FSBI N.N. Blokhin National Medical Research Center of Oncology, Russia*

The number of deaths in clear cell renal cell carcinoma (ccRCC) without metastases reaches 30-40% and with metastases is up to 90%. The methylation of promoter CpG islands of miRNA genes involved in the pathogenesis of ccRCC could be a possible marker for its early diagnosis. The purpose of this work was to assess the involvement of CpG methylation of miRNA genes in the development and progression of ccRCC and to identify markers for diagnosis at the early stages.

Methylation analysis was carried out by quantitative methyl-specific PCR on a set of 100 paired (tumor/normal) ccRCC samples, as well as 19 samples of the control group (without oncopathology). The significance of the results was assessed using the R statistic (Mann-Whitney U-test, Kolmogorov-Smirnov test, SPSS19; p 0.05).

We showed a statistically significant (p0.001) increase in the level of methylation of the MIR34C, MIR9-1/3, MIR129-2, MIR124-1/2/3, MIR130B genes in ccRCC samples compared to the adjacent tissue and control. Methylation of the MIR107, MIR130B, and MIR148A genes in ccRCC was studied for the first time. Using ROC analysis, a system for the early diagnosis of ccRCC (MIR34C, MIR9-1, MIR129-2, MIR124-3 and MIR130B) characterized by 91% sensitivity and 95% specificity, AUC=0.936 was compiled. Thus, 8 miRNA genes hypermethylated in ccRCC have been identified, which can be used as new biomarkers for early diagnosis of the disease.

The study was supported by the state task FGFU-2022-0007 of the Ministry of Science and Higher Education of the Russian Federation.

214

De novo mutation rates at the single mutation resolution in a human *HBB* gene region associated with adaptation and genetic disease

Adi Livnat[1], Daniel Melamed[1], Yuval Nov[2], Assaf Malik[3], Michael B. Yakass[4], Evgeni Bolotin[1], Revital Shemer[6], Edem K. Hiadzi[5], Karl L. Skorecki[7]

[1]*Institute of Evolution, University of Haifa, Israel*; [2]*Statistics, University of Haifa, Israel*; [3]*Bioinformatics Unit, University of Haifa, Israel*; [4]*Biochemistry, Cell and Molecular Biology, University of Ghana, Ghana*; [5]*Assisted Conception Unit, Lister Hospital and Fertility Centre, Ghana*; [6]*Faculty of Medicine, Technion, Israel*; [7]*Faculty of Medicine, Bar-Ilan University, Israel*

While it has been known that the mutation rate varies across the genome, previous estimates were based on averages across many genomic positions. We have developed a method to measure the de novo origination rates of target mutations of interest at target base positions and applied it to a 6-bp region in the human hemoglobin subunit beta (*HBB*) gene that contains the site of the hemoglobin S (HbS) mutation, as well as to the identical, paralogous hemoglobin subunit delta (*HBD*) region, in sperm cells from both African and European donors. The HbS mutation, which protects against malaria in heterozygotes and causes sickle cell anemia in homozygotes, is common in Africa and has served as a classic example of adaptation by random mutation and natural selection. Results showed, first, that mutation rate variation at the single mutation resolution is far higher and different than expected from previous studies. Mutation-specific rates varied between genes and between populations even when the mutations appear on the identical local genetic sequence. Second, the overall point mutation rate in the narrow *HBB* region studied is significantly higher in Africans than in Europeans, which is of interest given the importance of this region to adaptation and genetic disease. Finally, a combination of statistically significant observations showed that the malaria resistant HbS mutation originates de novo more frequently in *HBB* compared to the non-resistant but otherwise identical mutation in *HBD*, and in sub-Saharan Africans, who have been experiencing intense malaria pressure for many generations, than in Northern Europeans.

215

A negative exome is not the end of the story

Batel Terespolsky[1], Lina Berkun[1], Reeval Segel[1], Michal Barzily[1], Stanely Korman[2], David Zeevi[1], Omer Murik[1], Pinhas Renbaum[1], Ephrat Levy-Lahad[1]

[1]*Medical Genetics, Shaare Zedek Medical Center, Israel*; [2]*Metabolic Disease Unit, Shaare Zedek Medical Center, Israel*

Background

GSD3 is a glycogen accumulation disease causing damage to muscles, liver and heart, caused by biallelic pathogenic mutations in the AGL gene encoding Amylo 1,6 glucosidase. Family `H` has 2 children diagnosed with GSD3, exome sequencing identified one VUS, an intronic (+130bp) homozygous variant.

Methods

RT-PCR, qRT-PCR was used to analyze AGL expression. Long-PCR and NGS was used to confirm the structural variant and identify the exact molecular defect.

Results

We examined whether this lone intronic VUS was associated with changes in AGL expression. Different RT-PCR primer sets showed that exons 28-30 were not expressed in the affected children, whereas in the parents, all exons were transcribed. To determine whether this was due to exon skipping or a deletion of this region in the gDNA, we sequenced a 12Kb region including these exons. Combining long-range PCR, and NGS enabled us to sequence the entire segment and map the exon deletion with clear boundaries. This deletion results in a frameshift, loss of function mutation.

Conclusions

We identified a pathogenic mutation in an exome negative case, by combining expression analysis with NGS. This finding will now allow prenatal or preimplantation diagnosis and allow this couple to plan a family with healthy offspring. While we often rely on clinical exome sequencing, this study points out that in exome negative cases, when there is only a single suspected disease gene, it pays to invest additional means, to look deeper for the genetic basis of the disease.

217

Whole genome sequencing applied in familial hamartomatous polyposis identifies novel structural variations

Revital Kariv[1], Dvir Dahary[2], Yuval Yaron[1], Yael Petel-Galil[1], Mira Malcov[1], Guy Rosner[1]

[1]*Gastroenterology, IVFUunit, Tel Aviv Sourasky Medical Center, Israel*; [2]*Genome analysis, Geneyx Ltd., Israel, Israel*

Hamartomatous polyposis syndromes (HPS) are rare cancer-predisposing disorders including Juvenile polyposis (JPS), Peutz-Jeghers (PJS) and PTEN-Hamartomatous-Syndromes (PHS). Penetrant-mutations in corresponding genes (SMAD4, BMPR1A, STK11, PTEN and AKT1), are usually diagnosed by next-generation-sequencing gene panel (NGS-GP) for tailored surveillance and pre-implantation-testing for monogenic-disorders (PGT-M). Five probands with HPS phenotype, with no genetic diagnosis per genetic workup underwent whole-genome-sequencing (WGS) that identified structural genetic alterations: Two novel inversions in BMPRA1 and STK11, two BMPR1A-deletions, known as founder among Bukharan-Jews and BMPR1A micro-deletion. BMPR1A inversion was validated by "junction fragment" amplification and direct testing. PGT-M was performed by multiplex-PCR and enabled successful birth of non-carrier baby. WGS may be considered for HPS patients with no NGS-GP findings to exclude structural alterations.

219

Non-canonical DNA structures in the regulation of *Lhb* transcription

Tal Refael, Lilach Pnueli, Philippa Melamed

Faculty of Biology, Technion-Israel Institute of Technology, Haifa, Israel

Non-canonical DNA structures are often found at regulatory genomic regions and are gaining attention as potential drug targets for modifying gene expression. We identified, through sequence analysis, circular dichroism and chromatin immunoprecipitation (ChIP), a G-quadruplex (G4)/iMotif (iM)-forming sequence at the central untranscribed region of a novel enhancer of *Lhb* which encodes the gonadotropin, luteinizing hormone. This distal transcriptional enhancer is found in physical proximity to the *Lhb* promoter, carries characteristic histone modifications, and is transcribed to two eRNAs, one of which (eRNA2) directs open chromatin at the *Lhb* promoter and is required for *Lhb* transcription. Guide RNA-mediated recruitment of dCas9- KRAB or dCas9-VP64 to the enhancer, altered *Lhb* mRNA levels accordingly, confirming its activity. However, eRNA2 is not sufficient for *Lhb* transcription, and a distinct downstream lncRNA is required, which has G4/iM forming sequences at its promoter and first intron. These various G4/iM DNA structures are affected differentially by local conditions including hormonal treatments, and they appear to have diverse effects on transcription. HMGB2, which binds structured DNA, was found at several of these *loci* in ChIP experiments and bound the enhancer iM *in vitro*. Furthermore, HMGB2 knockdown affected some of the iMs and reduced dramatically levels of *Lhb* mRNA, the lncRNA and eRNA2, while the divergent eRNA1 levels increased. The involvement of DNA structures in determining the levels of these various regulatory non-coding RNAs and their effects on *Lhb* transcription, expands our understanding of how reproduction is regulated while emphasizing the complexity of transcriptional regulation via context-dependent cis-acting elements.

220

Adaptive sequencing using nanopores and deep learning of mitochondrial DNA

Artem Danilevsky, Avital Polsky, Noam Shomron

Department of Cell and Developmental Biology, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Nanopore sequencing is an emerging technology that reads DNA and identifies the various chemical modifications. “Adaptive sequencing” is an implementation of targeted sequencing, intended for use on the nanopore sequencing platform. In this study, we demonstrated an alternative method of software-based targeted sequencing that is performed in real time by combining nanopore sequencing and deep learning. Our results showed the feasibility of using deep learning for classifying signals from only the first 200 nucleotides in a raw nanopore sequencing signal format. This was further demonstrated by comparing the accuracy of our deep learning classification model across data from several human cell lines and other eukaryotic organisms. We used custom deep learning models and a script that utilizes a "Read Until" framework to target mitochondrial molecules in real time from a human cell line sample. This achieved an enrichment ability of 2.3-fold. In a series of very short sequencing experiments (10, 30, and 120 minutes), we identified genomic and mitochondrial reads with accuracy above 90%. The uniqueness of our method is its ability to distinguish two groups of DNA, even without a labeled reference. This contrasts with studies that required a well-defined reference, whether of a DNA sequence or of another type of representation. Additionally, our method showed higher correlation to the theoretically possible enrichment factor, compared to other published methods. We believe that our results will lay the foundation for rapid and selective sequencing using nanopore technology, and will pave the way for clinical applications that use nanopore sequencing data.

221

A meta-analysis study of the universality of gut microbiome-metabolome associations

Efrat Muller[1], Elhanan Borenstein[1,2,3]

[1]*The Blavatnik School of Computer Science, Tel Aviv University, Israel*; [2]*Sackler Faculty of Medicine, Tel Aviv University, Israel*; [3]*Santa Fe Institute, Santa Fe Institute, USA*

Microbiome-metabolome studies of the human gut have been gaining popularity in recent years, mostly due to accumulating evidence of the interplay between gut microbes, metabolites, and host health. Yet, it remains unclear whether findings of microbiome-metabolite associations identified in any single study carry over to other studies or cohorts, and how robust are microbiome-metabolites links.

Here, we performed a comprehensive meta-analysis to identify human gut metabolites that are 'universally' well-predicted by the gut microbiota. To this end, we processed data from 1733 samples from 10 independent human gut microbiome-metabolome studies, and implemented a machine learning pipeline to predict metabolite levels in each dataset based on the microbiota`s composition. Comparing the predictability of each metabolite across datasets, we found 97 universally well-predicted metabolites. These include metabolites involved in important microbial pathways such as bile acid transformations and polyamines metabolism. Importantly, however, other metabolites exhibited large variation in predictability across datasets, suggesting a cohort- or study-specific microbiome-metabolite relationship. Furthermore, models trained on the control group of a study, occasionally failed to transfer to the disease group of the same study, indicating a shift in microbial metabolism in disease-associated dysbiosis.

Accompanying this study, we are distributing an expanded user-friendly dataset collection of paired microbiome-metabolome data, publically available here: https://github.com/borenstein-lab/microbiome-metabolome-curated-data. Combined, this data resource, as well as our meta-analysis findings, can allow researchers to put identified microbially-associated metabolites within the context of other studies, facilitate the discovery of other robust microbe-metabolite associations, and perform rigorous data analysis on this unique multi-omic data.

222

Gene correction cannot repair aberrant epigenetic modifications in myotonic dystrophy type 1 (DM1) differentiated cells

Tayma Handal[1,2], Manar Abu Diab[1,2], Shira Yanovsky-Dagan[1], Fouad Zahdeh[1], Shulamit Sebban[2], Yosef Buganim[2], Walther J.A.A. van den Broek[3], Derick G. Wansink[3], Silvina Epsztejn-Litman[1], **Rachel Eiges**[1,2]

[1]*Stem Cell Research Laboratory, Medical Genetics Institute, Shaare Zedek Medical Center, Israel*; [2]*School of Medicine, The Hebrew University of Jerusalem, Israel*; [3]*Department of Cell Biology, Radboud Institute for Molecular Life Sciences (RIMLS), Radboud University Medical Center, Netherlands*

Epigenetic mistakes are occasionally secondary to heritable/de novo mutations and can cause a variety of human pathologies. To date, it remains unclear whether gene correction can restore epigenetic defects in patient cells. The aim of this study was to examine whether excision of a large CTG expansion from the DMPK gene, which results in congenital and most severe form of myotonic dystrophy type 1 (DM1), would abolish abnormal epigenetic modifications, and if so, whether it depends on the differentiation state of the cell.

Using bisulfite DNA sequencing and ChIP experiments, we show that the excision of CTGs in undifferentiated mutant hESCs (2000CTG) epigenetically resets the locus by abolishing DNA methylation and repressive histone modifications. This contrasts sharply with repeat deletion in affected myoblasts (2600CTG), where the levels of aberrant epigenetic modifications remain unchanged. In addition, we provide evidence for a switch from reversible to an irreversible heterochromatin state by in vivo differentiation of hESCs (teratomas), which can be set back in induced pluripotent stem cells (iPSCs) by reprogramming of DM1-affected gene-edited myoblasts.

Altogether, our findings suggest that the repair of the DNA sequence in cells of patients by gene editing may not be sufficient to therapeutically address the epigenetic aspects of DM1. The findings of our study may have wider implications as they may apply to a long list of disease-causing mutations that coincide with aberrant epigenetic modifications.

224

CAPRIN1 links embryonic stem cell differentiation with RNA metabolism

Juliane Viegas[1], Gajendra Kumar Azad[1,3], Yuan Lv[2], Sundararaghavan Pattabiraman[4], Jung Eun Park[5], Daniel Kaganovich[4], Siu Kwan Sze[5], Michal Rabani[1], Miguel Esteban[2], Tal Paltiel[1], Eran Meshorer[1]

[1]*Genetics, Hebrew University of Jerusalem, Israel*; [2]*Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, China*; [3]*Department of Zoology, Patna University, India*; [4]*Department of Experimental Neurodegeneration, University Medical Center Goettingen, Germany*; [5]*School of Biological Sciences, Nanyang Technological University, Singapore*

Embryonic stem cells (ESCs) are self-renewing and pluripotent. In recent years, factors controlling pluripotency, mostly nuclear, have been identified. To reveal non-nuclear regulators of ESCs, we screened an endogenously-labelled fluorescent fusion-protein library in mouse ESCs. One of the more compelling hits was the cell cycle-associated protein, CAPRIN1. CAPRIN1, a Stress Granule (SG) component, exhibited a strikingly cyclical localization pattern in sync with mitosis, and localized to SGs, in response to stress. CAPRIN1 knockout had little effect in ESCs, but dramatically skewed differentiation and gene expression programs. RIP-seq and SLAM-seq revealed that CAPRIN1 associates with, and promotes the degradation of thousands of RNA transcripts. CAPRIN1 interactome identified XRN2 as the likely ribonuclease. Upon early differentiation or stress, XRN2 colocalizes with CAPRIN1 inside SGs in a CAPRIN1-dependent manner. We propose that CAPRIN1 regulates an RNA degradation pathway operating during early ESC differentiation, eliminating undesired spuriously transcribed transcripts in ESCs.

226

Understanding the role of lncRNAs as regulators of their neighboring protein-coding genes in the gonadotroph cells of the anterior pituitary

Hadas Gruber, Tal Refael, Lilach Pnuel, Philippa Melamed

Biology, Technion - Israel Institution of Technology, Haifa, Israel

Long non-coding RNAs (lncRNAs) participate in various regulatory processes in the cell, both in the cytosol and the nucleus. In the nucleus, they can mediate gene activation or repression, both in cis and in trans, by the recruitment of regulatory proteins such as transcription factors and chromatin remodeling complexes. In the pituitary gonadotropes (murine) which regulate mammalian reproduction, we have identified several lncRNAs that show cell-type specific enriched expression and are transcribed adjacent to protein-coding genes key to reproduction and fertility. We hypothesize that these lncRNAs can play a role in regulating their neighboring protein-coding genes (PCGs) through localized mechanisms, which depend on sequence, structure and/or location of the lncRNA relative to that of the target gene. The lncRNA, Ak138425, upregulates Lhb (encoding the luteinizing hormone subunit) but only when over-expressed in cis, and this effect is dependent on splicing of the lncRNA. Foxl2os and Nr5a1os are lncRNAs transcribed overlapping and diverging from genes encoding the transcription factors, Foxl2 and Nr5a1 (SF-1) that determine gonadotroph differentiation and identity. Furthermore, genes with a similar structure and orientation as Ak138425 and Foxl2os are found in the human genome (ENSG00000268655 and FOXL2NB, respectively), as well as in other species, suggesting they might be conserved. Our three candidate lncRNA-encoding regions are also predicted (by QmRLFS-finder) to form DNA-RNA hybrid R-loops and G quadruplexes (by pqsfinder), structures that are known to bind PRC2 which represses gene expression. These features suggest several different mechanisms through which the lncRNAs might regulate transcription of their neighboring PCGs.

227

Mosaicism by somatic L1 retrotransposition in normal human cells

Young Seok Ju

Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, South Korea

In the course of an individual's lifetime, genomic mutations are accumulated in somatic lineages. Here, we show the universal ongoing retrotransposition of long interspersed nuclear element-1 (L1), a widespread mobile element in the human genome, through analyzing whole-genome sequences of 880 single-cell clones established from various tissues from 28 individuals and 19 matched colorectal cancers. Remarkably, 88% of colorectal epithelial cells in adult tissues acquire somatic L1 retrotranspositions (soL1Rs). The soL1R burden is ~3 events per cell on average with a substantial variance, which is ~10-fold accelerated in tumorigenesis. Genomic breakpoints suggest a few variations of the molecular mechanisms in the resolution of L1-induced DNA damages. Fingerprinting of donor L1s revealed 34 hot-L1s in the genome, 44% of which are new, often private in an individual, implying many more undiscovered active L1 sources in populations. Integration analysis with the early embryonic phylogenies of the clones, genome-wide methylation and gene expression profiles clearly demonstrate that (1) soL1Rs occur from early embryogenesis at a substantial rate, (2) hot-L1s are transcriptionally activated by early global epigenomic changes, rather than sporadic loss-of-methylation at the late stage, and (3) the post-transcriptional retrotransposition processes are largely inefficient. In sum, this study provides insights into the activities and underlying mechanisms of somatic L1 transcription and retrotransposition events in normal human cells.

228

A LHX2-OTX2 feed-forward regulatory network controls RPE differentiation in mammals

Mazal Cohen-Gulkar, Ahuvit David, Naama Messika-Gold, Mei Eshel, Shai Ovadia, Maria Idelson, Yamit Cohen Tayar, Benjamin Reubinoff, Tamar Ziv, Ran Elkon, Ruth Ashery-Padan

Department of Human Molecular Genetics and Biochemistry, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

A major challenge in understanding tissue differentiation is to uncover the tissue specific transcriptional complexes and to understand how they operate on genomic scale to establish the cell type specific transcription profile. We approach this topic by investigating two evolutionary conserved transcription factors LHX2 and OTX2 in the development of the retinal pigmented epithelium (RPE), a highly specialized central nervous system lineage which is essential for the development and function of the retina. By combining functional studies in vivo in mice and genomic and proteomic analyses in human RPE generated from stem cells we reveal, for the first time, that LHX2 and OTX2 function together, in a feed forward regulatory network on multiple co-bound distant RPE enhancers, regulating both TFs and RPE functional genes, many of which known to be involved in human retinal diseases. The proteomic analyses present evidence that the association between LHX2 and OTX is mediated by the co-factor LDB1. Moreover, the detection of direct physical interaction of OTX2 with multiple subunits of the Swi/Snf (BAF) chromatin remodeling complexes infers key function of OTX2-BAF in modulating the chromatin landscape in a tissue specific pattern. This work identifies a novel LHX2-OTX2 transcriptional complex of the RPE, provide global view on the bound enhancers and the multilevel regulatory network that is essential for acquiring the tissue specific transcription during tissue differentiation.

229

Y chromosomal STRs demonstrate high resolution of male lineages in North Eurasian people on population, tribal, and family levels

Vadim Stepanov[1], Vladimir Kharkov[1], Nikita Kolesnikov[1], Maksim Skalin[2], Vladimir Orekhov[2], Nick Yankovsky[3]

[1]*Laboratory of Evolutionary Genetics, Institute for Medical Genetics, Tomsk National Medical Research Center, Tomsk, Russia, Russia*; [2]*R&D, GORDIZ Ltd., Russia*; [3]*Laboratory of Genome Analysis, Institute of General Genetics, Russian Academy of Sciences, Russia*

The set of 38 YSTR loci combining fast and slow mutated markers was developed to characterize the genetic diversity of male lineages in human populations. The genotyping was performed via fragment length analysis of fluorescently labeled PCR products using commercially available genetic analyzers. Genotyping of 4500 male individuals representing 86 populations of Eastern Europe, Caucasus, Central Asia, Siberia and North-East Asia demonstrates high genetic diversity of Y-chromosomal haplotypes and specific structure of YSTR lineages in populations and sub-populations. In indigenous populations with pronounced tribal structure, the specific tribal haplotype clusters were found. In some native populations of Siberia, Central Asia and North-East Asia, the accumulation of specific variants in family clans within tribe of population subgroup, distinguishing them from other subgroups, has been shown.

230

Expression of Sphingosine-1-Phosphate (S1P)-associated genes at different anatomic sites of Ovarian Carcinoma, and its possible roles in the tumor progression

Aharon Shmuel Baskin[1,2], Ben Davidson[3], Tali Tavor Re'em[1], Reuven Reich[2]

[1]*Department of of Pharmaceutical Engineering, Azrieli College of Engineering, Jerusalem, Israel*; [2]*Institute of Drug Research, School of Pharmacy, The Hebrew University, Israel*; [3]*The Medical Faculty, University of Oslo, Norway*

Ovarian carcinoma (OC) presents a significant challenge both in designing effective drug treatment and in understanding cellular processes during disease progression, especially during the transition from a solid form of the tumor to a detached cellular spheroid form, in effusion. Emerging evidence shows the various functions of sphingolipids in cellular trafficking and cell motility. Specifically, Sphingolipid-1-phosphate (S1P) has been implicated as a potent regulator of cancer progression. Herein, our main objective was to perform mapping of the S1P-related genes, including five Spingosine 1 Receptors (1-5), Spingosine Kinase 1 (SK1), Spingosine Kinase 2 (SK2), Sphingosine 1-Phosphate Lyase , (SGPL1) and Sphingosine-1- Phosphate Phosphatas (SPP1).

Our preliminary results indicate that S1PR4 and S1PR5 were not detected in any of our specimens from patients with ovarian carcinoma. The mRNA expression levels of all other S1P-related genes were significantly elevated in the metastasis samples from OC patients, whereas most of those genes were down-regulated in effusion OC samples. Most interestingly, high mRNA expression levels of S1PR2 in effusion samples were found to be correlated with poor overall survival, evaluated from Kaplan-Meier test. Moreover, spearman correlations between S1P-related gene expression levels and downstream target proteins levels, as p-Akt, p-85, and p-ERK and JNK revealed possible activation of S1P signaling pathways in OC.

233

Uncovered novel recurrent deletions in AML are associated with replication stress

Shay Shilo[1], Aditee Kadam[1], Noa Chapal-Ilani[1], Ester Bnaya[1], Mark Minden[2], Liran Shlush[1]

[1]*Molecular Cell Biology, The Weizmann Institute of Science, Israel*; [2]*Medical Oncology & Hematology, Princess Margaret Cancer Centre, University Health Network, Canada*

Acute myeloid leukemia (AML) is an aggressive blood malignancy (80% mortality) that is characterized by a small number of mutations (n=10-20) (Ley et al., 2013). A single genetic mechanism, the microhomology-mediated end joining (MMEJ), is the predominant cause of most of the deletions in AML and other myeloid malignancies (Feldman et al., 2021). However, mainstream variant callers often overlook medium-sized deletions (50-100 bp), a frequent result of MMEJ repair (Cameron et al., 2019). Therefore many deletions might be ignored by the current analysis pipelines. We developed an algorithm that integrates prior knowledge about the genetic mechanism and can identify deletions even without alignment to a reference genome, to reveal cryptic deletions. The algorithm was tested over the Beat AML dataset (359 tumor-control paired exomes). All the previously reported deletions were detected, and the frequency of germline variant deletions in Beat AML as detected by the tool was correlated with the known population frequency. Identification of novel somatic deletions included validation with targeted sequencing on a different cohort of 500 samples from AML patients. Eventually, the new approach found 13 recurrent somatic mutations that were not reported previously. Moreover, the analysis detected a previously unreported class of MMEJ deletions with permissive homology with an extended base pairing. Characterization of the somatic and germline deletions demonstrated a strong association with genomic features influenced by replication stress, such as G-quadruplexes and minisatellites. No epigenetic change was observed in the DNA methylation before and after the MMEJ repair.

234

Non-complementary mismatched base pairs locally distort DNA structure, leading to increased DNA-binding by transcription factor proteins

Ariel Afek

Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel

Mismatched base-pairs represent a form of DNA damage frequently formed in living cells. Mismatches alter the local DNA structure, affecting interactions with DNA-binding proteins, including regulatory transcription factors (TFs).

Recent studies found that interactions between TFs and damaged DNA may play an essential role in mutagenesis. However, DNA damages significantly distort the DNA, and the structural impact of damage-induced or other distorted DNA shapes on protein-DNA recognition has not been well characterized.

We present Saturation Mismatch Binding Assay (SaMBA), a new technique to characterize the effects of mismatches on TF-DNA binding in high throughput. SaMBA generates DNA duplexes containing all possible single-base mismatches and quantitatively assesses the impact of the mismatches on TF-DNA interactions.

We applied SaMBA to measure the binding of 21 TFs to thousands of mismatched sequences and mapped the impact of mismatches on these TFs. Remarkably, for all TFs examined, the introduction of mismatches at certain positions resulted in significantly increased binding, with some mismatches creating high-affinity binding sites in nonspecific DNA and some converting known binding sites into “super-sites” stronger than any canonical Watson-Crick site.

Structural analyses revealed that these mismatches are frequently distorting the naked DNA such that its structure becomes similar to that of bound DNA sites, thus explaining the increased binding measured in our assay. Furthermore, our results reveal that the energy cost of deforming the DNA structure is a major determinant of protein-DNA recognition and reveal mechanisms by which mismatches can recruit TFs and thus modulate replication and repair activities in the cell.

235

Loss of function of FIGNL1, a DNA damage response gene, is a novel cause of human ovarian dysgenesis

Natan Florsheim[1,2], Larisa Naugolni[4], Paul Renbaum[1], Orit Lobel[1], Merav Gold[3], Michal Goldberg[3], Ephrat Levy-Lahad[1,2], David Zangen[2,5]

[1]*Medical Genetics Institute, Shaare Zedeq Medical Center, Israel*; [2]*Faculty of Medicine, The Hebrew University, Israel*; [3]*The Alexander Silberman Institute of Life Science, The Hebrew University, Israel*; [4]*Pediatric Endocrinology and Diabetes Institute, Shamir Medical Center, Israel*; [5]*Division of Pediatric Endocrinology, Hadassah Medical Center, Israel*

Background: Severe Ovarian Dysgenesis (OD), a rare heterogeneous XX disorder of sex development presents clinically with primary amenorrhea, hypergonadotrophic hypogonadism and infertility. The genetic basis of OD remains unknown in 70% of cases. To identify novel causes of OD, we study patients in which known genes have been excluded.

Methods: Whole-exome-sequencing was performed in a 14.5y old Ashkenazi Jewish, non-consanguineous origin patient, with OD. DNA damage response (DDR) was tested using Mitomycin-C (MMC) in chromosomal breakage assay, and analysis of ɣH2AX and DYK-Fignl1 foci in response to MMC and Phleomycin.

Results: We identified compound heterozygous frameshift deletions in FIGNL1, a DDR pathway gene: c.189delT and c.1519delTCTCA. Segregation was consistent with recessive inheritance. Western blots of DYK-tagged-FIGNL1-constructs showed no protein is produced by c.189delT, and 64% of WT (P=0.003) is produced by c.1519delTCTCA. Chromosomal breakage assays revealed significantly higher number of DNA breaks in patient-derived fibroblasts compared to control. DNA breaks increasement occurred both spontaneously (22.82±1.66 vs. 10.61±0.98 P=6X10-9) and following DNA damage induction (MMC, 5.00±0.59 vs. 2.50±0.25 P=0.003). HeLa cells transfected with DYK-tagged-FIGNL1-mutant or wild-type constructs showed overall decrease in nuclear foci (c.1519delTCTCA 5.6±1.5 vs. 14.8±0.9 P=0.002) and in recruitment of FIGNL1 to the nucleus in response to phleomycin (0±1.3 vs. 5.8±0.8 P=7.1X10-11) in the mutant transfected cells.

Conclusions: The novel compound heterozygous mutations in FIGNL1 are accompanied by impaired DDR and suggest that FIGNL1 loss of function is a novel genetic etiology for ovarian dysgenesis in humans. This further expands the crucial role of DDR pathway in normal ovariogenesis.

236

The transcriptional and regulatory identity of erythropoietin producing cells

Bjort Kragesteen[1], Amir Giladi[1], Eyal David[1], Laufey Geirsdottir[1], Chamutal Gur[1], Fernando Racimo[2], Eske Willerslev[2], Roland Wenger[3], Ido Amit[1]

[1]*Immunology, Weizmann Institute of Science, Israel*; [2]*GLOBE Institute, University of Copenhagen, Denmark*; [3]*Institute of Physiology and National Centre of Competence in Research, University of Zurich, Switzerland*

Erythropoietin (Epo) is the master regulator of erythropoiesis and oxygen homeostasis. Despite its physiological importance, the molecular and genomic contexts of the cells responsible for renal Epo production have not yet been resolved, limiting effective cell-based therapies for anaemia. Here, we performed single-cell profiling of an Epo reporter mouse to molecularly identify Epo-producing cells under hypoxic conditions. We report that a distinct and homogeneous population of kidney stromal cells, which we name Norn cells, are the sole source of Epo production in vivo. Extensive characterization of the Norn epigenetic and transcriptional landscape revealed Norn-specific markers, pathways, and transcription factor circuits conserved from mice to humans. These findings open new avenues to functionally dissect EPO gene regulation in human evolution and disease, and pave the way for the next generation of genetic and cell-based approaches for EPO therapies

237

Unraveling the dynamics of nuclear speckles and gene expression during Herpes virus infection

Shani Nadav Eliyahu[1], Sami Salminen[2], Vesa Aho[2], Salla Mattola[2], Maija Vihinen-Ranta[2], Yaron Shav-Tal[1]

[1]*The Mina & Everard Goodman Faculty of Life Sciences & Institute of Nanotechnology, Bar-Ilan University, Israel*; [2]*Department of Biological and Environmental Science & Nanoscience Center, University of Jyvaskyla, Finland*

Nuclear architecture undergoes dramatic changes during lytic viral infection, including changes in the distribution of chromatin and nuclear bodies, such as nuclear speckles that contain many types of splicing factors. DNA viruses induce development of nuclear Viral Replication Compartments (VRCs) which is accompanied by marginalization of host chromatin. Herpes Simplex Virus Type 1 (HSV-1) infection leads to the redistribution of nuclear speckles from the nuclear center to the nuclear periphery, and adjacent to VRCs. In this study we aimed to explore nuclear speckle function and dynamics with respect to viral and cellular RNAs during lytic infection of HSV-1. We observed that different splicing factors, which are nuclear speckle components, were redistributed under viral infection. Interestingly, transcription sites of viral RNA co-localized with nuclear speckles. Moreover, upregulated host gene transcription sites and single transcripts of this gene were associated with nuclear speckles. Taken together, the findings suggest that nuclear speckles play an important role in the regulation of cellular and viral gene expression during infection and might aid in splicing and export of viral and host mRNAs.

238

Understanding the complex genetic architecture connecting rheumatoid arthritis, osteoporosis, and inflammation: Discovering causal pathways

Melody Kasher[1], Frances Williams[3], Max Freidin[3], Ida Malkin[1], Stacey Cherny[1], Gregory Livshits[1,2,3]

[1]*Anatomy and Anthropology, Tel Aviv University, Israel*; [2]*School of Medicine, Ariel University, Israel*; [3]*Twin Research and Genetic Epidemiology, King's College London, UK*

Introduction: Rheumatoid arthritis (RA) and osteoporosis (OP) are two comorbid complex inflammatory conditions with evidence of shared genetic background and causal relationships. We aimed to clarify the genetic architecture underlying RA and various OP phenotypes whilst additionally considering an inflammatory component, C-reactive protein (CRP).

Material and Methods: Genome-wide association study summary (GWAS) statistics were acquired from the GEFOS consortium, CHARGE consortium, and UK Biobank. Mendelian randomization (MR) was used to detect the presence of causal relationships. Colocalization analysis was performed to determine shared genetic variants between CRP and OP phenotypes. Analysis of pleiotropy between traits due to shared causal SNPs was performed using pleiotropic analysis under composite null hypothesis (PLACO).

Results: MR analysis was suggestive of horizontal pleiotropy between RA and OP traits. RA was a significant causal risk factor for CRP (β= 0.027, 95%CI= 0.016 to 0.038). There was no evidence of CRPàOP causal relationship, but horizontal pleiotropy was apparent. Colocalization established shared genomic regions between CRP and OP including GCKR and SERPINA1 genes. Pleiotropy arising from shared causal SNPs revealed through the colocalization analysis were all confirmed by PLACO. These genes were found to be involved in the same molecular function “protein binding” (GO:0005515) associated with RA, OP and CRP.

Discussion: We identified three major components explaining the epidemiological relationship among RA, OP and inflammation: 1) Pleiotropy explains a portion of the shared genetic relationship between RA and OP, albeit polygenically; 2) RA contributes to CRP elevation; 3) CRP, which is influenced by RA, demonstrated pleiotropy with OP.

239

The archaic mutational load predicts the fate of introgressed fragments in humans

Alessio Gerussi[1,2,3], Rosanna Asselta[4], Viviane Slon[1,2], Pietro Invernizzi[3], Fabrizio Mafessoni[5]

[1]*Department of Anatomy and Anthropology and Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Israel*; [2]*The Dan David Center for Human Evolution and Biohistory Research, Tel Aviv University, Israel*; [3]*Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Italy*; [4]*Department of Biomedical Sciences, Humanitas University, Italy*; [5]*Department of Plant and Environmental Sciences, Weizmann Institute of Science, Israel*

Background and Aims: About 1-4% of the genome of humans living out of Africa entered in the human gene pool through interbreeding with Neanderthals. It has been suggested that Neanderthals accumulated deleterious variants that were swiftly selected against when entered the human gene pool, explaining the depletion of fragments originated from Neanderthals in genes of present-day individuals. Here we leverage the diversity of archaic genomes and deleteriousness measures of mutations to characterize the archaic mutational load along the genome. Methods: We combined publicly available genomic datasets of present-day humans and archaic hominins and measures of phylogenetic conservations to develop population-genetics aware statistics. These were used to test hypotheses using generalized linear mixed models and resampling-based methods. Results: We show that regions with more putatively deleterious mutations in archaic populations than in humans were more efficiently removed after introgression than regions with a lower mutational load. We found a similar pattern for variants influencing gene expression and immune-related variants, despite these are overrepresented in fragments of Neanderthal origin. Conclusion: Fragments carrying an excess of Neanderthal-derived mutations were largely purged by natural selection.

240

Ethical issues of forensic genealogical searching based on genomic information in Russia

Svetlana Borinskaya[1], Pavel Tishchenko[2]

[1]*Laboratory of Genome Analysis, Institute of General Genetics, Russian Academy of Sciences, Russia*;
[2]*Humanitarian Expertise and Bioethics Department, Institute of Philosophy, Russian Academy of Sciences, Russia*

Genealogical search based on autosomal SNP and/or Y STR loci alleles is being used to identify the criminal or the victim as a complementary approach based on autosomal CODIS STR loci alleles. Both of the approaches are being effectively used in many countries Russia included. The most known cases in Russia are the Russian Tsar family remains investigation in 2009 and the terrorist identification after the bomb explosion in Moscow Domodedovo airport in Moscow in 2011. The terrorist identification was based on Y STR allele database subpopulation analysis. Only population Y STR databases are strong enough to be used effectively in criminal investigations in Russia that implies academic scientist involvement in the investigations. Bioethical principle demands a reasonable balance of proportionality between the interest of the society and an individual. We designed a questionnaire and arranged an inquiry of 250 scientists during the Russian Genetic Society Meeting. The position of the 48.6% of the respondents is that there is no need to request anyone consent for the usage of the genomic database data in case of violent or terrorist crimes. The position of the 39.6 % of the respondents is that for the cases there is a need for the informed consent document from the donors of the data. Ethical issues and influences of lay public and expert views on DNA testing in the criminal field will be discussed.

241

Genetics, pathomics and radiomics analysis as a tool to study the molecular mechanisms of MET induced tumorigenesis and generate breast cancer personalized medicine

Michal Bloom[1], Dvora Kidron[3], Galia Tsarfaty[2], Eran Dotan[1], Ayman Iraqi[1], Judith Horev[1], Carrie R Graveel[4], Fuad A Iraqi[1] and Ilan Tsarfaty[1]

[1]Department of Clinical Microbiology and Immunology, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; [2]Department of Pathology, Meir Medical Center, Kfar Saba, Israel; Department of Diagnostic Imaging, MRI Unit, The Chaim Sheba Medical Center, Tel Hashomer, Israel; [4]Center for Cancer and Cell Biology, Van Andel Research Institute, Grand Rapids, MI, USA

Breast cancer is the second leading cause of cancer deaths among women. Breast tumorigenesis and progression arise through a sequential accumulation of mutations in oncogenes and tumor suppressor genes that serve as tumor driver genes. MET tyrosine kinase is the hepatocyte growth factor/scatter factor (HGF/SF) receptor. MET constitutive activation and mutations promote tumor growth and metastases. We hypothesize that MET activation, alteration of other driver genes, and Inherited Driver Modifier Gene products (IDMGs) generate Dynamic Driver Modifiers Protein-Protein Carcinogenic Networks (DMCNs) that Induce tumorigenesis and metastasis.

To isolate MET-IDMGs and DMCNs, we created a novel mice model in which 20 CC lines (Collaborative-Cross CC - mice with diverse genetic backgrounds) overexpress the activated MET receptor. We followed tumor development using CT imaging and characterized the pathology H&E staining analysis. Tumor follow-up revealed that the mutated MET CC lines showed wider diversity regarding tumor types (carcinoma, lymphoma, and sarcoma).

QTL and computational bioinformatics analyses (SNP's, ANAT and Human cBioPortal Database), together with qPCR and immunofluorescence staining, enabled the selection of 9 candidates IDMGs that demonstrate alteration by modulation of MET signaling. Several IDMGs served as very significant prognostic factors in human breast cohorts. Another set of candidate IDMGs code for Pathomics and Radiomics features. CRISPR-CAS9 KD and KO of several IDMGs significantly altered MET and p53 expression and reduced MET-induced motility.

We also developed a precision medicine tool using graphic AI and MET and IDMGs expression these results demonstrate that modifier genes play a significant role in tumor type development and tumor progression. The multidisciplinary analysis enables elucidating the molecular mechanisms of MET-induced tumorigenicity to generate novel prognostic factors and precision anti-MET therapy.

242

Candidate gene prioritization and disease-gene discovery through phylogenetic profiling

Christina Canavati[1], Dana Sherill-Rofe[1], Idit Bloch[1], Moien Kanaan[2], Ephrat Levy-Lahad[3], Yuval Tabach[1]

[1]Department of Developmental Biology and Cancer Research, The Hebrew University of Jerusalem, Jerusalem, Israel; [2]Molecular Genetics Laboratory, Istishari Arab Hospital, Ramallah, Palestine; [3]Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel

Recent studies estimate that the causative genes for hundreds of Mendelian diseases are still unknown. Accelerating disease gene discovery thus requires integrating information that reaches beyond the current knowledge base. Phylogenetic profiling (PP) offers such a new perspective. This led us to develop EvoRanker, which uses clade-wise PP across 1028 genomes to identify candidate disease-causing genes based on their global or local co-evolution with genes previously associated with the disease. We aim to prioritize patient candidate genes specifically in “unsolved” NGS case studies that may harbor genes that are either unannotated or have no clear link to the patient’s phenotype. Benchmarking of previously solved exome data revealed that the “true” gene was ranked among the top 5 candidates in ~78% of the cases based on PP alone. This approach yielded comparable results to other available tools. Yet, the results show complementarity amongst the top-ranked genes, indicating that PP analysis could pinpoint “true” causative genes that could not be identified by other existing tools. Remarkably, analysis of “unsolved” exome cases revealed two potential novel genes to be associated with previously undescribed genetic syndromes. Our platform, which scans global and local coevolution across hundreds of genomes, presents a complementary and innovative approach to pinpoint patient candidate genes that merit further investigation.

TALKS

112

Machine-learning of complex evolutionary signals improves classification of SNVs

Yuval Tabach[1], Sapir Labes[1], Doron Stupp[1], Naama Wagner[2], Idit Bloch[4], Michal Lotem[1], Ephrat Lahad[3], Paz Polak[3], Tal Pupko[2]

[1]*Department of Developmental Biology and Cancer Research, Faculty of Medicine, and Hadassah University Medical School, The Hebrew University, Israel*; [2]*The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Israel*; [3]*Medical Genetics Institute, Shaare Zedek Medical Center, Israel*; [4]*Department of Developmental Biology and Cancer Research, Hebrew University, Israel*

Conservation is a strong predictor for the pathogenicity of single nucleotide variants (SNVs). However, some positions that present complex conservation patterns across vertebrates stray from this paradigm. Here we analyzed the association between complex conservation patterns and the pathogenicity of SNVs in 115 disease-genes (i.e. all genes with 50 reported pathogenic or benign SNVs). We show that conservation is not a one-rule-fits-all solution since its accuracy highly depends on the analyzed set of species and genes. This context dependent concept is results from the evolutionary crosstalk between nucleotides, genes, and species. Those While conservation in certain species is highly informative to predict pathogenicity in sub-group of genes, other species are more informative to different sub-groups. This insight led to developing EvoDiagnostics which uses the conservations against each species as a feature within a random-forest machine-learning classification algorithm. EvoDiagnostics outperformed traditional conservation algorithms as new deep-learning approaches in most prediction-task. Overall, we suggest a new way to look at conservation which is more biological relevant and improves clinical prediction.

124

Human fetal kidney organoids enriched for notch dependent early epithelial differentiation

Michael Namestnikov[1,3], Osnat Cohen-Zontag[1], Dorit Omer[1], Yehudit Gnatek[1], David Schneider[4], Herzel Ben-Hur[4], Benjamin Dekel[1,2,3]

[1]*The Pediatric Stem Cell Research Institute and Pediatric Nephrology Division, Edmond and Lily Safra Childrens Hospital, Israel*; [2]*Pediatric Nephrology Division, Edmond and Lily Safra Childrens Hospital, Israel*; [3]*Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Israel*; [4]*Department of Obstetrics and Gynecology, Assaf Harofeh Medical Center, Israel*

The gold standard source for kidney development research is the fetal kidney. Transgenic mice have been utilized as a robust tool for modelling congenital abnormalities of the kidney and urinary tract (CAKUT), however, they do not precisely mimic human fetal kidney (hFK) development. In recent years, pluripotent stem cell derived-kidney organoids have shown great potential which is hindered by unspecific differentiation and incomplete maturation. In contrast, three-dimensional culture methods of human fetal kidney have been hardly reported. Utilizing a chemically defined serum free media we generated hFK organoids (hFKOs) expressing key markers of early nephron epithelium, differentiating renal epithelium, mature tubular markers and self-organize into polarized renal epithelium similarly to the native tissue. In addition, bulk RNA sequencing of hFKOs disclosed cellular lineages of early nephron epithelium at the renal vesical, comma and s-shape stage (e.g., "vesiculoids"), at levels that far exceed those in iPSC-derived kidney organoids and adult kidney tubuloids. Furthermore, Single-cell RNA sequencing and pseudotime analysis present diverse populations in hFKOs with a preserved differentiation axis, spanning nephron progenitors, glomeruli, developing and mature nephron segments. Importantly, vesiculoids were highly enriched for markers of the NOTCH signaling pathway and inhibition of NOTCH dampened vesiculoid growth and downregulated expression of LHX1, PAX2 and CDH6, suggestive of depletion of early proximal tubule progenitors. Vesiculoids are an in-vitro tool for the propagation of human fetal kidney tissue which recapitulate early differentiating renal structures superior to iPSC-derived kidney organoids. This system can be utilized for disease modelling of CAKUT inducing mutations and to study the effect of maternally administered drugs on the developing fetal kidney.

128

Comprehensive transcriptomic analysis of A-T skin fibroblasts reveals unique gene expression and pathway signature of cellular senescence

Amit Levon[1], Majd Haj[1], Yann Frey[1], Judith Campisi[2], Ran Elkon[1], Yael Ziv[1], Yosef Shiloh[1]

[1]*Department of Human Molecular Genetics and Biochemistry, Tel Aviv University, Tel Aviv, Israel*;
[2]*Campisi lab, Buck Institute for Research on Aging, USA*

Genome instability enhances the pace of aging. Therefore, genome instability syndromes are usually associated with accelerated segmental aging. Genome integrity in the face of ongoing DNA damage is maintained primarily by the DNA damage response (DDR). A chief mobilizer of DDR is the ATM protein kinase. ATM loss in humans leads to a multisystem genome instability syndrome, ataxia-telangiectasia (A-T). Part of the complex A-T phenotype is cellular senescence, a cell fate that includes cell cycle arrest, and plays a role in development and tissue homeostasis, as well as age-related degenerative and malignant diseases. Early studies demonstrated a rapid decline in the proliferation of cultured primary skin fibroblasts from A-T patients, which we confirm do indeed undergo premature senescence. Comparative RNA-sequencing analysis of A-T and control fibroblasts revealed unique transcriptional alterations in A-T cells and their dynamics over time in culture. Cluster analysis of differentially expressed genes (DEGs) allowed for the detection and categorization of these transcriptomic patterns. Functional enrichment analyses provided insight into the molecular pathways and cellular components associated with the premature senescence of A-T cells. The results were further corroborated by analyzing alterations in gene expression profiles using 'gene set enrichment analysis' (GSEA). Altogether, the analyses point to engagement of interferon signaling and extracellular matrix remodeling pathways, presumably caused by an accumulation of DNA damage. Notably, these pathways are known to be involved in cellular senescence and age-related pathologies, including fibrosis and cancer. Our data provide a molecular dimension to the segmental premature aging observed in A-T patients.

132

10K Study: Personalized medicine based on multi-omics population-level cohorts

Adina Weinberger, Eran Segal

Computer Science and Applied Mathematics, The Weizmann Institute of Science, Israel

Extraordinary technological advances in genomics in the past two decades led to efficient sequencing of many human genomes and to genome-scale measurements of gene expression, even at the single-cell level. These capabilities allowed the creation of unprecedented catalogues of novel genomes, functional DNA elements and non-coding RNAs from all kingdoms of life. But, perhaps with the exception of cancer and gene therapy for some monogenic diseases, genomics has yet to deliver on its promise to impact our everyday life. For example, drugs are still being developed in the traditional way, with screening assays to find lead compounds for targets that typically arise from animal studies, often without involving genomics in any of the steps. The10K study is a long-term observational study designed to gather information on lifestyle and diseases from a diverse sample of 10,000 Israeli adults (age 40-70). The project combines innovative medical tests and advanced artificial intelligence methods to discover personal characteristics that can help predict future medical conditions, even before they break out. It profiles participants at unprecedented depth, including physiological assessment of their muscular, skeletal, liver, blood, heart, vascular, immune, gastrointestinal, and cognitive body systems; coupling these medical records with molecular methods that profile the genetics, microbiome, metabolomic, transcriptomic, proteomic, and immune systems of each participant. We aim to apply machine learning methods to both the baseline variation in disease risk and the longitudinal data in order to identify novel therapeutic targets as well as means to modulate them by dietary, lifestyle, microbiome, and small molecules. We will also devise algorithms for predicting future onset of various diseases.

133

ENLIGHT: Pan-cancer response prediction to targeted and immunotherapies via tumor transcriptomics

Gal Dinstag[1], Eldad Shulman[1], Efrat Elis[1], Doreen Ben-Zvi[1], Omer Tirosh[1], Eden Maimon[1], Isaac Meilijson[1], Emmanuel Elalouf[1], Eyal Schiff[1], Danh-Tai Hoang[2], Sanju Sinha[3], Nishanth Ulhas Nair[3], Joo Sang Lee[4], Alejandro Schaffer[3], Ze'ev Ronai[5], Dejan Juric[6], Andrea Apolo[3], William Dahut[3], Stanley Lipkowitz[3], Raanan Berger[7], Razelle Kurzrock[8], Antonios Papanicolau-Sengos[3], Fatima Karzai[3], Mark Gilbert[3], Kenneth Aldape[3], Padma Rajagopal[3], Tuvik Beker[1], Eytan Ruppin[3], Ranit Aharonov[1]

[1]*Pangea Biomed, Pangea Biomed, Israel*; [2]*Biological Data Science Institute, The Australian National University, Australia*; [3]*National Cancer Institute, National Cancer Institute, USA*; [4]*Department of Precision Medicine, Sungkyunkwan University, South Korea*; [5]*Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, USA*; [6]*Massachusetts General Hospital Cancer Center, Massachusetts General Hospital Cancer Center, USA*; [7]*Cancer Center, Chaim Sheba Medical Center, Israel*; [8]*Worldwide Innovative Network (WIN) for Personalized Cancer Therapy, Worldwide Innovative Network (WIN) for Personalized Cancer Therapy, USA*

Precision oncology is gradually advancing into mainstream clinical practice, demonstrating significant survival benefits. However, eligibility and response rates remain limited in many cases, calling for better predictive biomarkers. Here we present ENLIGHT, a transcriptomics-based computational approach that identifies clinically relevant genomic interactions and uses them to predict patients' response to a variety of therapies in multiple cancer types without training on previous response data. We study ENLIGHT in two translationally oriented scenarios, Personalized Medicine (PM), aimed at prioritizing treatments for a single patient, and Clinical Trial Design (CTD), selecting the most likely responders in a patient cohort. Evaluating ENLIGHT's PM performance on 21 blinded clinical trial datasets, we show that it can effectively predict treatment response across multiple therapies and cancer types (obtaining an odds ratio of 2.59), substantially improving upon SELECT, a previously published transcriptomics-based approach, and performing as well as supervised predictors developed for specific indications and drugs, but on a much broader array of therapies and indications. In the CTD scenario, ENLIGHT can markedly enhance the success of clinical trials by excluding non-responders while achieving more than 90% of the response rate attainable under an optimal exclusion strategy. In sum, ENLIGHT is one of the first approaches to demonstrably predict therapeutic response across multiple cancer types by leveraging the transcriptome.

139

Bi-allelic variants in NAE1 cause intellectual disability, ischiopubic hypoplasia, stress-mediated lymphopenia and neurodegeneration

Imre Schene[1], Holger Rehmann[2], Maarten Massink[3], Maria Van Der Wal[4], Corinna Bauder[5,6], Martha Labeur[5], Maarten Lequin[7], Michiel Houben[8], Jaques Giltay[3], Saskia Haitjema[9], Albert Huisman[9], Fleur Vansenne[10], Judith Bluvstein[11], John Pappas[12], Laila Shailee[13], Yuri Zarate[14], Michal Mokry[15], Gijs Van Haaften[16], Edward Nieuwenhuis[17], Damian Refojo[18,19], Femke Van Wijk[4], Sabine Annemijn Fuchs[1], Peter Marin Van Hasselt[1]

[1]*Department of Metabolic Diseases, division Pediatrics,, Wilhelmina Childrens Hospital University Medical Center Utrecht, Netherlands*; [2]*Department of Energy and Biotechnology, Flensburg University of Applied Sciences, Germany*; [3]*Department of Genetics, Wilhelmina Childrens Hospital, University Medical Center Utrecht, Netherlands*; [4]*Center of Translational Immunology, Wilhelmina Childrens Hospital University Medical Center Utrecht, Netherlands*; [5]*Department of Neuroendocrinology, Max Planck Institute of Psychiatry, Germany*; [6]*Institute of Developmental Genetics, Helmholtz Zentrum Munchen, Germany*; [7]*Division Imaging and oncology, University Medical Center Utrecht, Netherlands*; [8]*Department of General Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Netherlands*; [9]*Central Diagnostics Laboratory, University medical center Utrecht, Netherlands*; [10]*Department of Medical Genetics, University Medical Center Groningen, Netherlands*; [11]*Dravet Center and Comprehensive Epilepsy Center, NYU Grossman School of Medicine, USA*; [12]*NYU Clinical Genetic Services, NYU Grossman School of Medicine, USA*; [13]*Department of Radiology, NYU Grossman School of Medicine, USA*; [14]*Section of Genetics and Metabolism, University of Arkansas for Medical Sciences, USA*; [15]*Laboratory of Experimental Cardiology, Department of Cardiology, University Medical Center Utrecht, Netherlands*; [16]*Department of Genetics, Division Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Netherlands*; [17]*Department of Biomedical and life sciences, University College Roosevelt, Netherlands*; [18]*CONICET, Instituto de Investigacion en Biomedicina de Buenos Aires (IBioBA) - CONICET - Partner Institute of the Max Planck Society, Argentina*; [19]*Molecular Neurobiology, Max Planck Institute of Psychiatry, Germany*

Neddylation has been implicated in various cellular pathways and in the pathophysiology of numerous diseases. We identified four individuals with bi-allelic variants in the NAE1 gene, encoding neddylation's E1 enzyme. Individuals showed decreased NAE1 expression, a changed ratio of non-neddylated to neddylated cullins and proteasome dysfunction. Next, we aimed to delinate the cellular consequences of faulty neddylation and how these would lead to the clinical phenotype. We chose to focus primarily on the most rare phenotypic features, that would best reflect the pathophysiology at stake. Rare features included neuronal loss during infections and lymphopenia that worsened to some extent during infections suggesting that NAE1 protects against stress-mediated cellular apoptosis. In support, NAE1 deficient cells showed decreased viability during various stress-inducing events, such as CD3/CD28 stimulation. The rarest phenotypic feature - delayed closure of the ischiopubic rami – correlated with significant downregulation of RUN2X and SOX9 expression in transcriptomic data of fibroblasts. Both genes are involved in the pathophysiology of ischiopubic hypoplasia. Thus, we show NAE1 plays a major role in (skeletal) development and cellular homeostasis during stress. Our approach suggests a focus on rare phenotypic features is able to capture the biological essence in diseases caused by mutations in pleiotropic genes.

148

Sex-specific regulation of metabolic health and vertebrate lifespan by AMP biosynthesis

Itamar Harel

Department of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel

Energy homeostasis is disrupted with age, which then fuels multiple age-related pathologies. The AMP-activated protein kinase (AMPK) is the primary sensor of cellular energy in eukaryotes. However, the genetic regulation of vertebrate aging by AMPK remains poorly understood. Here, we manipulate energy levels in the turquoise killifish by mutating APRT, a key enzyme in AMP biosynthesis. These manipulations produced a male-specific lifespan extension and restored metabolic plasticity. Exploring the observed sex differences using an integrated omics approach implicated the mitochondria as an important player. Mechanistically, APRT regulated mitochondrial functions and AMPK activity, mimicking energy starvation in heterozygous cells. A fasting-like state was also detected, particularly in heterozygous males, which leads to resistance to high-fat diet. Finally, life-long intermittent fasting eliminated the male-specific longevity benefits mediated by the APRT mutation. These observations identify the AMP/AMPK axis as a sex-specific regulator of vertebrate longevity and metabolic health.

156

Hormone-controlled cooperative binding of transcription factors within enhancer clusters drives synergistic gene induction during fasting

Dana Goldberg[1], Nufar Buchshtab[1], Meirav Bar Shimon[1], Meital Charni-Natan[1], Ido Goldstein[1]

Institute of Biochemistry, Food Science and Nutrition. the Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Israel

During fasting hepatocytes produce glucose in response to hormonal signals. Glucagon and glucocorticoids are principal fasting hormones that cooperate in regulating glucose production by gluconeogenesis. However, how these hormone signals are integrated and translated to a biological output is unknown.

We use genome-wide profiling of gene expression, enhancer dynamics and transcription factor (TF) binding in primary hepatocytes to uncover the mode of cooperation between glucagon and glucocorticoids.

We found that compared to a single treatment with each hormone, a dual treatment directs hepatocytes to a pro-gluconeogenic gene program by synergistically inducing gluconeogenic genes. The cooperative model driving synergistic gene expression is based on glucagon-mediated enhancer activation, leading to increased binding of the glucocorticoid receptor (GR) upon glucocorticoid stimulation. Thus, glucagon-mediated enhancer activation assists GR binding to specific enhancers. Glucagon did not only activate single enhancers but was also able to activate enhancer clusters, thereby assisting the loading of GR in other enhancer units within the cluster.

In summary, our data show that cells integrate extracellular signals by an enhancer-specific mechanism whereby one hormone activates enhancers, thereby assisting the loading of a TF stimulated by a second hormone. This leads to synergistic gene induction and a tailored response to fasting.

172

Utilization of mitochondrial augmentation technology to explore associations between mitochondrial haplogroup, function, and persistence

Yehuda Brody[1], Igor Pozdnyakov[1], Natalie Yivgi-Ohana[1], **Noa Sher**[1]

Minovia, Minovia, Israel

Mitochondria are critical for cellular homeostasis and function, as well as cell fate decisions including proliferation, differentiation, and cell death. Each mitochondrion harbors several copies of circular mitochondrial DNA (mtDNA), and the number of mitochondria in different cell types ranges from tens to hundreds of thousands. During evolution, mitochondrial DNA diversified and are now classified to different haplotypes, correlated with functional differences and disease frequency. Mitochondrial dysfunction and levels of heteroplasmy (the subset of cellular mtDNA content harboring mutations) are associated with numerous diseased states, as well as with the natural process of aging.

Mitochondria are critical regulators of hematopoietic function, shifting cellular energetic homeostasis between glycolytic and oxidative phosphorylation states, and regulating stemness of hematopoietic stem cells. Heteroplasmy, copy number and mitochondrial function of hematopoietic cell types is tightly controlled; it has been demonstrated that improved mitochondrial content and function of even a single hematopoietic lineage exerts multisystemic phenotypic improvements.

We have developed an ex vivo method to enrich cells with exogenous mitochondria, and an error-correcting nanopore-based analytical method to accurately quantify exogenous mtDNA in recipient cells. In parallel, we developed a mitochondrial bank encompassing the major haplogroups. By simultaneously enriching HSPCs with mitochondria of multiple haplogroups, we are exploring coexistence and persistence of exogenous mtDNA on the background of a single recipient cell. These methods, in conjunction with single-cell based methods assessing ATP content based on protein synthesis capacity, will enable a better understanding of the interplay between mitochondrial haplogroup, heteroplasmy level, and mitochondrial activity in hematopoietic cells.

187

Predicting molecular mechanisms of hereditary diseases by using their tissue-specific manifestation

Eyal Simonovsky[1], Moran Sharon[1], Maya Ziv[1], Omry Mauer[1], Idan Hekselman[1], Juman Jubran[1], Ekaterina Vinogradov[1], Chanan Argov[1], Omer Basha[1], Lior Kerber[1], Yuval Yogev, Ayellet Segre[4], Hae Kyung Im[3], Ohad Birk, Lior Rokach, **Esti Yeger-Lotem**[1]

[1]*Department of Clinical Biochemistry and Pharmacology, Ben-Gurion University of the Negev, Israel*; [3]*Section of Genetic Medicine, Department of Medicine, The University of Chicago, USA*; [4]*Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA., USA*

How do widely expressed disease genes and variants lead to tissue-specific hereditary diseases? Previous attempts to answer this question were limited to testing few candidate mechanisms. Moreover, although many hereditary diseases manifest clinically in specific tissues, variant interpretation schemes have been mostly oblivious to tissue contexts. To answer this question at larger scale, we developed a machine-learning platform that uses tissue-selectivity as a means to predict disease genes along with their disease-causality features. The platform was based on thousands of biologically interpretable features derived from heterogeneous tissue omics dataset, hence dramatically expanding previous efforts, and outperformed other tools. Using this platform we uncovered known and novel disease-causality features, the most common of which was previously overlooked. Next, we created a catalog of the tissue-specific risks of 18,927 protein-coding genes (https://netbio.bgu.ac.il/trace/). As proof-of-concept, we successfully prioritized candidate disease genes in 50 genetically-diagnosed patients with rare diseases. Thus, machine learning combined with tissue omics data and information on patients' affected tissues enhances genetic and clinical understanding of tissue-specific phenotypes.

196

Monitoring health and COVID-19 using Big Data

Michael Snyder

Department of Genetics, Stanford School of Medicine, Stanford, CA, USA

Recent technological advances as well as longitudinal monitoring not only have the potential to improve the treatment of disease (Precision Medicine) but also empower people to stay healthy (Precision Health). We have been using advanced multiomics technologies (genomics, immunomics, transcriptomics, proteomics, metabolomics, microbiomics) as well as wearables for monitoring health in 109 individuals for up to 12 years and made numerous major health discoveries covering cardiovascular disease, oncology, metabolic health and infectious disease. We have found that individuals have distinct aging patterns that can be measured in an actionable period of time as well as seasonal patterns of health markers. We have also explored the effects of fiber using multiomics profiling. Finally, we have used wearable devices for early detection of infectious disease, including COVID-19 and built an alerting system for detecting health stressors that is scaleable to the entire planet. We believe that advanced technologies have the potential to transform healthcare.

209

Genetics of SSRI antidepressant use and implications for COVID19 risk

Joel Gelernter[1], Daniel Levey[1], Marco Galimberti[1], Stefany L.L. Empke[1], Kelly Harrington[2], J. Michael Gaziano[3], The Million Veteran Program[5], Murray B. Stein[4]

[1]*Psychiatry, Yale University School of Medicine and VA CT Healthcare Center, USA*; [2]*Psychiatry, Boston University School of Medicine and VA Boston Healthcare System, USA*; [3]*Medicine, Harvard Medical School, Brigham and Women's Hospital and VA Boston Healthcare System, USA*; [4]*Psychiatry, UC San Diego School of Medicine and San Diego VA, USA*; [5]*Research, US Department Veterans Affairs, USA*

The antidepressants fluvoxamine and fluoxetine have been shown to prevent hospitalization and other severe outcomes for COVID19, and these effects may extend to other selective serotonin reuptake inhibitor (SSRI) medications. One way to understand the SSRI-COVID19 relationship is via querying the underlying genetic relationships, including assessing causality using Mendelian randomization.

We undertook GWAS analyses of the trait of SSRI use in the US VA Million Veteran Program (MVP) sample. The analysis in EUR ancestry individuals included 177,494 cases (who had received SSRI prescriptions) and 268,353 controls (who had not). This resulted in discovery of 26 independent genomewide significant risk loci. SSRIs are most commonly used for depression, and therefore we might have expected similar discovery to what was seen for that trait, but there were many more risk loci for SSRI use, contrary to expectations (16 loci for depression). However, the Rg with depression (from PMID: 34045744) was 0.96; the Rg with citalopram (another SSRI) use (in UK Biobank) was 0.89. Other traits with Rgs 0.7 with SSRI use included headache, use of amitriptyline (a non-SSRI antidepressant), and inability to work due to disability. Associated SNPs mapped (for example) to DRD2 (lead locus), NRXN1, and MAD1L1.

With this information, we will investigate possible causality with respect to COVID19 outcomes and will interrogate biology and pathways with the aim of understanding what underlies the genetics of SSRI use and why it is under such strong apparent genetic influence compared to per se psychiatric traits.

218

Microbiome profiling in patients with Adenomatous Polyposis compared to sporadic subjects

Revital Kariv[1], Naomi Fliss[1], Dana Ivankovski[2], Laura Sol Grinshapon[2], Dror Hadas[3], Shira Zelber-Sagi[2], **Elhanan Borenstein**[3]

[1]*Department of Gastroenterology, Tel Aviv Sourasky Medical Center and Faculty of Medicine,Tel Aviv University, Israel*; [2]*School of Public Health, Haifa University, Israel*; [3]*Department of Clinical Microbiology and Immunology,, Faculty of Medicine, School of Computer Science, Faculty of Exact Sciences Edmond J. Safra Center for Bioinformatics Tel Aviv University, Israel*

Adenomatous-polyposis syndromes are rare inherited conditions caused by mutations in cancer-related genes with a phenotype of numerous adenomatous polyps along the gastrointestinal (GI) tract. High GI-cancer risk in polyposis requires endoscopic surveillance, while chemoprevention and lifestyle have not been shown to significantly affect polyp burden and cancer risk. In contrast, sporadic colorectal cancer (CRC) and its precursors have been robustly linked to lifestyle, metabolic-profile, and nutrition, as well as to the gut microbiome. Both polyposis and CRC patients were shown to have distinct microbiome composition and with increase specific taxa abundance. These profiles may be used for early prevention, diagnosis, and therapeutic decisions. To better characterize the microbiome of adenomatous-polyposis patients, we aimed to conduct a comprehensive comparison between polyposis and sporadic subjects, using 16S rRNA metagenomic profiling. To this end, we obtained, sequenced, and analyzed samples from 19 adenomatous polyposis patients, and 49 controls. Samples were sequenced by Merck and were analyzed using QIIME2. Our analysis identified significant differences between polyposis and controls (including subjects with and without sporadic polyps), in terms of both overall diversity and composition. Specifically, polyposis patients were associated with a significantly less diverse microbiome. Furthermore, comparing polyposis patients to controls additionally identified an increase in the abundance of Blautia and Bifidobacterium and a decrease in Bacteroides among polyposis patients. These initial findings suggest that adenomatous polyposis-specific microbiome patterns can be clearly detected even in a relatively small, heterogeneous cohort. Further studies to validate these findings and to pinpoint specific markers for clinically-relevant stratification are currently underway.

231

Prime editing to correct POLG-related epilepsy

Gautam Kok[1], Remi Stevelink[2,3], Eline Kormelink[1], Bobby Koeleman[2], Kees Braun[3], Sabine Fuchs[1]

[1]*Metabolic Diseases, UMC Utrecht, Netherlands*; [2]*Genetics, UMC Utrecht, Netherlands*; [3]*Pediatric Neurology, UMC Utrecht, Netherlands*

Background

Alpers-Huttenlocher syndrome is a severe metabolic disorder with intractable, therapy-resistant epilepsy, muscle weakness and liver disease. 70% is caused by the POLG c.1399GA (p.Ala467Thr) variant. Most children die within months to years after seizure onset due to status epilepticus or liver failure. Genetic repair of this variant may form a curative treatment strategy. We aim to use the CRISPR/Cas9 derived gene editing technique prime editing to genetically repair the most common pathogenic POLG-variant.

Methods

We developed guide RNAs (gRNAs) against the most common pathogenic POLG-variant c.1399AT using in-silico prediction algorithms. We then transfected patient-derived fibroblasts with base and prime editor plasmids, the gRNAs and a mutation-specific dual-fluorescent prime editing and enrichment reporter (fluoPEER). Editing efficiency was checked using Sanger sequencing. We analyzed various mitochondrial readouts, including mtDNA levels, mitochondrial density, oxidative phosphorylation, and complex levels.

Results

With prime editing, we were able to correct the c.1399AT mutation in 46% of patient-derived fibroblasts. We observed no unwanted editing of other nucleotides near the target site. Repaired cells showed an improvement in mitochondrial readouts.

Conclusion

We are the first to show that base and prime editing can be used to genetically correct the most common pathogenic POLG-variant causing Alpers-Huttenlocher syndrome. This technique holds promise to causal treatment for this disease.

243

The effect of the gut microbiome of adult and aging mice on metabolic characteristics

Dana Binyamin[1,*], Nir Werbner[1,*], Meital Nuriel-Ohayon[1], Atara Uzan[1], Hadar Mor[1], Atallah Abbas[2], Oren Ziv[1], Roee Guttman[2], Omry Koren[1]

[1]Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel; [2]Laboratory of Integrative Physiology (LIP). MIGAL - Galilee Research Institute, Kiryat Shmona, Israel; and Department of Animal Sciences, Faculty of Sciences and Technology, Tel-Hai College, Upper Galilee, Israel

During aging, there is a physiological decline, an increase of morbidity and mortality and a natural change in the gut microbiome. In this study, we investigated the influence of the gut microbiome on different metabolic parameters in adult and aging mice. Fecal and blood samples from adult (n=42, 100-300 days) and aging (n=32, 550-750 days) mice were collected. 16S rRNA gene sequence analysis was carried out using Qiime2. Mice weight and body composition were measured by using NMR, and insulin and leptin levels in the blood were measured using the Mouse Adipokine Magnetic Bead Panel kit. We then preformed Fecal Microbiota Transplantation (FMT) from adult and aging mice into young germ-free (GF) mice in order to examine the effect of the gut microbiome of adult and aging mice on weight, body composition and insulin and leptin levels. We also monitored food consumption and RQ 10 days after FMT using Metabolic Cages. We found that adults and aging mice have different microbiomes. We observed a high Firmicutes/Bacteroidetes ratio in aging mice compared to adult mice in addition to several genera that were significantly different between the groups. In the examined metabolic parameters, we observed significantly higher weight and fat mass and lower lean mass in aging mice along with high insulin and leptin levels in the blood. In the FMT mice, the gut microbiome from aging mice caused several metabolic changes in the young transplanted mice. Fat body mass and insulin levels were higher in the mice who received aging feces than mice receiving adult feces. In addition, they consumed more food and had higher metabolic activity (average RQ) compared to mice receiving adult feces. We conclude that aging mice have a gut microbiota that is associated with obesity, and they also exhibit metabolic parameters related to obesity. In addition, the gut bacterial population itself is sufficient to induce some of the manifestations of obesity.

244

Whole coding genome inter-clade comparison to predict global cancer-protecting variants

Lamis Naddaf and Yuval Tabach

Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada (IMRIC), Hebrew University School of Medicine, Jerusalem, Israel

In this research, we identified the missense genetic variants that have the potential to enhance resistance against cancer. Such field has not been widely explored, as researchers tend to investigate mutations that cause diseases, in response to the suffering of patients, rather than those mutations that protect from them. In conjunction with the genomic revolution, and the advances in genetic engineering and synthetic biology, identifying the protective variants will increase the power of genotype-phenotype predictions and can have significant implications on improved risk estimation, diagnostics, prognosis and even for personalized therapy and drug discovery. To approach our goal, we systematically investigated the sites of the coding genomes and picked up the alleles that showed a correlation with the species' cancer resistance. We predicted 250 protecting variants (PVs) with a 0.01 false discovery rate and more than 20 thousand PVs with a 0.25 false discovery rate. Cancer resistance in Mammals and reptiles was significantly predicted by the number of PVs a species has. Moreover, Genes enriched with the protecting variants are enriched in pathways relevant to tumor suppression like pathways of Hedgehog signaling and silencing, which its improper activation is associated with the most common form of cancer malignancy. We also showed that the PVs are more abundant in healthy people compared to cancer patients within different human races.