Année de publication : 2018

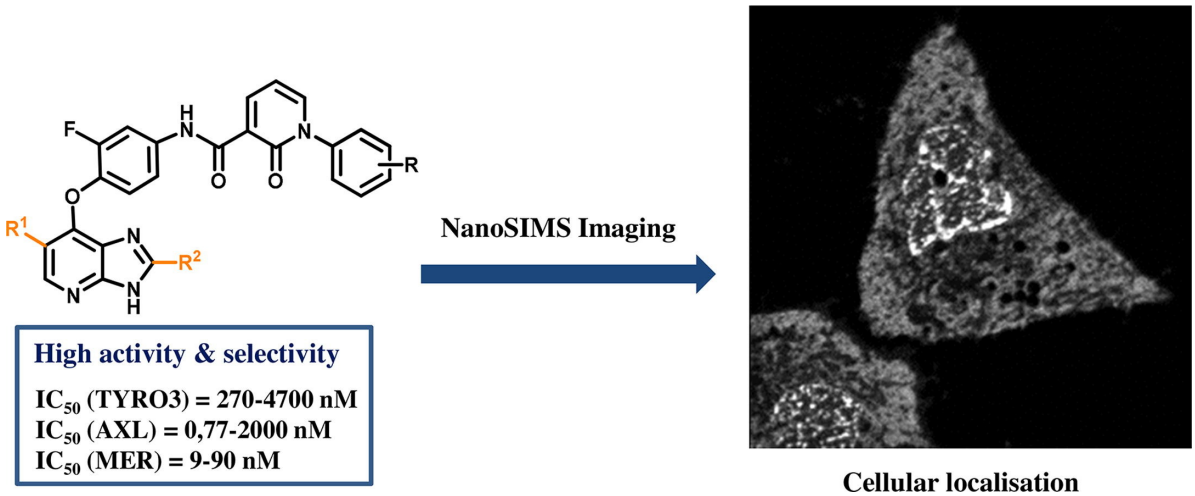
Tom Baladi, Jessy Aziz, Florent Dufour, Valentina Abet, Véronique Stoven, François Radvanyi, Florent Poyer, Ting-Di Wu, Jean-Luc Guerquin-Kern, Isabelle Bernard-Pierrot, Sergio Marco Garrido, Sandrine Piguel (2018 Nov 1)

Design, synthesis, biological evaluation and cellular imaging of imidazo[4,5-b]pyridine derivatives as potent and selective TAM inhibitors.

Bioorganic & medicinal chemistry : 26 : 5510-5530 : DOI : 10.1016/j.bmc.2018.09.031

Résumé

The TAM kinase family arises as a new effective and attractive therapeutic target for cancer therapy, autoimmune and viral diseases. A series of 2,6-disubstituted imidazo[4,5-b]pyridines were designed, synthesized and identified as highly potent TAM inhibitors. Despite remarkable structural similarities within the TAM family, compounds 28 and 25 demonstrated high activity and selectivity in vitro against AXL and MER, with IC value of 0.77 nM and 9 nM respectively and a 120- to 900-fold selectivity. We also observed an unexpected nuclear localization for compound 10Bb, thanks to nanoSIMS technology, which could be correlated to the absence of cytotoxicity on three different cancer cell lines being sensitive to TAM inhibition.



Benoit Playe, Chloé-Agathe Azencott, Véronique Stoven (2018 Oct 5)

Efficient multi-task chemogenomics for drug specificity prediction.

PloS one : e0204999 : DOI : 10.1371/journal.pone.0204999

Résumé

Adverse drug reactions, also called side effects, range from mild to fatal clinical events and significantly affect the quality of care. Among other causes, side effects occur when drugs bind to proteins other than their intended target. As experimentally testing drug specificity

against the entire proteome is out of reach, we investigate the application of chemogenomics approaches. We formulate the study of drug specificity as a problem of predicting interactions between drugs and proteins at the proteome scale. We build several benchmark datasets, and propose NN-MT, a multi-task Support Vector Machine (SVM) algorithm that is trained on a limited number of data points, in order to solve the computational issues or proteome-wide SVM for chemogenomics. We compare NN-MT to different state-of-the-art methods, and show that its prediction performances are similar or better, at an efficient calculation cost. Compared to its competitors, the proposed method is particularly efficient to predict (protein, ligand) interactions in the difficult double-orphan case, i.e. when no interactions are previously known for the protein nor for the ligand. The NN-MT algorithm appears to be a good default method providing state-of-the-art or better performances, in a wide range of prediction scenario that are considered in the present study: proteome-wide prediction, protein family prediction, test (protein, ligand) pairs dissimilar to pairs in the train set, and orphan cases.

Nicolas Servant, Nelle Varoquaux, Edith Heard, Emmanuel Barillot, Jean-Philippe Vert (2018 Sep 8)

Effective normalization for copy number variation in Hi-C data.

BMC bioinformatics : 313 : DOI : 10.1186/s12859-018-2256-5

Résumé

Normalization is essential to ensure accurate analysis and proper interpretation of sequencing data, and chromosome conformation capture data such as Hi-C have particular challenges. Although several methods have been proposed, the most widely used type of normalization of Hi-C data usually casts estimation of unwanted effects as a matrix balancing problem, relying on the assumption that all genomic regions interact equally with each other.

C-A Azencott (2018 Aug 8)

Machine learning and genomics: precision medicine versus patient privacy.

Philosophical transactions. Series A, Mathematical, physical, and engineering sciences : DOI : 20170350

Résumé

Machine learning can have a major societal impact in computational biology applications. In particular, it plays a central role in the development of precision medicine, whereby treatment is tailored to the clinical or genetic features of the patient. However, these advances require collecting and sharing among researchers large amounts of genomic data, which generates much concern about privacy. Researchers, study participants and governing bodies should be aware of the ways in which the privacy of participants might be compromised, as well as of the large body of research on technical solutions to these issues. We review how breaches in patient privacy can occur, present recent developments in computational data protection and discuss how they can be combined with legal and ethical

perspectives to provide secure frameworks for genomic data sharing.This article is part of a discussion meeting issue 'The growing ubiquity of algorithms in society: implications, impacts and innovations'.

Kévin Vervier, Pierre Mahé, Jean-Philippe Vert (2018 Jul 22)
MetaVW: Large-Scale Machine Learning for Metagenomics Sequence Classification.
Methods in molecular biology (Clifton, N.J.) : 9-20 : DOI : 10.1007/978-1-4939-8561-6_2

Résumé

Metagenomics is the study of microbial community diversity, especially the uncultured microorganisms by shotgun sequencing environmental samples. As the sequencers throughput and the data volume increase, it becomes challenging to develop scalable bioinformatics tools that reconstruct microbiome structure by binning sequencing reads to reference genomes. Standard alignment-based methods, such as BWA-MEM, provide state-of-the-art performance, but we demonstrate in Vervier et al. (2016) that compositional approaches using nucleotides motifs have faster analysis time, for comparable accuracy. In this work, we describe how to use MetaVW, a scalable machine learning implementation for short sequencing reads binning, based on their k-mers profile. We provide a step-by-step guideline on how we trained the classification models and how it can easily generalize to user-defined reference genomes and specific applications. We also give additional details on what effect parameters in the algorithm have on performances.

Evelien M Bunnik, Kate B Cook, Nelle Varoquaux, Gayani Batugedara, Jacques Prudhomme, Anthony Cort, Lirong Shi, Chiara Andolina, Leila S Ross, Declan Brady, David A Fidock, Francois Nosten, Rita Tewari, Photini Sinnis, Ferhat Ay, Jean-Philippe Vert, William Stafford Noble, Karine G Le Roch (2018 May 17)
Changes in genome organization of parasite-specific gene families during the Plasmodium transmission stages.
Nature communications : 1910 : DOI : 10.1038/s41467-018-04295-5

Résumé

The development of malaria parasites throughout their various life cycle stages is coordinated by changes in gene expression. We previously showed that the three-dimensional organization of the Plasmodium falciparum genome is strongly associated with gene expression during its replication cycle inside red blood cells. Here, we analyze genome organization in the P. falciparum and P. vivax transmission stages. Major changes occur in the localization and interactions of genes involved in pathogenesis and immune evasion, host cell invasion, sexual differentiation, and master regulation of gene expression. Furthermore, we observe reorganization of subtelomeric heterochromatin around genes involved in host cell remodeling. Depletion of heterochromatin protein 1 (PfHP1) resulted in loss of interactions between virulence genes, confirming that PfHP1 is essential for

maintenance of the repressive center. Our results suggest that the three-dimensional genome structure of human malaria parasites is strongly connected with transcriptional activity of specific gene families throughout the life cycle.

Peiying Ruan, Morihiro Hayashida, Tatsuya Akutsu, Jean-Philippe Vert (2018 Mar 6)
Improving prediction of heterodimeric protein complexes using combination with pairwise kernel.
BMC bioinformatics : 39 : DOI : 10.1186/s12859-018-2017-5

Résumé

Since many proteins become functional only after they interact with their partner proteins and form protein complexes, it is essential to identify the sets of proteins that form complexes. Therefore, several computational methods have been proposed to predict complexes from the topology and structure of experimental protein-protein interaction (PPI) network. These methods work well to predict complexes involving at least three proteins, but generally fail at identifying complexes involving only two different proteins, called heterodimeric complexes or heterodimers. There is however an urgent need for efficient methods to predict heterodimers, since the majority of known protein complexes are precisely heterodimers.

Koen Van den Berge, Fanny Perraudeau, Charlotte Soneson, Michael I Love, Davide Risso, Jean-Philippe Vert, Mark D Robinson, Sandrine Dudoit, Lieven Clement (2018 Feb 27)
Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications.
Genome biology : 24 : DOI : 10.1186/s13059-018-1406-4

Résumé

Dropout events in single-cell RNA sequencing (scRNA-seq) cause many transcripts to go undetected and induce an excess of zero read counts, leading to power issues in differential expression (DE) analysis. This has triggered the development of bespoke scRNA-seq DE methods to cope with zero inflation. Recent evaluations, however, have shown that dedicated scRNA-seq tools provide no advantage compared to traditional bulk RNA-seq tools. We introduce a weighting strategy, based on a zero-inflated negative binomial model, that identifies excess zero counts and generates gene- and cell-specific weights to unlock bulk RNA-seq DE pipelines for zero-inflated data, boosting performance for scRNA-seq.

Davide Risso, Fanny Perraudeau, Svetlana Gribkova, Sandrine Dudoit, Jean-Philippe Vert (2018 Jan 20)
A general and flexible method for signal extraction from single-cell RNA-seq

data.
Nature communications : 284 : DOI : 10.1038/s41467-017-02554-5

Résumé

Single-cell RNA-sequencing (scRNA-seq) is a powerful high-throughput technique that enables researchers to measure genome-wide transcription levels at the resolution of single cells. Because of the low amount of RNA present in a single cell, some genes may fail to be detected even though they are expressed; these genes are usually referred to as dropouts. Here, we present a general and flexible zero-inflated negative binomial model (ZINB-WaVE), which leads to low-dimensional representations of the data that account for zero inflation (dropouts), over-dispersion, and the count nature of the data. We demonstrate, with simulated and real data, that the model and its associated estimation procedure are able to give a more stable and accurate low-dimensional representation of the data than principal component analysis (PCA) and zero-inflated factor analysis (ZIFA), without the need for a preliminary normalization step.

Jiao Y., Vert J.P. (2018 Jan 1)
The Weighted Kendall and High-order Kernels for Permutations
Proceedings of the 35th International Conference on Machine LearningProceedings of the 35th International Conference on Machine Learning : 80 : 2314-2322

Résumé

Boyd J., Pinhiero A., Nery E.D., Reyal F., Walter T. (2018 Jan 1)
Analysing double-strand breaks in cultured cells for drug screening applications by causal inference
IEEE International Symposium on Biomedical ImagingIEEE International Symposium on Biomedical Imaging

Résumé

Pauwels E., Bach F., Vert J.P. (2018 Jan 1)
Relating Leverage Scores and Density using Regularized Christoffel Functions
Neural Information Processing SystemsNeural Information Processing Systems

Résumé

Le Morvan M., Vert J.P. (2018 Jan 1)
WHInter: A Working set algorithm for High-dimensional sparse second order

Interaction models.
Proceedings of the 35th International Conference on Machine LearningProceedings of the 35th International Conference on Machine Learning : 80 : 3635-3644

Résumé

Année de publication : 2017

Chloé-Agathe Azencott, Tero Aittokallio, Sushmita Roy, , Thea Norman, Stephen Friend, Gustavo Stolovitzky, Anna Goldenberg (2017 Sep 30)
The inconvenience of data of convenience: computational research beyond post-mortem analyses.
Nature methods : 937-938 : DOI : 10.1038/nmeth.4457

Résumé

Elsa Bernard, Yunlong Jiao, Erwan Scornet, Veronique Stoven, Thomas Walter, Jean-Philippe Vert (2017 Sep 27)
Kernel Multitask Regression for Toxicogenetics.
Molecular informatics : DOI : 10.1002/minf.201700053

Résumé

The development of high-throughput in vitro assays to study quantitatively the toxicity of chemical compounds on genetically characterized human-derived cell lines paves the way to predictive toxicogenetics, where one would be able to predict the toxicity of any particular compound on any particular individual. In this paper we present a machine learning-based approach for that purpose, kernel multitask regression (KMR), which combines chemical characterizations of molecular compounds with genetic and transcriptomic characterizations of cell lines to predict the toxicity of a given compound on a given cell line. We demonstrate the relevance of the method on the recent DREAM8 Toxicogenetics challenge, where it ranked among the best state-of-the-art models, and discuss the importance of choosing good descriptors for cell lines and chemicals.