Title

Computational quantification of immune cell subpopulation to predict clinical outcomes in bladder cancer

Authors

Marcela Aguilera-Flores,[1,2] Robert S. Svatek,[3] and Jianhua Ruan[1]
[1]Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX, [2]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and [3]Department of Urology, The University of Texas Health Science Center at San Antonio, San Antonio, TX

Abstract

To test our novel hypothesis that anti-tumor immune activity in cancer patient is predictive of clinical outcomes, we propose two complementary approaches to computationally determine immune cell subpopulation profiles from patient gene expression data. In both approaches, the data was analyzed with scripts written in MATLAB and Stata 10.0. The first approach was to use regression based analysis to find a correlation between immune cell subpopulation in a human genome database (Fantom5) and the clinical outcome of patients with bladder cancer in the Cancer Genome Atlas consortium. In this case, the gene expression levels of each patient are assumed to be a weighted average of gene expression levels in the constituting cell subpopulations. In the second approach we used a signature based analysis to investigate how treatment affects immune cell subpopulations. We used data from patients' before and after treatment and immune cell subpopulation signature genes. With the last approach we have seen preliminary results that suggest that some immune cell sub-populations have higher variations than others. High correlation between patient's survival was found using t-test and p-values < 0.1. Some of the cell subpopulations with p-values < 0.1 are CD9+ B cells (0.0803), Gama Delta T Cells (0.0824), Granulocyte Macrophage (0.0811), CD4+CD25+-CD45RA – Memory Conventional T Cells (0.0747), Macrophage (-0.0953), CD133+ Stem cells (-0.0949), and CD8+ T Cells (-0.0901). Research is underway to confirm this result, and to continue analyzing the data collected in both approaches. Future work will include building a model for predicting patient survival and drug responses using the immune cell sub-population profiles as predictors.

Title

Review Presentation: “A novel approach for multi-SNP GWAS and its application in Alzheimer’s disease”

Presenter

Abel Alemeshet, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

This poster is a review of the paper “A novel approach for multi-SNP GWAS and its application in Alzheimer’s disease” by Bodily *et al*. (2016). The abstract as it appears in the original publication is as follows:

“Genome-wide association studies (GWAS) have effectively identified genetic factors for many diseases. Many diseases, including Alzheimer’s disease (AD), have epistatic causes, requiring more sophisticated analyses to identify groups of variants which together affect phenotype. Based on the GWAS statistical model, we developed a multi-SNP GWAS analysis to identify pairs of variants whose common occurrence signaled the Alzheimer's disease phenotype. Despite not having sufficient data to demonstrate significance, our preliminary experimentation identified a high correlation between GRIA3 and HLA-DRB5 (an AD gene). GRIA3 has not been previously reported in association with AD, but is known to play a role in learning and memory.”

Title

Comparison of AutoDock and Glide in docking Suvorexant to the crystal structure of human OX_2 receptor (hOX_2R)

Authors

Madan Baral,[1] Ricardo Avila,[1] John Mohl,[1,2,3] Ming-Ying Leung,[1,2,3,4] and Rachid Skouta[5]
[1]Bioinformatics Program, [2]Computational Science Program, [3]Border Biomedical Research Center, [4]Department of Mathematical Sciences, and [5]Department of Chemistry, The University of Texas at El Paso, El Paso, TX

Abstract

G protein-coupled receptors (GPCRs) are a ubiquitous family of seven-transmembrane proteins that play important roles in signal transduction pathways making them potential targets of modern drugs. In the past decades, a vast array of tools for ligand-protein docking have become available, and given their number and diversity, few comparative studies have been made. It is desirable to develop a pipeline specific to GPCRs, using the best available software for validating the binding affinity of potential GPCR targeted drugs. Here, we present the comparative docking results of the potent FDA-approved narcolepsy drug Suvorexant along with another three high-affinity antagonists on the recently crystallized OX2 receptor (a GPCR) using two different software packages: AutoDock and Glide. The experimental crystal structure of OX2 bound to Suvorexant, was obtained from the Protein Data Bank (PDB ID: 4S0V). Potential binding sites were screened using both software packages, and used to generate a receptor grid. The docking results were validated by calculating the Root Mean Square Deviation (RMSD) of the crystal and predicted ligand conformation, as well as the free energy of binding. Both docking tools presented the binding of Suvorexant to the same binding pocket in the protein that was observed in the crystal structure, and had small RMSD values indicating a close match to the crystal conformation. AutoDock yielded a lower free energy score compared to Glide, and although both packages reproduced hydrogen bonding, electrophilic, and hydrophobicity interactions between the receptor and ligands, only Autodock was able to reproduce the expected pi-pi stacking interactions.

Title

Identification of cell cycle proteins in divergent eukaryotes using HITUHMM (Homolog Identification Tool Using Hidden Markov Modeling)

Authors

Daniel Carillo,[1,2] Jerry Duran,[1,3] Jon Mohl,[4,5] Ming-Ying Leung,[1,3,4,5] Jennifer Apodaca[3]

[1]Undergraduate Participation in Bioinformatics Training Program (UPBiT), [2]Department of Mathematical Sciences, [3]Department of Biological Sciences, [4]Border Biomedical Research Center, [5]Bioinformatics Program, UTEP, El Paso, TX

Abstract

The purpose of this project is to develop a bioinformatics pipeline to identify proteins important for cell division across distantly related phyla. To date, much of the existing knowledge about mitosis is derived from model organisms found within the animal and fungal kingdoms such as yeast (*S. cerevisiae*), fruit fly (*D. melanogaster*), worm (*C. elegans*), and mouse (*M. musculus*). We are developing a homolog identification tool that primarily uses hidden Markov (HITUMM) modeling to identify genes associated with well characterized features that are associated with the nuclear division of cells. This work focuses on identifying cell division proteins in distantly related microbial eukaryotes, specifically in parasitic protists such as *Giardia lamblia, Trypanaosoma bruciae* and *Plasmoidium falciparum.* Our bioinformatics pipeline combines BLAST, orthoMCL, MUSCLE, HMMER and network analysis to create a unique convenient search tool capable of identifying homologs in distantly related organisms.

Title

Assessing the prediction accuracy of the UTEP GPCR prediction pipeline

Authors

Daniel Castaneda-Mogollon,[1] Jon Mohl,[1,2,3] Sergio Munoz,[1] and Ming-Ying Leung,[1,2,3,4]
[1]Bioinformatics Program, [2]Computational Science Program, [3]Border Biomedical Research Center, and [4]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

G-Protein Coupled Receptors (GPCRs) are a large family of protein receptors. They are the targets of many pharmaceutical drug development research studies because of their vital importance in cell signal transduction. These receptors are mostly found in eukaryotes, and are also known as seven transmembrane domain receptors. GPCRs are responsible for sending signals inside a cell (triggering a biochemical cascade reaction) after a ligand has docked to it. So far, the size of the GPCR protein family remains unknown, but it is estimated that around 800 genes are responsible for coding them from the human genome. Various bioinformatics software exist for predicting whether a protein is a GPCR or not based on its amino acid sequence. Since there are many other types of proteins located in a cell's membrane, it is important to identify those belonging to the GPCR protein family. The purpose of my study is to assess the prediction accuracy, positive predicted value (PPV) and sensitivity of the test by testing three GPCR prediction approaches (Pfam, TMHMM, GPCRpred) used in our GPCR prediction pipeline, which is currently under development at UTEP. Two protein datasets, collected from public protein databases, were analyzed: experimentally proven GPCR (2887 sequences) and non-GPCR (1614 sequences). When the three prediction tools were used together, they correctly identified 99.82% of the positive set as GPCRs and 90.70% as non-GPCRs in the negative set, giving an overall prediction accuracy of 96.68%.

Title

Electronic and structural properties of C_{60} and $Sc_3N@C_{80}$ graphene supported fullerenes

Authors

Nakul N. Karle,[1] Tunna Baruah,[1,2] Rajendra R. Zope,[1,2] and Jose U. Reveles[3]
[1]Computational Science Program and [2]Department of Physics, The University of Texas at El Paso, El Paso, TX, and [3]Department of Physics, Virginia Commonwealth University, Richmond, VA

Abstract

A theoretical study on the geometric and electronic structure of C_{60} and $Sc_3N@C_{80}$ absorbed on pristine single layer graphene (SLG) is presented. C_{60} is found to adsorb in two nearly degenerate configurations: with a pentagon facing the SLG which is the most stable one, and with a hexagon facing the SLG in a face-to-face perfect alignment, rarely common in π-π interactions, 0.06 eV higher in energy. The calculated binding energy of 0.76 eV, which includes dispersion effects, is in good agreement with previous theoretical and experimental reports. On the other hand, $Sc_3N@C_{80}$ adsorption on the SLG resulted in a higher binding energy of 1.00 eV for nearly degenerate isomers that have a pentagon and a hexagon facing the SLG. This larger binding energy is explained in terms of a higher dispersion interaction between the larger metallofullerene and the SLG and due to the fact that charge separation in $Sc_3N@C_{80}$, which results in a positively charged Sc_3N inside a negatively charged C_{80}, favors binding with the SLG. Further, the Sc_3N moiety is found to rotate inside the supported C_{80} fullerene which in combination with the orientation of the fullerene on the SLG leads to a series of isomers with binding energies ranging from 0.76 to 1.00 eV. Our results show that it could be possible to adsorb metallofulleres on graphene with an energy large enough to prevent diffusion and therefore opening the possibility to potential applications.

Title

The effects of population stratification on the Quantitative Transmission Disequilibrium Test

Authors

Kyle A. Long,[1] Anthony M. Musolf,[2] and Joan E. Bailey-Wilson[2]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX and [2]National Human Genome Research Institute, National Institutes of Health, Baltimore, MD

Abstract

The Transmission Disequilibrium Test, TDT, is a family-based association test that looks for genetic correlation between some genetic variant and a trait, often times a disease. This test performs well when examining two or more combined populations when analyzing binary traits. However, it begins to break down when looking at quantitative traits that exhibit different distributions across populations (i.e., it may have different population means, different ranges, etc.). Due to differing allele frequencies of genetic variants across populations, which causes population stratification, combining the populations may result in the inflation of the false positive rate, skewing the relationship between genotype and phenotype. The study of quantitative traits requires a correction to be applied in order combat the high number of false positives. Here, we use PLINK QFAM, by Shaun Purcell of Harvard, and QTDT, by Goncalo Abecasis of the University of Michigan, two programs that employ the Transmission Disequilibrium Test, to study how the TDT algorithm handles quantitative traits. In order to test this, we used populations with both a small difference in population means and a large difference in the means. We found that PLINK QFAM requires a correction in both instances. QTDT, on the other hand, seems to have a built in correction. In the case of the small difference, it applies the correction appropriately. In the case of the large difference, though, QTDT's algorithm seems to overcorrect. This test reveals a difference in the underlying algorithm of the two programs, leading to future work in understanding their approach.

Title

Constructing an iterative method in predicting O-glycosylation sites of a protein with different ppGalNAc-transferases working in concert

Authors

Jon Mohl,[1,2,3] Thomas A Gerken,[4] and Ming-Ying Leung[1,2,3,5]

[1]Border Biomedical Research Center, [2]Bioinformatics Program, and [3]Computational Science Program, The University of Texas at El Paso, El Paso, TX, [4]Departments of Pediatrics, Biochemistry, and Chemistry, Case Western Reserve University, Cleveland, OH, and [5]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

The ISOGlyP program predicts whether a Threonine or Serine residue will be glycosylated by human ppGalNAc-transferases. Neighboring glycosylation can affect the likelihood of a site being glycosylated by certain transferases. To account for this, a version of the ISOGlyP program is needed to look at a stepwise prediction that will allow for adjacent predicted glycosylation to influence the sites that had lower scores in prior steps. In short, for each round of prediction, the program modifies the sequence to contain a glycosylated residue at the position with the highest EVP score from the prior prediction. This continues until no remaining Threonine or Serine residues have predicted EVP values above a stated cutoff value. The algorithm was designed to minimize computing resources by only modifying regions in which the glycosylation predicted to have occurred. The iterative ISOGlyP method was implemented in Python and was parallelized so that large number of sequences could be processed in a single submission. To test the different methods, results from Steentoft, et al. (2013) was compiled because both the specific ppGalNAc-transferases present in the cells and the resulting O-GalNAc glycoproteome were experimentally confirmed for the SimpleCell HepG2 cell line. Using the Steentoft's results, the accuracy, sensitivity and specificity were determined by randomly sampling confirmed positive and negative positions. In the analysis, overall accuracy increased from 50.2% to 64.9% in the iterative method. This increase was due to reducing the number of false positives, but some of the true positives were not identified as a tradeoff.

Title

Identifying gene expression pathways associated with Squamous Lung Cancer

Author

Roshani Rajapaksha, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

Squamous lung cancer (SQLC) is a common type of lung cancer and a leading cause of cancer related death. Due to the lack of the knowledge of its oncogenesis mechanisms, there are so few drugs that are both active and tolerable in SQLC patients. Therefore, the research and treatment of SQLC is of great importance. The objective of this study was to identify deregulated gene pathways in SQLC, which will provide new insights into newer and more specific therapeutic intervention.

Gene expression data was downloaded from the GEO database and analyzed using the Tmev, microarray data analyzing software to identify differentially expressed genes in cancer and normal cells. We identified 28 genes as significantly different in gene expression in cancer cells. Among them 9 genes were up-regulated in the cancer cells and 19 genes were down-regulated. Pathway enrichment analysis was conducted using DAVID Functional Annotation Tool to identify the deregulated biological pathways of these differentially expressed genes. DNA replication, cell cycle, metabolism of xenobiotics, and p53 signaling pathways were identified as significantly altered in cancer cells. MCM2 gene may induce lung cancer via the DNA replication pathway whereas CCNB1 and CDC6 genes will alter the p53 signaling and cell cycle pathways respectively. Thus MCM2, CCNB1, CDC6 genes have key roles in the progression and development of SQLC and may potentially be used as specific therapeutic targets.

Title

pfsearch: A potential tool for the prediction of the GPCR protein

Authors

Dewan Shrestha,[1] Jon Mohl,[1,2] Sergio Munoz,[1] and Ming-Ying Leung[1,2,3]
[1]Bioinformatics Program, [2]Computational Science Department, and [3]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

G-protein-coupled receptors (GPCRS) are the largest and most diverse group of membrane receptors in eukaryotes. GPCR are the seven transmembrane protein structure with an external N terminus and internal C-terminus. Since, GPCR are one of the most important receptor protein for signal transduction cascade and about fifty percent of the pharmacological drugs target these protein, prediction of this protein has been a prime interest.

For this study, we used a computational tool called pfsearch, provided by Swissprot and developed by Philipp Bucher, which uses the GPCR protein profiles obtained from PROSITEto predict GPCR proteins. Each of the GPCR protein profiles have their own cutoff score, so the pfsearch compares the profile against the protein sequence and generates the score based on alignment. If the score is within the cut off value then it is predicted as the specific GPCR protein family based on the profile we used. For testing of the tool, we used the positive data set of 2887 GPCR sequences that were experimentally verified, out of which only 64 of the sequences were not predicted as GPCR. For the non-GPCR data set (1614 sequences), all of them were predicted as non-GPCR sequences. Based on the statistical analysis, the tool had sensitivity of 0.977 and accuracy of 0.98 and positive predictive value of 100%. So, this shows that pfsearch can be one of the potential tool for predicting the GPCR protein.

Title

Differential gene expression analysis of the transcriptomic alterations in tumor cell lines following V-ATPase inhibition

Authors

Regina Umarova,[1,2] Ben Wheaton,[3] Sashi Palaniswamy,[4] Sergio Cordova,[5] and Anitha Sundararajan[6]

[1]Department of Biology, New Mexico Highlands University, Las Vegas, NM, [2]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, [3]Department of Biology and Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, NM, [4]School of Computer Science Program, The University of Manchester, Mancheser, UK, [5]Department of Computer Science, University of New Mexico, Albuquerque, NM, and [6]National Center for Genome Resources, Santa Fe, NM

Abstract

The goal of this study was to evaluate alterations in the transcriptome in several human tumor cell lines (MDA MB231, T47D, MDA MB453, LNCaP, PC3 and LaPC4) following vacuolar ATPase (V-ATPase) inhibition. V-ATPase is a multi-subunit proton pump that maintains the pH gradient across the endomembrane system; its inhibition alters pH homeostasis causing far-reaching downstream consequences. V-ATPase is upregulated in tumors and tumor cell lines. Tumor cell lines were treated with 10nM Concanamycin A (a highly specific and potent V-ATPase inhibitor), sequenced and then aligned against GRCh38 (new human genome assembly). The data analysis was performed using DESeq tool revealing 2% differentially expressed genes.

In the previous research it was shown that pharmacologic inhibition of V-ATPase alters the expression of several genes including Hypoxia Inducible Factor 1α (HIF1α) and the estrogen and androgen receptors (ER and AR, respectively). HIF1 α, ER and AR all regulate transcription. To identify the biological processes in which V-ATPase may be involved in expression regulation, the pathway analysis was performed using the differentially expresses genes. The PANTHER pathway analysis revealed 330 pathways, where V-ATPase was part of the PTHR31792 (Vacuole ATPase Assembly Integral Membrane Protein VMA21) Subfamily.

Title

Identification of hepatitis C virus-host interactions using functional genomics and bioinformatics approaches

Authors

Kristin Valdez,[1,2] Brianna Lowey,[1] Qisheng Li,[1] Fathi Elloumi,[3] Maggie Cam,[3] and T. Jake Liang[1]
[1]Liver Diseases Branch, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, [3]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and [2]Collaborative Bioinformatics Resource, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD

Abstract

Hepatitis C virus is a positive strand RNA virus belonging to the *Flaviviridae* family, and is now the leading cause of liver cirrhosis and liver cancer. Hepatitis C virus depends heavily on host factors for efficient replication and infection. There are several stages of hepatitis C virus infection: entry, translation, replication, assembly and secretion. The assembly stage of infection requires effective lipid droplet biosynthesis, and previous research has indicated that N-Myc Downstream Regulated Gene 1 (NDRG1) is an antiviral host factor that decreases hepatitis C virus infection at this stage. In addition, studies have shown that NDRG1 plays a regulatory role in cells, and is downregulated by the hepatitis C virus during infection, helping the virus to circumvent NDRG1's antiviral effects. However, underlying mechanisms including specific gene interactions and related pathways have yet to be uncovered. We present an RNA interference study using hepatocytes, or liver cells, to uncover mechanistic details regarding NDRG1. Using small interfering RNAs, we prepared triplicates of NDRG1 knock down samples along with triplicates of control cells. Messenger RNA was extracted from the cells and quantified with a microarray scan. Subsequently, we analyzed the results using a novel implementation of a microarray pipeline and Ingenuity Pathway Analysis software. Through these tools, we were able to elucidate pathways specific to NDRG1's role in lipid biosynthesis and uncover key genes involved in those pathways. These findings may provide not only insights into hepatitis C virus-related disease mechanisms but also valuable therapeutic targets.

Title

Identifying genetic sequence variations from exome datasets

Authors

Mariana Vasquez,[1,2] Alejandro Diaz,[3] Jon Mohl,[1,4,5] and Ming-Ying Leung[1,4,5,6]
[1]Bioinformatics Program, [2]Department of Biological Sciences, [3]Department of Computer Science, [4]Border Biomedical Research Center, [5]Computational Science Program, [6]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

As whole exome sequencing is becoming a more accessible way to obtain an abundance of biological information, scientists can use it to study mutations in the genome that can lead to various types of cancers and other diseases. Genetic sequence variations (GSVs) are mutations found within the genome and can vary among different populations. GSVs can include small insertions or deletions (InDels) and single nucleotide polymorphisms (SNPs). These variations can occur either within the transcript, including exons, introns or untranslated regions, or upstream and downstream of the transcript. GSVs in exomes can be obtained through next generation sequencing (NGS), which is a high-throughput DNA sequencing procedure that can generate millions of DNA sequences per sample per run. To identify the GSVs, the sequences obtained from NGS need to be aligned to a human reference genome. The Bowtie2 software, an efficient aligner, was run to identify the locations of GSVs. A script was created to parse and format Bowtie2 output for OncoMiner. It was now possible to map the locations of the GSVs to specific types of mutations within genes, such as SNPs or InDels, which occur in the transcript and surrounding regions. A table containing the necessary information to perform downstream statistical analyses by the OncoMiner pipeline was created. The script for parsing Bowtie2 output and isolating the GSV information will be modified to run in parallel on either the Blue Waters Supercomputer at the University of Illinois at Urbana-Champaign or the High-Performance Cluster at UTEP.

Title

Review Presentation: "Genomic characterization and phylogenetic analysis of Zika virus circulating in the Americas"

Presenter

Bofei Wang, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

This poster is a review of the paper "Genomic characterization and phylogenetic analysis of Zika virus circulating in the Americas" by Qing Ye *et al*. (2016) published in the journal *Infection, Genetics and Evolution.* The abstract, as originally published by the authors, is as follows:

“Motivation: The rapid spread and potential link with birth defects have made Zika virus (ZIKV) a global public health problem. The virus was discovered 70 years ago, yet the knowledge about its genomic structure and the genetic variations associated with current ZIKV explosive epidemics remains not fully understood.

Results: In this review, the genome organization, especially conserved terminal structures of ZIKV genome were characterized and compared with other mosquito-borne flaviviruses. It is suggested that major viral proteins of ZIKV share high structural and functional similarity with other known flaviviruses as shown by sequence comparison and prediction of functional motifs in viral proteins. Phylogenetic analysis demonstrated that all ZIKV strains circulating in the America form a unique clade within the Asian lineage. Furthermore, we identified a series of conserved amino acid residues that differentiate the Asian strains including the current circulating American strains from the ancient African strains. Overall, our findings provide an overview of ZIKV genome characterization and evolutionary dynamics in the Americas and point out critical clues for future virological and epidemiological studies.”