

Title

Mathematical modelling and sensitivity analysis of the transmission dynamics of lassa fever, with a view of selected Nigerian states

Authors

Tolulope Samuel Adeyina^{1,2} and Oyamakin S. Oluwafemi²

¹Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and ²Department of Statistics, University of Ibadan, Ibadan, Nigeria

Abstract

Lassa fever accounts for an estimated 200,000 to 500,000 cases and 5000 deaths yearly in West Africa. Understanding the dynamics of the disease and formulating a model for the movement from one compartment to the other will help provide an effective and lasting solution for the eradication of the disease. A non-linear deterministic model was considered to study the dynamics transmission and control of Lassa fever virus. The total population was divided into seven mutually exclusive classes between human and rodents. Existence and uniqueness of the solution of the model were determined, the basic reproduction number is derived, and the model threshold parameter was examined using next-generation operator method. The model result shows that diseases free equilibrium is local asymptotically stable at $R_0 < 1$ and unstable at $R_0 > 1$, the model is globally asymptotically stable. Sensitivity analysis of the model parameters was carried out; the most sensitive parameter are $\eta(\mathbf{a}_1)$ and ϵ_h which are effective transmission rate in susceptible humans by infected humans and progression rate of humans from the exposed state to the infectious state respectively. Data was collected from NCDC, and the value of R_0 was estimated for different states in Nigeria with the occurrence at three levels illustrating varying intervention levels. At minimum intervention (0.1), Edo state and Ondo state have their $R_0 = 1.515655$ and 1.065444 respectively which implies instability ($R_0 > 1$). Further investigation shows that timely intervention and proximity to diagnostic centers will help reduce the basic reproduction number of Lassa fever in the selected Nigerian states.

Title

Computer vision-based tools to segment gray and white matter regions in experimental tissue sections and to analyze tracer injection sites mapped in digital atlas space: Use cases for the hypothalamus and ventral tegmental area for circuits related to feeding control

Authors

Alexandro Arnal,¹ Olac Fuentes,² and Arshad M. Khan³

¹Computational Science Program, ²Department of Computer Science, and ³Department of Biological Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

We have scripted two computer vision-based programs that complete two tasks: (1) process and analyze raw images of experimental tissue sections stained for Nissl substance and segment gray and white matter regions from them; and (2) calculate the percent overlap of a mapped injection deposit area in relation to underlying brain regions within a standardized rat brain atlas (LW Swanson, 2018; Brain Maps 4.0, J Comp Neurol). First, we show that automating Task 1 can be achieved with Fully Convolutional Networks (FCN), a family of deep learning models trained end-to-end for pixel-wise classification of images. The proposed FCN was trained on a dataset augmented from nine Nissl-stained coronal images of a rat and parcellated using the Swanson atlas as a guide. For simplicity, the model is trained to segment the fornix, a distinct structure found in all levels of the available parcellations. We experiment with different resolutions, number of augmented samples, augmentation methods and network architectures. The Task 1 script can classify pixels based on texture with higher density classification ascribed to regions in and around the fornix. For Task 2, we illustrate the significance of parcellation or semantic segmentation with a quantitative analysis of brain regions and overlapping artifacts. Often, rigorous benchmark standards to document ground truth - such as atlas-based mapping of a tracer injection site in relation to the underlying cytoarchitecture - do not take into account the overlap of tracer injection deposits with the boundaries of the underlying brain regions. As a result, the neural connections defined by the deposit are usually inferred as being traced from a single assigned brain region to the exclusion of other regions within the tracer deposit footprint. Task 2 workflow involves the creation of two annotated images, one representing a given level or portion thereof of the Swanson atlas, and the other representing any given artifact superimposed on the atlas map. The tool allows areal calculations to be made of the fraction of deposit within a target site and the fraction that spills over to other sites. We demonstrate the feasibility of our script using injection site deposits mapped for tracer deposits delivered into the ventral tegmental area. Collectively, these tools provide bench scientists interested in delineating cytoarchitecture from their experimental tissue sets - and to map their tracer deposits over such cytoarchitecture onto atlas maps - better ways to rigorously process and analyze their data.

Title

Compiling structural data for predicting functional roles of G-protein coupled receptors

Authors

Owusu Kwabena Dankwah,¹ Khodeza Begum,¹ and Ming Ying Leung^{1,2}

¹Computational Science Program, ²Border Biomedical Research Center, ²Bioinformatics Program and ²Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

G-protein coupled receptors (GPCRs) are the largest class of cell-surface receptors and are encoded by more than 800 genes in the human genome. They are a major target for the pharmaceutical industry as they mediate therapeutic effects of ~34% of the marketed drugs. The rapid growth of available three-dimensional (3D) structural data of the biologically and pharmaceutically important G-protein coupled receptors has revealed useful information for studying GPCR-ligand binding modes, and G-protein binding mechanisms. These structures shed light on structural similarity and diversity of the GPCR ligand recognition, GPCR functional states, and characteristics of a receptor structure that is competent for G-protein binding. Studies have shown that structural pattern defined by transmembrane intramolecular interactions are suitable for comparison of GPCR 3D structures and unsupervised distinction of the receptor states. 3D structural information has been used to predict coupling probabilities for GPCRs to individual G-proteins. We have constructed a searchable MySQL database, GPCR-PEnDB (GPCR Prediction Ensemble Database; gpcr.utep.edu/database), of confirmed GPCRs and non-GPCRs with plans to link them to their Protein Data Bank entries. GPCR-PEnDB contains 3129 confirmed GPCR, 3575 non-GPCR sequences from over 1200 species. To date, the Protein Data Bank contains 546 3D structures related to 105 unique GPCRs. We will integrate these data with protein sequence features to improve computational algorithms for predicting the functional roles of GPCRs.

Title

Analysis on the fairness of large class B medical equipment in China

Authors

Wenhan Guo^{1,2} and Qingfeng Tian²

¹Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and ²Department of Social Medicine and Health Service Management, College of Public Health, Zhengzhou University, Zhengzhou, China

Abstract

In China, large-scale medical equipment classified as class A is approved by the Ministry of Health of the State, while large class B medical equipment is approved at the provincial level. The present study is to understand the fairness about the distribution of large class B medical equipment in several regions and provinces, and to provide a basis for the Departments of Health and Family Planning to develop allocation plan of large-scale medical equipment. An investigation was taken on the current status on large class B medical equipment. Gini coefficient was used to analyze and evaluate the fairness and scientificity about the allocation of large class B medical equipment from two angles of population and geography. Theil index was used to measure cross-regional fairness and analyze the relevant factors. The Gini coefficient of large class B medical equipment was below 0.3 and the allocation efficiency was quite well. In accordance with the geographical distribution of the Gini coefficient, the configuration of CT, MRI and DSA was more than 0.6 and was relatively high. The Theil index of CT, MRI, DSA were 0.0153, 0.0171 and 0.0305, respectively, which indicated good fairness in general. Allocation of large class B medical equipment has a better fairness when it is distributed according to population than according to geographical. The fairness of the allocation of large class B medical equipment in different provinces is influenced by the local area. It is suggested to adopt the macro control policy and scientific strategy to allocate large class B medical equipment.

Title

In vitro cytotoxicity of imidacloprid and thiamethoxam formulations (Actara®25 WG and Iron®70 WG)

Authors

Raisa Nkweteyim,^{1,2} Faustin Pascal T. Manfo,² and Njotu Fabrice²

¹Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and ²Department of Biochemistry and Molecular Biology, University of Buea, Buea, Cameroon

Abstract

Pesticides are widely used in Cameroon to improve yield of crops especially as a good proportion of the labor force by occupation are in agriculture as their primary source of livelihood. Thus, it is important to investigate the effects these commonly used pesticides have on human health. Since imidacloprid is the most widely used pesticide in the world, this research was focused on the cytotoxic effects of imidacloprid and thiamethoxam formulations on human red blood cells *in vitro*.

To carry out this research, blood was collected from healthy blood donors, red blood cells isolated, and then incubated with various concentrations of the pesticide formulations (Iron® 70 WG and Actara® 25 WG) for 30 minutes. Following lysis of the red blood cells, hemoglobin elevation was quantified using spectrophotometric techniques.

Exploratory data analysis was carried out using Microsoft Excel 2013 software while the dose-response curve and cytotoxic concentration 50 (CC50) were computed using Sigma Plot Software. Results showed that Iron® 70 WG is significantly cytotoxic with a low cytotoxic concentration 50 (CC50) of 949.8µM, but there was no significant cytotoxicity for Actara® 25 WG. These results project that continued use of Iron 70 WG will degrade human health, thus it is advised to avoid exposure to this chemical.

Title

Identifying novel DNA variants potentially associated with human hydrocephalus

Authors

Raymond Otoo,¹ Jun Zhang,³ and Jon Mohl^{1,2}

¹Bioinformatics Program, ²Border Biomedical Research Center, ²Computational Science Program, and ²Department of Mathematical Sciences, The University of Texas at El Paso, and ³Paul L. Foster School of Medicine, Texas Tech Health Sciences Center at El Paso, El Paso, TX

Abstract

Human hydrocephalus is a common medical condition that is characterized by abnormalities in the flow or resorption of cerebrospinal fluid (CSF), resulting in ventricular dilatation in the brain. To date genetic etiologies of this disease is elusive. Dr. Zhang's lab is determined to understand the genetic etiology of ETiNPH. DNA samples were collected from eight unrelated patients with four manifesting the disease and four healthy patients. The samples were sequenced using Illumina 2000 and the data was assembled in .xlsx format. A python script together with a shell script and R -script were used to identify the common SNPs among the diseased patients. The analysis of the data has been started with Chromosome 19 as it has the highest gene density of all human chromosomes, more than double the genome-wide average. The chromosome positions of the common SNPs were extracted and compared against the NCBI Single Nucleotide Polymorphism database (dbSNP). 31 SNPs were found in all the diseased but not the healthy individuals and 30 of these were found not reported in dbSNP. For the SNPs that were found in 3 diseased individuals against no healthy individual or found in all the diseased and only one healthy individual, 345 were found not to be reported. From this results, further investigations can be performed on the unreported chromosome locations for this disease to develop drugs to manage or treat the disease. In future studies, the study can be extended to other chromosomes as they may contain certain variations as well.

Title

On the performance of variable selection and classification via rank-based classifier

Authors

Md Showaib Rahman Sarker,¹ Michael Pokojovy,^{1,2} and Sangjin Kim^{1,2}

¹Computational Science Program and ²Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

In high-dimensional gene expression data analysis, the accuracy and reliability of cancer classification and selection of important genes play a very crucial role. To identify these important genes and predict future outcomes (tumor vs. non-tumor), various methods have been proposed in the literature. But only few of them take into account correlation patterns and grouping effects among the genes. In this article, we propose a rank-based modification of the popular penalized logistic regression procedure based on a combination of ℓ_1 and ℓ_2 penalties capable of handling possible correlation among genes in different groups. While the ℓ_1 penalty maintains sparsity, the ℓ_2 penalty induces smoothness based on the information from the Laplacian matrix, which represents the correlation pattern among genes. We combined logistic regression with the BH-FDR (Benjamini and Hochberg false discovery rate) screening procedure and a newly developed rank-based selection method to come up with an optimal model retaining the important genes. Through simulation studies and real-world application to high-dimensional colon cancer gene expression data, we demonstrated that the proposed rank-based method outperforms such currently popular methods as lasso, adaptive lasso and elastic net when applied both to gene selection and classification.

Title

A computational method for predicting functional effects of cancer-related genetic sequence variants

Authors

Bofei Wang,¹ Jonathon Mohl,^{1,2} and Ming Ying Leung^{1,2}

¹Computational Science Program, ²Border Biomedical Research Center, ²Bioinformatics Program and ²Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

Rapid advances in next generation sequencing (NGS) technologies provide many opportunities to identify associations between genetic sequence variants (GSV) and diseases, which may lead to better clinical diagnosis and treatments. OncoMiner is a bioinformatics pipeline developed at UTEP (OncoMiner.utep.edu) for mining NGS data and predicting their associations with cancers. Since the current input for OncoMiner is limited to a specific file format set by the Otogenetics sequencing company, we have developed a Python script to handle the more common variant calling format (VCF) NGS files and convert them to the OncoMiner input (OMI) format. To improve efficiency, the script splits the VCF file by chromosomes into smaller files for parallel processing. The script was tested on 148 VCF files containing data from prostate cancer patients downloaded from The Cancer Genome Atlas. We used 1, 2, 4, 8, 16 and 24 cores. The highest speedup achieved was 4.0001 using 24 cores. Also, a 7 GB VCF file containing 29 samples from local patients, normal people and cell lines was split into 35 subsets of various sizes to test program's capability of handling big datasets and analyze runtime behavior. The runtimes of 35 files using 8 cores range from 35.7 mins to 39.7 hours. Then, we incorporated the script into OncoMiner using web.py. For future work, we are going to compare the results of OncoMiner against those from other common prediction tools. By assessing these tools, we are trying to create an ensemble approach to achieve higher accuracy.

Self-similar models: How close the Diffusion Entropy Analysis and Detrended Fluctuation Analysis are from other models

William Kubin,¹ Peter K. Asante,¹ Maria C. Mariani,^{1,2} and Osei K. Tweneboah¹

¹Computational Science Program and
²Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

Financial and seismic data, like many other high frequency data are known to exhibit memory effects. Current trend is to apply the concepts of the Lévy process, Hurst exponent, Diffusion Entropy Analysis (DEA) and the Detrended Fluctuation Analysis (DFA) to find out the memory behavior (whether short or long) that the time series data exhibits. It has been widely used in literature that the Lévy process is well known to describe long memory processes. The Lévy process (where the increments are independent and follow the Lévy distribution) is self similar. This is described by the relation

$$\alpha = \frac{1}{H}.$$

Our goal in this work is to examine how close the DEA and DFA models are to a self-similar model. We do this by investigating the relationships between α characterizing the DFA and the resulting H parameter characterizing the self-similar property as well as the δ characterizing the DEA and the resulting H parameter characterizing the self-similar property.

We analyzed time series data from several financial markets and recordings of volcanic eruptions from a seismic station.

Title

Review Presentation: "Systems-wide analysis unravels the new roles of CCM signal complex (CSC)"

Presenter

Dristi Adhikari, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

This poster is a review of the paper "Systems-wide analysis unravels the new roles of CCM signal complex (CSC)" by Abou-Fadel *et al.* (2019) published in the preprint server "bioRxiv"; doi.org/10.1101/631424. The abstract originally published by the authors is as follows:

"Cerebral cavernous malformations (CCMs) are characterized by abnormally dilated intracranial capillaries that result in increased susceptibility to stroke. Three genes have been identified as causes of CCMs; KRIT1 (CCM1), MGC4607 (CCM2) and PDCD10 (CCM3); one of them is disrupted in most CCM cases. It was demonstrated that both CCM1 and CCM3 bind to CCM2 to form a CCM signaling complex (CSC) to modulate angiogenesis. In this report, we deployed both RNA-seq and proteomic analysis of perturbed CSC after depletion of one of three CCM genes to generate interactomes for system-wide studies. Our results demonstrated a unique portrait detailing alterations in angiogenesis and vascular integrity. Interestingly, only indirect overlapped alterations between RNA and protein levels were detected, supporting the existence of multiple layers of regulation in CSC cascades. Notably, this is the first report identifying that both β 4 integrin and CAV1 signaling are downstream of CSC, conveying the angiogenic signaling. Our results provide a global view of signal transduction modulated by the CSC, identifies novel regulatory signaling networks and key cellular factors associated with CSC."

Title

Review Presentation: “Application of an NGS-based 28-gene panel in myeloproliferative neoplasms reveals distinct mutation patterns in essential thrombocythaemia, primary myelofibrosis and polycythaemia vera”

Presenter

Kristen Arce, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

This poster is a review of the research paper “Application of an NGS-based 28-gene panel in myeloproliferative neoplasms reveals distinct mutation patterns in essential thrombocythaemia, primary myelofibrosis and polycythaemia vera” by Delic *et al.* (2016) in *British Journal of Haematology*, 175:419–426; doi.org/10.1111/bjh.14269. The abstract originally published by the authors is as follows:

“Molecular routine diagnostics for BCR-ABL1-negative myeloproliferative neoplasms (MPN) currently focusses on mutations in JAK2, CALR and MPL. In recent years, recurrent mutations in MPNs have been identified in several other genes. We here present the validation of a next generation sequencing (NGS)-based 28-gene panel and its use in MPN. We analysed the mutation status of 28 genes in 100 MPN patients [40 essential thrombocythaemia (ET), 30 primary myelofibrosis (PMF), 30 polycythaemia vera (PV)] and found two or more mutated genes in 53 patients. Moreover, significantly more mutated splicing genes (SF3B1, SRSF2 and U2AF1) were present in PMF (0.60 mutated genes/patient) compared to ET (0.15) while no mutations in splicing genes were found in PV. Additionally, chromatin modification genes (ASXL1 and EZH2) were frequently mutated in PMF patients (0.50) and, to a significantly lesser extent, in ET (0.13) and PV (0.07). Contrarily, DNA methylation genes (DNMT3A, IDH1, IDH2 and TET2) were mutated most often in PV (0.5) and less frequently in ET (0.23) and PMF (0.20), but without reaching statistical significance. Our results demonstrate the feasibility and utility of NGS-based panel diagnostics for MPN. With 53% of the patients bearing two or more mutated genes, their prognostic relevance needs further studies.”

Title

Review Presentation: “ATD: a comprehensive bioinformatics resource for deciphering the association of autophagy and diseases”

Presenter

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

Abstract

This poster is a review of the article “ATD: a comprehensive bioinformatics resource for deciphering the association of autophagy and diseases” by Wang *et al.* (2018) in *Database*, 2018: article ID bay093; doi.org/10.1093/database/bay093. The abstract in the original article is as follows:

“Autophagy is the natural, regulated, destructive mechanism of the eukaryotes cell that disassembles unnecessary or dysfunctional components. In recent years, the association between autophagy and diseases has attracted more and more attention, but our understanding of the molecular mechanism about the association in the system perspective is limited and ambiguous. Hence, we developed the comprehensive bioinformatics resource Autophagy To Disease (ATD, <http://auto2disease.nwsuaflmz.com>) to archive autophagy-associated diseases. This resource provides bioinformatics annotation system about genes and chemicals about autophagy and human diseases by extracting results from previous studies with text mining technology. Based on the big data from ATD, we found that some classes of disease tend to be related with autophagy, including respiratory disease, cancer, urogenital disease and digestive system disease. We also found that some classes of autophagy-related diseases have a strong association among each other and constitute modules. Furthermore, we extracted the autophagy–disease-related genes (ADGs) from ATD and provided a novel algorithm Optimized Random Forest with Label model to predict potential ADGs. This bioinformatics annotation system about autophagy and human diseases may provide a basic resource for the further detection of the molecular mechanisms of autophagy pathway to disease.”