Tuller-con 2021: poster abstracts

section	text
Title	1. Modeling the effect of rRNA-mRNA interactions and mRNA folding on mRNA translation in organelles
Authors	Stav Carmel Ezra, Tamir Tuller
Abstract	The process of translation initiation in prokaryotes is mediated via the folding of the mRNA and its binding to the 16S rRNA component of the small ribosomal subunit. Recently it was shown that in organelles, which have originated from formerly autonomous bacteria, such as the chloroplast the 16S rRNA structure underwent evolution that affects the chloroplast's translation initiation.
	Thus, in this project we are developing a novel computational pipeline that can be used for inferring and engineering translation regulation in organelles. This pipeline should have important contribution to basic science, including the understanding of the genomic evolution and the biophysics of translation in organelles, and to biotechnology.

section	text
Title	2. Effect of human microbiome supplementation on metabolic disease and other conditions.
Authors	Shimshi Atar, Tamir Tuller
Abstract	Human gut microbiome comprises the collective genome of microbes inhabiting the gut. It is known to have protective effects against pathogens and have various important interactions with digestion and our immune system. Indeed, there is growing evidence that modifications in the microbiota composition may lead to several illnesses.
	Here, as part of a clinical study, we study the predictive effect of patient microbiome composition on the success of treating it with microbes' supplements aiming at dealing with various diseases such as Clostridium difficile infections (CDI), Anxiety and Depression, Ulcerative Colitis, Metabolic Disease, and Solid Tumor.

section	text
Title	3. Circulating Micro-RNAs As Biomarkers of Visceral Obesity-Related Adipose Tissue Disease (OrAD)
Authors	Nataly Makarenkov, Yulia Haim, Assaf Rudich, Isana Veksler-Lublinsky
Abstract	Background: Individuals with obesity differ in the degree of increased health risk imposed by obesity, which is known to associate with remodeling of adipose tissue, particularly of the visceral-fat depot (VAT). VAT's Obesity related Adipose tissue Disease (OrAD) may therefore assist in patients' risk stratification/classification. However, visceral fat is not accessible for direct sampling. Here we explored whether circulating miRNAs can serve as clinically-accessible biomarkers of visceral-fat OrAD in obesity.
	Methods: In n=14 patients with obesity undergoing elective abdominal surgery, VAT-OrAD was assessed histologically for adipocyte hypertrophy and macrophage infiltration, inflammation-related mRNA expression, and adipocyte dysfunction (serum adiponectin/leptin). Circulating and VAT miRNAs were analyzed by Next Generation Sequencing.
	Results: 103 circulating miRNAs with ≥5 copies per million correlated significantly to one or more VAT-OrAD features. Hierarchical clustering revealed 35 miRNAs that correlated only with VAT inflammation, 26 only with histology features, and 11 solely with adipocyte dysfunction. 30 miRNAs correlated with 2 OrAD features, while 1 correlated with all parameters. Correlation of these circulating miRNA with their expression in VAT identified a sub-set for which VAT may be a major direct source. Bioinformatic analyses showed that miR-486-3p, miR-486-5p, miR-22-5p and other miRNAs were previously implicated in obesity and obesity related disorders such as Metabolic Syndrome, Diabetes Mellitus Type 2, and obesity-related cancers (hepatocellular carcinoma, colorectal cancer).
	Conclusions: Circulating miRNAs reflective of VAT-OrAD features are candidate clinically-accessible biomarkers for VAT health, and may assist in better classifying/risk-stratifying patients with obesity.

section	text
Title	4. Machine Learning for prediction of sRNA-mRNA interactions
Authors	Shani Cohen
Abstract	Bacterial small RNAs (sRNAs) are relatively short non-coding RNA molecules (~50-500 nt) that play a significant role in the regulation of various bacterial functions, such as virulence, environmental sensing, metabolism, and gene expression. The sRNAs act as post-transcriptional regulators either by binding specific proteins to alter their activity or by base-paring with target mRNAs. The two major classes of base-pairing sRNA are commonly called cis-encoded and trans-encoded. The cis-encoded sRNAs, a.k.a. antisense RNA, are transcribed from the strand complementary to the mRNA they regulate, whereas the trans-encoded share only a partial sequence complementarity with their targets and thus enable to regulate multiple genes. Similar to microRNAs in eukaryotic, the trans-encoded sRNAs modulate the translation, processing, and/or stability of their target mRNAs by short interactions.
	Computational prediction of sRNA-mRNA interactions can provide candidates for biological validation, and speed up the discovery of the sRNA regulation mechanism in bacteria. Although several approaches have already been developed, this task is still challenging. In this research, we propose a novel method for sRNA-mRNA interactions prediction that is based on a weighted non-negative matrix factorization. As opposed to previous methods, our model doesn't require the synthesis of negative data. Our model was trained over interaction data collected from very recent researches that applied high-throughput technologies, while evaluation was done over an independent dataset of biologically validated interactions.

section	text
Title	5. Evolutionary Stability Optimizer (ESO): A Novel Approach to Identify and Avoid Mutational Hotspots in DNA Sequences while Inducing High Expression Levels
Authors	Itamar Menuhin1, Matan Arbel2, Niv Amitay3, Karin Sionov4, Shaked Bergman4, Omer Edgar2 ,Doron Naki2, Itai Katzir2 and Tamir Tuller4
Abstract	Modern synthetic biology procedures rely on the ability to generate stable genetic constructs that keep their functionality over long periods of time. However, maintenance of these constructs requires energy from the cell and thus reduces the host's fitness. Natural selection results in loss-of-functionality mutations that negate the expression of the construct in the population. Current approaches for prevention of this phenomena focus on either small scale, manual design of evolutionary stable constructs; or the detection of mutational sites with unstable tendencies. We designed a tool we named the Evolutionary Stability Optimizer (ESO), that enables large-scale automatic design of evolutionary stable constructs with respect to both mutational and epigenetic hotspots, and allows users to define custom hotspots to avoid. Furthermore, our tool takes the expression levels of the input constructs into account by considering the GC content and the codon usage of the host organism, defining a trade-off between stability and gene expression. In this study we present the many features of the ESO and show that it accurately predicts the evolutionary stability of endogenous genes. The ESO was created as an easy-to-use, flexible platform based on the notion that directed genetic stability research will continue to evolve and revolutionize current applications of synthetic biology.

section	text
Title	6. Automatic modeling of tissue specificity of viruses
Authors	Alma Davidson
Abstract	As viruses cannot self-replicate, they are highly dependent on their hosts to carry out their cellular mechanisms. As a result, the differences between the machineries of the host cells leads to viral affinity for specific tissues, which can be exhibited on the genomic level. In this project, we focus on the coronavirus variants and perform genomic analysis to indicate the viral affinity of the coronavirus.
	This analysis is based on the ChimeraARS algorithm, which enables us to estimate the tendency of the virus to share common long substrings with its host. Thus, we can explore novel tissue specific patterns indicating viral specificity.
	The discovered data may be utilized to construct a computational model, uncovering profound aspects related to viral tissue-specific-adaptation, both in the coronavirus and in general.

section	text
Title	7. Accelerating whole cell simulations of mRNA translation using a dedicated hardware
Authors	David Shallom
Abstract	In recent years intracellular biophysical simulations have been used with increasing frequency not only for answering basic scientific questions but also in the field of synthetic biology. However, since these models include networks of interaction between millions of components, they are extremely time consuming and cannot run easily on parallel computers.
	In this study, we demonstrate for the first time a novel approach addressing this challenge by using a dedicated hardware designed specifically to simulate such processes. As a proof of concept, we specifically focus on mRNA translation, which is the process consuming most of the energy in the cell. We design a hardware that simulates translation in Escherichia coli for thousands of mRNAs and ribosomes which is in orders of magnitude faster than a similar software solution.
	With the sharp increase in the amount of genomic data available today and the complexity of the corresponding models inferred from them, we believe that the strategic suggested here will become common.

section	text
Title	8. Non-optimal codon usage determines the protein level dynamics during cell-cycle
Authors	Mahua Bhattacharya, Sumit Mukherjee, Milana Frenkel-Morgenstern
Abstract	Translation of mRNA is a major factor in protein expression, and the adaptation to the optimal codon usage is a substantial part of it. Codon usage pattern of genes has an implication on post-transcriptional regulation to changes in translational profiles. Some studies have indicated that the codon usage pattern determines the stability of proteins, which in turn affects the post-translational mechanism and functional activity of proteins. Since synonymous codons are attached to the anti-codons in tRNA with different affinities, the codons are translated in different efficiencies. It has been shown that cell-cycle regulated genes are biased towards low-affinity codons and adopt non-optimal codon usage. That observation has been made in-silico for human genes, which led us to hypothesize that protein encoded with low-affinity codons will result in oscillations in the protein levels during the cell cycle, while similar proteins encoded with high-affinity codons will present constant levels throughout the different cell-cycle phases. To analyze the steady-state RNA expression at the different phases of the cell cycle, we downloaded the publicly available RNA-seq data of three human cell lines, where the authors performed mRNA sequencing of individual G0/G1, G1/S, and M phases. We selected the top two genes that are cell cycle-dependent and changed their codons to optimized codons and assessed the stability of proteins stability. We have found that the stability of proteins is higher in non-optimal codon usage compared to their synonymous counterparts. We performed Our findings suggest that there is a difference in the dynamics patterns of proteins and RNA encoded by the non-optimal vs optimal codons.

section	text
Title	9. Computational Modeling of the effect of miRNA binding sites distribution on mRNA stability
Authors	Sharon Bader, Tamir Tuller
Abstract	The modeling of miRNA-mRNA interactions has various important applications in synthetic biology and human health. However, this research question is specifically challenging since the efficiency of mRNA down regulation by miRNA is affected by dozens of features including either competition or synergism among miRNAs and mRNAs.
	In this project we are developing a predictive computational model that considers for the first time both the distribution of miRNA binding sites along the mRNA and their strength on the efficiency of mRNA stability regulation.

section	text
Title	10. Determinants of Efficient Modulating of Ribosomal Traffic Jams
Authors	Sophie Vinokour
Abstract	mRNA translation is the process consuming most of the cellular energy. Thus, this process is under strong evolutionary selection for its optimization and rational optimization or reduction of the translation process efficiency can impact the cell growth rate. Algorithms for modulating growth rate can have various application in biotechnology, medicine, and agriculture. In this study, we demonstrate that the analysis of these algorithms can also be used for understanding translation.
	We specifically describe and analyze various generic algorithms, based on comprehensive computational models and whole cell simulations of translation for introducing silent mutations that either reduce or expand ribosomal traffic jams along the mRNA. As a result, more/less resources are available, respectively, for the cell, promoting improved or reduced cells growth-rate. We then explore the cost of these algorithms' performance, in terms of their computational time, the number of mutations they introduce, the modified genomic region, the effect on local translation rates, and the properties of the modified genes.
	Among others, we show that mRNA levels of a gene is a much stronger predictor of the effect of its engineering on the ribosomal pool than the ribosomal density of the gene. We also demonstrate that the mutations at the ends of the coding regions that to have much stronger effect on the ribosomal pool. In addition, we report two optimization algorithms that exhibit a tread-off between the number of mutations they introduce and their running time.

section	text
Title	11. Engineering and testing the translation process in prokaryotes via examination of plasmids library, TCP-seq and ribosome profiling
Authors	Larissa Fine, Rachel Cohen-Kupiec, Tamir Tuller
Abstract	Various genome-based studies from recent years have provided new predictions related to the regulation of translation in prokaryotes and the way it is affected by features of the transcript sequence. Based on various computational analyses that have been performed in our lab (Tuller's) it was suggested that in prokaryotes the interactions between the rRNA (ribosomal RNA) and the mRNA both UTRs and coding regions affect all the mRNA translation steps. In addition, the local folding of the mRNA in different parts of the transcript, can also affect all translation steps. While computational models based on these discoveries can potentially help modulating the translation rate of an mRNA via the modification of its nucleotide composition, various aspects of such models need yet to be investigated for them to be fully understood.
	The aim of my research program is to perform a set of experiments in E. coli to investigate these aspects of translation in depth. These will include a screen of a large library of heterologous gene modifications and their impact on protein translation using state of the art techniques as ribosome profiling (Ribo-seq) and Translation Complex Profile sequencing (TCP-seq) a technique which was not used before in prokaryotes.
	The new data will elucidate how specific features of the mRNA sequences affect translation and will help to verify and confirm parts of our mentioned above models. This in turn will lead to better tuning of our models related to all translation steps (pre-initiation, termination, re-initiation and more), and to a better understanding of the way they are encoded in the transcripts of prokaryotes such as E. coli.

section	text
Title	12. Estimating the Predictive Power of Silent Mutations on Cancer Classification and Prognosis
Authors	Tal Gutman, Guy Goren, Omri Efroni, Tamir Tuller
Abstract	In recent years it has been shown that silent mutations, in and out of the coding region, can affect gene expression and may be related to tumorigenesis and cancer cell fitness. However, the predictive ability of these mutations for cancer type diagnosis and prognosis has not been evaluated yet. In the current study, based on the analysis of 9,915 cancer genomes and approximately three million mutations, we provide a comprehensive quantitative evaluation of the predictive power of various types of silent and non-silent mutations over cancer classification and prognosis. The results indicate that silent-mutation models outperform the equivalent null models in classifying all examined cancer types and in estimating the probability of survival 10 years after the initial diagnosis. Additionally, combining both non-silent and silent mutations achieved the best classification results for 68% of the cancer types and the best survival estimation results for up to nine years after the diagnosis. Thus, Silent mutations hold considerable predictive power over both cancer classification and prognosis, most likely due to their effect on gene expression. It is highly advised that silent mutations are integrated in cancer research in order to unravel the full genomic landscape of cancer and its ramifications on cancer fitness.

section	text
Title	13. Analysis of selection on mRNA secondary structure strength in protein-coding sequences within conserved protein families
Authors	Michael Peeri and Tamir Tuller
Abstract	The mRNA molecule can form secondary structure through short-range base-pairing interactions determined by the nucleotide sequence. These structures compete with other interactions of the mRNA strand and are suspected to influence many gene expression processes. In this study we attempt to study selection acting on mRNA secondary structure strength in finer resolution than done before, by analyzing families of orthologous proteins. Within a conserved protein family, the nucleotide sequences observed are the result of multiple selective pressures maintaining the folded protein's function, but also efficient and accurate translation, protein folding and degradation and other steps in the gene expression process. Focusing on an homologous position within the family, (i.e. at the nucleotides encoding a homologous amino-acid), we can assume many of these selective pressures act similarly on all members of the family, justifying their analysis as samples taken from a single distribution. We can therefore perform statistical tests (given sufficient data) to infer which selective pressures are needed to explain the observations at any homologous position. The strength of purifying selection on the amino-acid level can be measured against codon bias, mRNA secondary structure bias and other characteristics of the coding sequence to reveal how these processes are regulated by the coding sequence and answer the following questions:
	• Which factors explain the huge variation between genes and regions within a genome?
	• What are the relationships between different traits selected for in different regions of the coding sequences? do some traits tend to be selected together in the same regions?
	• Does the coding sequence provide enough flexibility to allow arbitrary traits to be selected for in the same region, or are there trade-offs between traits?
	• Does selection for secondary structure strength accompany specific protein domains or protein secondary structures?

section	text
Title	14. The experimental validation of gene-gene fusions as a diagnostic tool for the detection of Glioblastoma Multiforme (GBM).
Authors	Olawumi D. Giwa, Gidi Baum, Milana Frenkel-Morgenstern
Abstract	Gliomas are tumors with etiologies in glial cells resident in the brain. They account for approximately 80% of all malignant brain tumors. High-grade gliomas such as glioblastoma multiforme (GBM) are the most common types of primary malignant brain tumors. The standard treatment paradigm for patients with GBM is still very limited in terms of survival.
	Therefore, we selected GBM as a model to investigate and subsequently develop an early non-invasive (liquid biopsies) diagnostic tool and precision medicine methodology to monitor patients using blood plasma-derived circulating cell-free DNA (cfDNA). Our laboratory analyzed cfDNA samples obtained from GBM patients (n=27) and compared them with samples from healthy individuals (n=20).
	We have demonstrated that the concentration of cfDNA in the plasma of GBM patients is significantly higher than in the plasma of healthy individuals. In addition, Next-generation sequencing (NGS) analysis of the cfDNA revealed several putative chimeric DNA in the GBM samples that are not present in samples from healthy individuals. We are currently trying to confirm and validate the NGS findings by several different methods 1) Probe-based Polymerase Chain Reaction (qPCR) assays on the cfDNA samples in which the probe is located directly on the fusion site; 2) cloning and Sanger sequencing of PCR product/s from the cfDNA; and 3) Oxford NanoPore sequencing.
	The future aspect of these findings can be used for novel drug targets and precision medicine for GBM patients to improve the patient's prognostic outcomes.

section	text
Title	15. How environment shapes the global codon and amino acid usage pattern of microbial communities: A cross-biome metagenomic analysis.
Authors	Arup Panda, Prof. Tamir Tuller
Abstract	Microbes are everywhere in the world. However, understanding the attributes that characterize different microbial communities is still very challenging. Metagenomics is an emerging field that helps to understand the genetic and functional capabilities of microbial communities at the system level without any need for culturing. Metagenomics is also increasingly being recognized as a powerful tool to understand the community-level attributes that distinguish different communities. One such community-level attribute is codon usage which was widely used to characterize microbes at the species level and was shown to be related to various fundamental processes such as gene expression, metabolism, and horizontal gene transfer. However, little is known about the codon usage of microbial communities as a whole. Specifically, how the microbes from different ecological niches are related in terms of their codon usage is still not clear. Here in this study we considered protein-coding DNA sequences of more than 100 metagenomic samples collected from diverse environmental sources and analyzed their codon usage pattern. To compare and contrast codon usage of microbes in these communities here we mainly considered three matrices namely, codon adaptation index (CAI), effective number of codons (ENCs) and directional codon bias score (DCBS). Here we found cohesive signals in the codon usage of studied communities, however further studies are ongoing to understand the trend more clearly.

section	text
Title	16. Detecting and Understanding Meaningful Cancerous Mutations based on Computational Models of mRNA Splicing
Authors	Nicolas Lynn, Prof. Tamir Tuller (MSc supervisor)
Abstract	Currently, most of the predictive tools for cancer diagnosis and prognosis are based on mutations that have effects on the amino acid content of proteins. However, in recent years it has been shown that silent mutations (i.e. mutations that apparently do not change the amino acid content), can affect gene expression and the resultant proteins, and thus tumorigenesis and cancer cell fitness. Therefore, developing approaches for the deciphering of silent mutations should have major contribution to disease characterization and therapeutic development.
	One fundamental aspect that can significantly affect the produced proteins is mRNA splicing. The regulatory codes of this process appear both in the coding and non-coding parts of the transcript; mutations that affect these regulatory codes may seem silent in some cases but they have major effect on the nature of the generated proteins.
	In this work we developed a computational pipeline for detecting meaningful cancer mutations that affect splicing. Among others, the pipeline is based on a highly accurate deep neural network that predicts splice sites from arbitrary mature mRNA sequences and on models that predict the functional contribution of different part of the protein sequences.
	Based on this pipeline we propose a novel way to classify silent mutations as either drivers or passengers in cancer progression which has various immediate applications.

section	text
Title	17. Modeling evolution of local genetic instability.
Authors	Hadar Ben Shoshan and Tamir Tuller
Abstract	Genome instability can be defined as an enhanced tendency for the genome to acquire mutations, especially in mutational hotspots. Currently there are no accurate models that can capture the evolution of local genetic instability and can be used for engineering this aspect.
	The aim of our study is to understand and model for the first time how local genetic instability undergoes evolution in microorganisms and how it is affected by various features of the genome. This will be used for developing novel approaches for engineering genetic stability and thus has vast biotechnological potential.

section	text
Title	18. Functional interplay between microRNAs, RNA binding proteins, and alternative polyadenylation in gene regulation in C.elegans
Authors	Stav Lutzky & Isana Veksler-Lublinsky
Abstract	Regulation of gene expression is fundamental for proper development, homeostasis, and adaptation to the environment for all living organisms. Post-transcriptional gene regulation, which takes place between the transcription and the translation of a gene, is largely controlled by two classes of regulators, microRNAs (miRNAs) and RNA-binding proteins (RBPs). MiRNAs are short non-coding RNA molecules that hybridize to complementary sequences on target mRNAs, usually located in the 3'UTR region, and repress their translation or mediate their degradation. RBPs also perform their function by binding to mRNAs; however, their binding sites can be located in various regions including 5'UTRs, coding sequences, and 3'UTRs. Unlike miRNAs' repressive role, the regulatory activity of RBPs may be positive or negative, depending on the protein, the mRNA, and the biological context. In recent years, a wide repertoire of functional connections between miRNAs and RBPs has been discovered, uncovering a new level of complexity in gene expression regulation. These connections include mutual regulation of the same target mRNAs by both miRNAs and RPBs.
	Since both miRNAs and RBPs bind to the 3'UTR region of mRNAs, its length may add a layer of complexity to the interplay between miRNAs and RBPs. 3'UTR length is regulated by alternative polyadenylation (APA). By changing the position of polyadenylation, APA can generate transcripts with multiple 3' UTR isoforms, each containing distinct regulatory elements (e.g., miRNA and RBP binding sites). In this study, we used C.elegans UTRome.org data that provides 3'UTR variants across different tissues, to study the coordination between miRNAs and RBPs in tissue-specific gene regulation. We identified pairs of miRNA-RBPs that show statistically significant differential co-targeting across tissues. Our results provide evidence for complex gene regulatory networks that involve multiple factors such as miRNAs and RBPs.

section	text
Title	19. Synthetic Rational Design of Potential Zika Virus Vaccine Based on a Computational Model
Authors	Zafrir Zohar, Roopin Modi, Zarai Yoram, Siridechadilok Bunpote, Julander Justin, Tuller Tamir
Abstract	Synthetic virology is an important multidisciplinary scientific field, with emerging applications in biotechnology and medicine, aiming at developing methods to generate, understands, and engineer synthetic viruses. Many viruses of the Flaviviridae family, including the Zika virus (ZIKV), are widespread pathogens of significant importance to human health. Yet, and despite extensive research, there are currently no approved vaccines available for this virus. Therefore, designing attenuated synthetic virus versions and controlling their virulence and rate of replication in various models is a fundamental milestone in the continuing efforts in fighting the diseases they cause. Specifically, this approach helps improving our understanding of the genomes of additional viruses and may promote developing potential vaccines and virus based therapies. Using a computational based pipeline model for the rational design of attenuated synthetic RNA viruses, we generated dozens of ZIKV variants, based on procedures that preserve viral amino acid content, but affect various functional silent aspects of their genome. Results in Vero cells show a gradient of attenuations of the examined synthetic ZIKV strains, as well as mRNA levels, which correlate with our model predictions. Furthermore, results in mouse model demonstrated active immune response and higher survival ratio for mice vaccinated with the synthetic attenuated ZIKV variants, as well as significant long-term protection. Further study is warranted to determine whether these attenuated synthetic viruses would be suitable for clinical use. Our tested model based abilities in design and synthesis of such viruses may well contribute to the development of vaccines and virus based therapies.

section	text
Title	20. Genomic determinants of viral host diversity
Authors	Marina Parr, Tamir Tuller, Dmitrij Frishman
Abstract	The efficiency of the viral infection depends on the success of the attachment of the virus to the host cell, effective recruitment of the host genetic apparatus and avoidance of host defense system. The relative high viral mutation rate promotes the generation of new viral variants; those mutations that provide viruses any advantages during the infection become fixed in the viral genomes. Some of these mutations can facilitate the binding between receptors at the surface of the host cells and viral antireceptors [1]. Other mutations that are accumulated during the long passaging of the virus in the host cells can also play a role in virus-host adaptation - viruses are known to mimic host's nucleotides, codons, amino acids usage and even more complex codes [2], RNA structures [3]; this allows viruses to use the host protein synthesis system more effectively and avoid the host defense system and thus improve their replication efficiency. Alterations in the viral genomes can not only increase the fitness of the virus to the specific host but also broaden its host range. Sometimes even a point mutation is enough to grant the virus the ability to infect a new host. For example, with the substitution E627K in PB2 protein bird strain of Influenza A virus becomes able to infect cells of mammals [4].
	Many viruses are known to infect several different organisms. The reasons for that might occur at both host and virus sides: high similarity between host organisms might reduce the challenge for the virus to infect all of them; on the other hand, a virus can be adapted to infect several very different species. In this project we aim at finding fundamental features of viral genomes that can enable a virus to infect different organisms. For quantitative characterization of how wide is the hosts range of the specific virus we suggest to use a new measure that we named host diversity. Host diversity of a virus is calculated as the sum of branch lengths of the phylogenetic tree that is built using the hosts SSU rRNA sequences. Thus, this measure characterizes not only the number of the hosts but also how different are these species. Analysis of correlation between host diversity and sequence features in orthologous groups of viral genes and proteins enabled us to hypothesize about the mechanisms of viral adaptation to many different hosts.
	References
	1. S. Szepanski et al. (1992) Virology, 188(1):85-92
	2. E. Goz et al. (2018) Bioinformatics, 34(19): 3241-3248
	3. J. Witteveldt et al. (2014) Nucleic Acids Res., 42(5):3314-29
	4. M. Hatta et al. (2007) PLoS Pathog, 3(10):1374-9

section	text
Title	21. Microbiome Engineering
Authors	Communique, iGEM 2021 team
Abstract	The microbiome is a network of various microbial organisms organized in interconnected communities present in synthetic and natural environments, from simple life forms to humans. These communities have extensive effects on their systems, including plant growth, oil bioremediation, and even human conditions such as diabetes and autoimmune diseases.
	In order to accurately engineer microbiomes, genetic modifications must be limited to certain members of the community - selectivity is a key factor in changing functions in microbial environments. Moreover, the introduction of genes into unwanted hosts can have unintended consequences such as large ecological impacts, posing major biosafety issues.
	We aim to develop a software tool that automatically designs a microbiome-specific plasmid that is selectively expressed in certain parts of the bacterial population. Our algorithm fine-tunes genetic information, from single genes to whole genetic circuits, to be expressed optimally for target organisms while preventing expression in other species by impairing the same parameters for them. This ensures the safety of microbial engineering despite horizontal gene transfer, allowing GMOs to progress beyond the borders of the supervised lab while maintaining safety standards.

section	text
Title	22. Selection for efficient gene expression in lncRNA
Authors	Yuri Klayman under supervision of Prof. Tamir Tuller
Abstract	lncRNA are relatively long (more than 200 nucleotides) RNA producing genes that are believed to have a functional transcriptional product that doesn't encode for a functional protein.
	lncRNA are common in most organism with what seems like a trend for being more common in more "complex" organisms. lncRNA molecules are transcribed by RNA Pol II similarly to mRNA and most undergo splicing as well as capping and the addition of a Poly-A tail. While much of the lncRNA remains in nucleus and is presumed to have a role in gene expression regulation, lncRNA are also known to leave the nucleus.
	While transcription level per lncRNA "gene" is relatively low, in humans as much as half of the mature RNA in the cytosol seems to be lncRNA and not mRNA.
	Since the lncRNA molecules reach the cytosol they are bound to encounter ribosomes but translating RNA molecules that do not actually encode for proteins is costly and wasteful in energy and resources and potentially detrimental to the cell as the random peptide can accumulate and cause disruption.
	We plan to explore the lncRNA sequences looking for evidence of sequence features, motifs and other factors such as the RNA structure that affect the recruitment of ribosomes by these genes, the unintended translation efficiency and degradation of the RNA and resulting peptide.
	Furthermore, we hope finding such features will allow us to predict expression levels of lncRNA genes.

section	text
Title	23. Designing DNT sensing bacteria based on computational models
Authors	Shir Bahiri Elitzur, Etai Spigel, Shimshon Belkin, Tamir Tuller
Abstract	The detection of buried landmines is a humanitarian issue of global proportions in acute need of a practical solution. Current mine detection technologies require personnel in the immediate area of the mines, along with the obvious risks involved. Innovative computational algorithms were employed to enhance the performance of a bacterial (E. coli) biosensor genetically engineered to detect explosive vapors above buried landmines. New bioreporters engineered with these modifications exhibited over a 10-fold increase in signal intensity in the presence of 2,4-dinitrotoluene and a two-fold reduction in its detection threshold. This research presents that the whole-cell biosensor has succeeded in dramatically enhancing the remote detection capabilities of buried explosive devices, thereby providing a potential answer to the problem of detecting that presently has no technical solution. Furthermore, this breakthrough is expected to be further utilized in other applications of cell-based environmental monitoring.

section	text
Title	24. Computational analysis of novel sense-antisense chimeric transcripts reveals their abundance and potential regulatory roles in human cells
Authors	Sumit Mukherjee, Rajesh Detroja, Milana Frenkel-Morgenstern
Abstract	Many human genes are transcribed from both strands and produce sense-antisense gene pairs. Sense-antisense (SAS) chimeric transcripts are produced upon the coalescing of exons/introns from both sense and antisense transcripts of the same gene. SAS chimera was first reported in prostate cancer-cell. Subsequently, numerous SAS chimeras have been reported in the ChiTaRS database. Still, the functional implications and evolutionary significance of SAS chimeras remain elusive. We investigated the structural and functional aspects of SAS chimeras. We found that long palindromic sequences present at or near the junction sites of most SAS chimeras. We also predicted that most of SAS chimeric transcripts are long non-coding RNA (lncRNA) transcripts. Analysis of RNA secondary structure often plays a significant role in determining the function of lncRNA transcripts. Hence, we were interested in studying structural aspects of these SAS chimeras, particularly the functional role of the palindromic sequences. We found that SAS chimeras form long hairpin-like structures along the length of the palindromic regions, indicative of their possible function as double-stranded RNA (dsRNA) that can serve to inhibit gene expression. Furthermore, RNA-RNA interaction analysis by free energy minimization uncovered the potential interaction of SAS chimeras with their parental mRNAs. This result indicates that this long hairpin-like structure could enable SAS chimeras to interact with their parental mRNA transcripts and regulate their expression in response to different cellular conditions. Finally, we found several SAS chimeras in the RNA-seq data of different healthy human tissues and detected their potential orthologs in mice, highlighting their possible regulatory roles. Our study is the first comprehensive analysis of SAS chimeras in humans and established their potency in functional regulation.

section	text
Title	25. Modelling the phenotypic outcomes of CRISPR-Cas genome editing technologies
Authors	Shai Cohen, Shaked Bergman and Tamir Tuller
Abstract	Silencing genes is a common procedure in many gene-engineering endeavors and has become easier with the introduction of CRISPR-based tools. These tools often work by creating double-stranded breaks (DSB's) in the DNA and letting the cell's repair-mechanisms induce mutations that'll silence the target gene, for example by mutating its start codon. Many articles deal with where to create these DSB's in the genome and how to make sure there'll be no off-target activity in other genes we'd like to leave untouched. However, these articles usually fail to address the effect of the induced mutations on many elements of gene expression, if they indeed have any effect at all. Instead, they assume that if a gene is mutated in any way then its expression will be changed as well, be it the expression of the on\off target gene. In this article we introduce a tool that will check the effect of any induced mutation in any gene on its translation, transcription and splicing. In addition, the tool will check if the induced mutation is in a region that is conserved across species or in humans. Furthermore, it will check if the induced mutation is likely to be cancerous. Using this tool people will be able to plan better where and how to mutate genes in a manner that will maximize the chance of silencing them with minimal effect on off-target genes.

section	text
Title	26. Exosomal DNA as a novel biomarker in brain cancers diagnostics
Authors	Lee Azolai, Vikrant Palande, Milana Frenkel-Morgenstern
Abstract	Gliomas are the most prevalent form of brain tumors, representing 30% of brain and CNS tumors, which progress into 80% of malignant brain tumors. To date, the only standard diagnostic technique available for this class of tumor is tissue biopsy, which can be pernicious to patient health. Liquid biopsy, a recently developed technique, will reduce such effects of biopsy. Liquid biopsy uses cell-free DNA, cell-free RNA, cell-free proteins and exosomes circulating in the blood as tumor-derived material. These materials thus serve as tumor biomarkers for diagnosis and prognosis. Exosomes are encapsulated membranous vesicles that are continuously released into body fluids from tumorous and non-tumorous tissues. Tumor-derived exosomes bear tumor-specific molecules which can provide detailed data about the development and progression of particular tumors. Our efforts observed DNA fragments with exosomes secreted by the LN229 and CCFSTTG1 glioblastoma cell lines into the culture medium. Our findings raise the possibility that exosomes from different cell lines contain unique sets of DNA fragments, differing in terms of both cell and molecular distribution. We further noted that exosomes secreted by the two cell lines into the culture medium included trace amounts of exosomal DNA fragments. Further studies revealed that the two cell lines produce different numbers of exosomes and that differences also exist with respect to the quantities of exosomal DNA and cell-free DNA released by the glioma cell lines considered. On the basis of our findings, it would appear that specific cell-derived exosomes contain unique sets of exosomal DNA. We, moreover, concluded that analysis of exosomal DNA from plasma cells of glioma patients might help establish liquid biopsy as a diagnostic test for identifying particular types of tumors.

section	text
Title	27. Seasonal UV exposure and Vitamin D: Association with the dynamics of COVID-19 transmission in Europe
Authors	Sunanda Biswas Mukherjee, Alessandro Gorohovski, Eugene Merzon, Eliad Levy, Sumit Mukherjee, and Milana Frenkel-Morgenstern
Abstract	Several recent studies demonstrated that low plasma 25(OH) vitamin D levels are associated with the risk of COVID-19 infection. The primary source of vitamin D production in humans is environmental UV. In many viral respiratory diseases, peak infection rates are observed during winter due to reduced UV exposure and low temperatures. In Europe, the second wave of COVID-19 began early in the winter of 2020. Investigating the impact of seasonal temperature and UV exposure on COVID-19 transmission could thus aid in prevention and intervention. As such, we first performed a comprehensive meta-analysis of all related published literature based on association of vitamin D and COVID-19, which supported the concept that the level of vitamin D level is a critical risk factor for COVID-19 infection. Next, to understand the potential impact of seasonal UV and temperature levels on COVID-19 cases, we analyzed meteorological data and daily COVID-19 cases per million in the populations of 26 European countries. We observed that low temperature, UV index, and cloud-free vitamin D UV dose (UVDVF) levels are negatively correlated with COVID-19 prevalence in Europe. Further, a distributed lag non-linear model (DLNM) was used to assess the non-linear delayed effects of individual seasonal factors on COVID-19 cases. Such analysis highlighted a significant delayed impact of UVDVF on the cumulative relative risk (RR) of COVID-19 infection. The findings of this study suggest that low UV exposure can affect the required production of vitamin D in the body, which substantially influences the dynamics of COVID-19 transmission and severity.

section	text
Title	28. Analyzing Recurrent Mutations in Cancer Patients
Authors	Yoram Zarai and Tamir Tuller
Abstract	One of the most critical criterion for identifying driver cancer mutations, i.e. mutations that drive the process of cancer progression, is their recurrence among cancer patients.
	Here, we utilize a TCGA dataset of somatic cancer mutations from thousands of patients with diverse cancer types to analyze pan-cancer recurrent mutations. Specifically, we examine recurrent mutations among alternative allele frequencies in the (healthy) population, and their distribution among conservation regions. In addition, we investigate the role of regulatory recurrent mutations.
	We find that most of the non-silent highly recurrent mutations are located in highly conserved sites, whereas highly silent ones are mostly located within lowly conserved sites. In addition, highly recurrent mutations exhibit low alternative allele frequencies in the population, and a significant negative correlation (-0.7) was found between alternative frequency alleles in the population and in cancer patients. As compared to non-recurrent mutations, the average alternative allele frequency is higher within recurrent mutations, even when considered separately over different transcript regions.