

Année de publication : 2017

---

Héctor Climente-González, Eduard Porta-Pardo, Adam Godzik, Eduardo Eyras (2017 Aug 31)

**The Functional Impact of Alternative Splicing in Cancer.**

*Cell reports* : 2215-2226 : [DOI : S2211-1247\(17\)31104-X](https://doi.org/10.1016/j.celrep.2017.08.014)

**Résumé**

Alternative splicing changes are frequently observed in cancer and are starting to be recognized as important signatures for tumor progression and therapy. However, their functional impact and relevance to tumorigenesis remain mostly unknown. We carried out a systematic analysis to characterize the potential functional consequences of alternative splicing changes in thousands of tumor samples. This analysis revealed that a subset of alternative splicing changes affect protein domain families that are frequently mutated in tumors and potentially disrupt protein-protein interactions in cancer-related pathways. Moreover, there was a negative correlation between the number of these alternative splicing changes in a sample and the number of somatic mutations in drivers. We propose that a subset of the alternative splicing changes observed in tumors may represent independent oncogenic processes that could be relevant to explain the functional transformations in cancer, and some of them could potentially be considered alternative splicing drivers (AS drivers).

Marine Le Morvan, Andrei Zinovyev, Jean-Philippe Vert (2017 Jun 27)

**NetNorM: Capturing cancer-relevant information in somatic exome mutation data with gene networks for cancer stratification and prognosis.**

*PLoS computational biology* : e1005573 : [DOI : 10.1371/journal.pcbi.1005573](https://doi.org/10.1371/journal.pcbi.1005573)

**Résumé**

Genome-wide somatic mutation profiles of tumours can now be assessed efficiently and promise to move precision medicine forward. Statistical analysis of mutation profiles is however challenging due to the low frequency of most mutations, the varying mutation rates across tumours, and the presence of a majority of passenger events that hide the contribution of driver events. Here we propose a method, NetNorM, to represent whole-exome somatic mutation data in a form that enhances cancer-relevant information using a gene network as background knowledge. We evaluate its relevance for two tasks: survival prediction and unsupervised patient stratification. Using data from 8 cancer types from The Cancer Genome Atlas (TCGA), we show that it improves over the raw binary mutation data and network diffusion for these two tasks. In doing so, we also provide a thorough assessment of somatic mutations prognostic power which has been overlooked by previous studies because of the sparse and binary nature of mutations.

Helga Paula Török, Victor Bellon, Astrid Konrad, Martin Lacher, Laurian Tonenchi, Matthias Siebeck, Stephan Brand, Enrico Narciso De Toni (2017 Apr 8)

**Functional Toll-Like Receptor (TLR)2 polymorphisms in the susceptibility to inflammatory bowel disease.**

*PloS one* : e0175180 : [DOI : 10.1371/journal.pone.0175180](https://doi.org/10.1371/journal.pone.0175180)

**Résumé**

The recent genome-wide association studies (GWAS) in inflammatory bowel disease (IBD) suggest significant genetic overlap with complex mycobacterial diseases like tuberculosis or leprosy. TLR variants have previously been linked to susceptibility for mycobacterial diseases. Here we investigated the contribution to IBD risk of two TLR2 polymorphisms, the low-prevalence variant Arg753Gln and the GTn microsatellite repeat polymorphism in intron 2. We studied association with disease, possible correlations with phenotype and gene-gene interactions.

Naylor P., Laé M., Reyat F., Walter T. (2017 Jan 1)

**Nuclei segmentation in histopathology images using deep neural networks**

*2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*

**Résumé**

**Année de publication : 2016**

---

Anne-Sophie Hamy, Hélène Bonsang-Kitzis, Marick Lae, Matahi Moarii, Benjamin Sadacca, Alice Pinheiro, Marion Galliot, Judith Abecassis, Cecile Laurent, Fabien Reyat (2016 Dec 23)

**A Stromal Immune Module Correlated with the Response to Neoadjuvant Chemotherapy, Prognosis and Lymphocyte Infiltration in HER2-Positive Breast Carcinoma Is Inversely Correlated with Hormonal Pathways.**

*PloS one* : e0167397 : [DOI : 10.1371/journal.pone.0167397](https://doi.org/10.1371/journal.pone.0167397)

**Résumé**

HER2-positive breast cancer (BC) is a heterogeneous group of aggressive breast cancers, the prognosis of which has greatly improved since the introduction of treatments targeting HER2. However, these tumors may display intrinsic or acquired resistance to treatment, and classifiers of HER2-positive tumors are required to improve the prediction of prognosis and to develop novel therapeutic interventions.

Nikolay Tsanov, Aubin Samacoits, Racha Chouaib, Abdel-Meneem Traboulsi, Thierry Gostan, Christian Weber, Christophe Zimmer, Kazem Zibara, Thomas Walter, Marion Peter, Edouard Bertrand, Florian Mueller (2016 Sep 8)

**smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability.***Nucleic acids research* : e165**Résumé**

Single molecule FISH (smFISH) allows studying transcription and RNA localization by imaging individual mRNAs in single cells. We present smiFISH (single molecule inexpensive FISH), an easy to use and flexible RNA visualization and quantification approach that uses unlabelled primary probes and a fluorescently labelled secondary detector oligonucleotide. The gene-specific probes are unlabelled and can therefore be synthesized at low cost, thus allowing to use more probes per mRNA resulting in a substantial increase in detection efficiency. smiFISH is also flexible since differently labelled secondary detector probes can be used with the same primary probes. We demonstrate that this flexibility allows multicolor labelling without the need to synthesize new probe sets. We further demonstrate that the use of a specific acrydite detector oligonucleotide allows smiFISH to be combined with expansion microscopy, enabling the resolution of transcripts in 3D below the diffraction limit on a standard microscope. Lastly, we provide improved, fully automated software tools from probe-design to quantitative analysis of smFISH images. In short, we provide a complete workflow to obtain automatically counts of individual RNA molecules in single cells.

Solveig K Sieberts, Fan Zhu, Javier García-García, Eli Stahl, Abhishek Pratap, Gaurav Pandey, Dimitrios Pappas, Daniel Aguilar, Bernat Anton, Jaume Bonet, Ridvan Eksi, Oriol Fornés, Emre Guney, Hongdong Li, Manuel Alejandro Marín, Bharat Panwar, Joan Planas-Iglesias, Daniel Poglayen, Jing Cui, Andre O Falcao, Christine Suver, Bruce Hoff, Venkat S K Balagurusamy, Donna Dillenberger, Elias Chaibub Neto, Thea Norman, Tero Aittokallio, Muhammad Ammad-Ud-Din, Chloe-Agathe Azencott, Víctor Bellón, Valentina Boeva, Kerstin Bunte, Himanshu Chheda, Lu Cheng, Jukka Corander, Michel Dumontier, Anna Goldenberg, Peddinti Gopalacharyulu, Mohsen Hajiloo, Daniel Hidru, Alok Jaiswal, Samuel Kaski, Beyrem Khalfaoui, Suleiman Ali Khan, Eric R Kramer, Pekka Marttinen, Aziz M Mezlini, Bhuvan Molparia, Matti Pirinen, Janna Saarela, Matthias Samwald, Véronique Stoven, Hao Tang, Jing Tang, Ali Torkamani, Jean-Phillipe Vert, Bo Wang, Tao Wang, Krister Wennerberg, Nathan E Wineinger, Guanghua Xiao, Yang Xie, Rae Yeung, Xiaowei Zhan, Cheng Zhao, , Jeff Greenberg, Joel Kremer, Kaleb Michaud, Anne Barton, Marieke Coenen, Xavier Mariette, Corinne Miceli, Nancy Shadick, Michael Weinblatt, Niek de Vries, Paul P Tak, Danielle Gerlag, Tom W J Huizinga, Fina Kurreeman, Cornelia F Allaart, S Louis Bridges, Lindsey Criswell, Larry Moreland, Lars Klareskog, Saedis Saevarsdottir, Leonid Padyukov, Peter K Gregersen, Stephen Friend, Robert Plenge, Gustavo Stolovitzky, Baldo Oliva, Yuanfang Guan, Lara M Mangravite, S Louis Bridges, Lindsey Criswell, Larry Moreland, Lars Klareskog, Saedis Saevarsdottir, Leonid Padyukov, Peter K Gregersen, Stephen Friend, Robert Plenge, Gustavo Stolovitzky, Baldo Oliva, Yuanfang Guan, Lara M Mangravite (2016 Aug 24)

**Crowdsourced assessment of common genetic contribution to predicting anti-**

**TNF treatment response in rheumatoid arthritis.**

*Nature communications* : 12460 : [DOI : 10.1038/ncomms12460](https://doi.org/10.1038/ncomms12460)

**Résumé**

Rheumatoid arthritis (RA) affects millions world-wide. While anti-TNF treatment is widely used to reduce disease progression, treatment fails in ~one-third of patients. No biomarker currently exists that identifies non-responders before treatment. A rigorous community-based assessment of the utility of SNP data for predicting anti-TNF treatment efficacy in RA patients was performed in the context of a DREAM Challenge ([http://www.synapse.org/RA\\_Challenge](http://www.synapse.org/RA_Challenge)). An open challenge framework enabled the comparative evaluation of predictions developed by 73 research groups using the most comprehensive available data and covering a wide range of state-of-the-art modelling methodologies. Despite a significant genetic heritability estimate of treatment non-response trait ( $h^2=0.18$ ,  $P$  value=0.02), no significant genetic contribution to prediction accuracy is observed. Results formally confirm the expectations of the rheumatology community that SNP information does not significantly improve predictive performance relative to standard clinical traits, thereby justifying a refocusing of future efforts on collection of other data.

Mayumi Isokane, Thomas Walter, Robert Mahen, Bianca Nijmeijer, Jean-Karim Hériché, Kota Miura, Stefano Maffini, Miroslav Penchev Ivanov, Tomoya S Kitajima, Jan-Michael Peters, Jan Ellenberg (2016 Mar 9)

**ARHGEF17 is an essential spindle assembly checkpoint factor that targets Mps1 to kinetochores.**

*The Journal of cell biology* : 647-59 : [DOI : 10.1083/jcb.201408089](https://doi.org/10.1083/jcb.201408089)

**Résumé**

To prevent genome instability, mitotic exit is delayed until all chromosomes are properly attached to the mitotic spindle by the spindle assembly checkpoint (SAC). In this study, we characterized the function of ARHGEF17, identified in a genome-wide RNA interference screen for human mitosis genes. Through a series of quantitative imaging, biochemical, and biophysical experiments, we showed that ARHGEF17 is essential for SAC activity, because it is the major targeting factor that controls localization of the checkpoint kinase Mps1 to the kinetochore. This mitotic function is mediated by direct interaction of the central domain of ARHGEF17 with Mps1, which is autoregulated by the activity of Mps1 kinase, for which ARHGEF17 is a substrate. This mitosis-specific role is independent of ARHGEF17's RhoGEF activity in interphase. Our study thus assigns a new mitotic function to ARHGEF17 and reveals the molecular mechanism for a key step in SAC establishment.

Víctor Bellón, Véronique Stoven, Chloé-Agathe Azencott (2016 Jan 19)

**Multitask feature selection with task descriptors**

*Pacific Symposium on Biocomputing* : 261-72

**Résumé**

Machine learning applications in precision medicine are severely limited by the scarcity of data to learn from. Indeed, training data often contains many more features than samples. To alleviate the resulting statistical issues, the multitask learning framework proposes to learn different but related tasks jointly, rather than independently, by sharing information between these tasks. Within this framework, the joint regularization of model parameters results in models with few non-zero coefficients and that share similar sparsity patterns. We propose a new regularized multitask approach that incorporates task descriptors, hence modulating the amount of information shared between tasks according to their similarity. We show on simulated data that this method outperforms other multitask feature selection approaches, particularly in the case of scarce data. In addition, we demonstrate on peptide MHC-I binding data the ability of the proposed approach to make predictions for new tasks for which no training data is available.

Jiao Y., Korba A., Sibony E. (2016 Jan 1)

**Controlling the distance to a Kemeny consensus without computing it**

*Proceedings of The 33rd International Conference on Machine Learning* : 2971-2980

**Résumé**

Machairas V., Baldeweck T., Walter T., Decencière E. (2016 Jan 1)

**New general features based on superpixels for image segmentation learning**

*International Symposium on Biomedical Imaging* *International Symposium on Biomedical Imaging*

**Résumé**

Chloé-Agathe Azencott (2016 Jan 1)

**Network-Guided Biomarker Discovery**

*Lecture Notes in Computer Science* *Machine Learning for Health Informatics* : 9605 : 319-336 :

[DOI : 10.1007/978-3-319-50478-0\\_16](https://doi.org/10.1007/978-3-319-50478-0_16)

**Résumé**

Identifying measurable genetic indicators (or biomarkers) of a specific condition of a biological system is a key element of precision medicine. Indeed it allows to tailor diagnostic, prognostic and treatment choice to individual characteristics of a patient. In machine learning terms, biomarker discovery can be framed as a feature selection problem on whole-genome data sets. However, classical feature selection methods are usually underpowered to process these data sets, which contain orders of magnitude more features than samples. This can be addressed by making the assumption that genetic features that are linked on a biological network are more likely to work jointly towards explaining the phenotype of interest. We review here three families of methods for feature selection that integrate prior

knowledge in the form of networks.

#### Année de publication : 2015

---

Kévin Vervier, Pierre Mahé, Maud Tournoud, Jean-Baptiste Veyrieras, Jean-Philippe Vert (2015 Nov 22)

#### **Large-scale machine learning for metagenomics sequence classification.**

*Bioinformatics (Oxford, England)* : 1023-32 : [DOI : 10.1093/bioinformatics/btv683](https://doi.org/10.1093/bioinformatics/btv683)

#### Résumé

Metagenomics characterizes the taxonomic diversity of microbial communities by sequencing DNA directly from an environmental sample. One of the main challenges in metagenomics data analysis is the binning step, where each sequenced read is assigned to a taxonomic clade. Because of the large volume of metagenomics datasets, binning methods need fast and accurate algorithms that can operate with reasonable computing requirements. While standard alignment-based methods provide state-of-the-art performance, compositional approaches that assign a taxonomic class to a DNA read based on the k-mers it contains have the potential to provide faster solutions.

Matahi Moarii, Valentina Boeva, Jean-Philippe Vert, Fabien Reyat (2015 Oct 30)

#### **Changes in correlation between promoter methylation and gene expression in cancer.**

*BMC genomics* : 873 : [DOI : 10.1186/s12864-015-1994-2](https://doi.org/10.1186/s12864-015-1994-2)

#### Résumé

Methylation of high-density CpG regions known as CpG Islands (CGIs) has been widely described as a mechanism associated with gene expression regulation. Aberrant promoter methylation is considered a hallmark of cancer involved in silencing of tumor suppressor genes and activation of oncogenes. However, recent studies have also challenged the simple model of gene expression control by promoter methylation in cancer, and the precise mechanism of and role played by changes in DNA methylation in carcinogenesis remains elusive.

Laurent Guyon, Christian Lajaunie, Frédéric Fer, Ricky Bhajun, Eric Sulpice, Guillaume Pinna, Anna Campalans, J Pablo Radicella, Philippe Rouillier, Mélissa Mary, Stéphanie Combe, Patricia Obeid, Jean-Philippe Vert, Xavier Gidrol (2015 Sep 19)

#### **$\Phi$ -score: A cell-to-cell phenotypic scoring method for sensitive and selective hit discovery in cell-based assays.**

*Scientific reports* : 14221 : [DOI : 10.1038/srep14221](https://doi.org/10.1038/srep14221)

**Résumé**

Phenotypic screening monitors phenotypic changes induced by perturbations, including those generated by drugs or RNA interference. Currently-used methods for scoring screen hits have proven to be problematic, particularly when applied to physiologically relevant conditions such as low cell numbers or inefficient transfection. Here, we describe the  $\Phi$ -score, which is a novel scoring method for the identification of phenotypic modifiers or hits in cell-based screens.  $\Phi$ -score performance was assessed with simulations, a validation experiment and its application to gene identification in a large-scale RNAi screen. Using robust statistics and a variance model, we demonstrated that the  $\Phi$ -score showed better sensitivity, selectivity and reproducibility compared to classical approaches. The improved performance of the  $\Phi$ -score paves the way for cell-based screening of primary cells, which are often difficult to obtain from patients in sufficient numbers. We also describe a dedicated merging procedure to pool scores from small interfering RNAs targeting the same gene so as to provide improved visualization and hit selection.