6th BioInformatics Workshop

« Cancer Genomics and Data Integration »

8-9 April 2013 - Strasbourg

Abstracts

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

BAHLAWANE Christelle
Université du Luxembourg
Collaborateur Scientifique
Luxembourg
christelle.bahlawane@uni.lu
+3524666446191

Gastrointestinal stromal tumors (GIST) are the most common mesenchymal tumors of the GI tract. Until 1998, surgery was the only treatment available and the disease outcome was very poor for patient with large tumors (>5cm). The discovery of the driver mutations of GISTs1 and the development of suitable inhibitors (imatinib, sunitinib, and more recently hsp90 inhibitors)2,3,4 have revolutionized the treatment of the patients. It turned out that 80-85% of the patients carry an activating mutation in the proto-oncogenic receptor tyrosine kinase (KIT) while 7-10% harbor a mutation in the related receptor PDGFRα (platelet-derived growth factor receptor)5. Until now, the success of the treatment depends on the type of mutations, the tumor stage and the development or not of secondary resistance to the drug. Consequently, targeted therapies are required! For this, it is essential to understand the biological mechanisms of response and resistance in these tumors.
We developed an "in-vitro model" allowing the investigation at the molecular level of the singularities of each type of mutation as well as the features shared by all of them, avoiding the high background found while collecting patient's samples. Using this model, we compared the signals resulting from different activating mutations in KIT and PDGFRα to the wild type signaling using classical molecular biology, combined with high throughput technologies such as transcriptomics and metabolomics. Integrating all these data will help developing personalized drugs and /or finding specific biomarkers allowing the identification of the best treatment for each patient.

1. Hirota, S. Gain-of-Function Mutations of c-kit in Human Gastrointestinal Stromal Tumors. Science 279, 577–580 (1998).
2. Joensuu, H. & Dimitrijevic, S. Tyrosine kinase inhibitor imatinib (STI571) as an anticancer agent for solid tumours. ANNALS OF MEDICINE 33, 451–455 (2001).
3. Floris, G. et al. The heat shock protein 90 inhibitor IPI-504 induces KIT degradation, tumor shrinkage, and cell proliferation arrest in xenograft models of gastrointestinal stromal tumors. Molecular cancer therapeutics 10, 1897–908 (2011).
4. Papaetis, G. S. & Syrigos, K. N. Targeted therapy for gastrointestinal stromal tumors: current status and future perspectives. Cancer metastasis reviews 29, 151–70 (2010).
5. Corless, C. L., Fletcher, J. a & Heinrich, M. C. Biology of gastrointestinal stromal tumors. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 22, 3813–25 (2004).

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

BAIG Komal
Université du Luxembourg
PhD Student
Luxembourg
komal.baig@uni.lu
00352 4666446966

Colorectal cancer (CRC) is one of the most frequent and deadly cancers in the western world. The development of CRC is a multistep progress characterized by the accumulation of genetic alterations and epigenetic alterations, often referred to as the adenoma-carcinoma sequence. The suppressor of cytokine signaling (SOCS) proteins are a family of Src homology 2 domain-containing proteins and consist of eight members (SOCS 1–7 and CIS, cytokine inducible SH2-containing protein). SOCS family members have been strongly implicated in the negative regulation of cytokine signal transduction pathways and are emerging as potential tumor suppressors. Previous work on the tumor suppressive role of some SOCS proteins has been carried out in mice models of CRC and gastrointestinal diseases but implications of the SOCS members in human forms of CRC still remain to be discovered. The aim of this work was to examine the role of the SOCS family genes in the context of human colon tumorigenesis. Gene expression profiling of SOCS family members was done by integrating publicly available microarray expression data on colon cancer and gastrointestinal diseases in humans from public databases.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

BLEIN Sophie
Centre de Recherche en Cancérologie de Lyon
Doctorante
LYON
sophie.blein@gmail.com

Factors associated with oxidative stress and cancer risk in the Breast and Prostate Cancer Cohort Consortium (BPC3)

Sophie Blein a,b,c,d,e, Sonja Berndt f, Amit D. Joshi g, Daniele Campa h, Regina G. Ziegler f, Elio Riboli i, and David G. Cox a,b,c,d,e,i,* on Behalf of the NCI Breast and Prostate Cancer Cohort Consortium

a Université de Lyon, F-69000 Lyon, France
b Université Lyon 1, ISPB, Lyon, F-69622, France, F-69000 Lyon, France
c INSERM U1052, Centre de Recherche en Cancérologie de Lyon, F-69000 Lyon, France
d CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, F-69000 Lyon, France
e Centre Léon Bérard, F-69008 Lyon, France
f Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD
g Department of Epidemiology, Harvard School of Public Health, Boston, MA
h German Cancer Research Center (DKFZ), Heidelberg, Germany School of Public Health
I Imperial College, London, UK

Background Both endogenous factors (genomic variations) and exogenous factors (environmental exposures, lifestyle) have an impact on the balance of reactive oxygen species (ROS). Furthermore, the electron transport chain and oxidative respiration of the mitochondria is a natural source of ROS. Variants of the ND3 (rs2853826 or G10398A) gene of the mitochondrial genome are hypothesized to differ with respect to electron transport chain efficiency and generation of ROS. Manganese superoxide dismutase (MnSOD) and glutathione peroxidase (GPX-1) are enzymes responsible for the first steps in the detoxification of ROS. Single nucleotide polymorphisms, rs4880 Val16Ala in the MnSOD gene and rs1050450 Pro198Leu in GPX-1, are associated with changes in enzymatic activity and detoxification of ROS. In this study, we test the hypothesis that these variants, particularly with respect to statistical interactions between rs4880 - rs1050450, and alcohol consumption - rs2853826 are associated with risk of breast and prostate cancer.
Methods Case-control studies were conducted in the framework of the Breast and Prostate Cancer Cohort Consortium (BPC3), consisting of nine well-established cohorts. 13511 women and 8490 men diagnosed with breast and prostate cancer respectively were included in analyses to examine the interaction between rs4880 and rs1050450 in MnSOD and GPX-1. Similarly, 10726 women and 7532 men with breast and prostate cancer were included in our study to investigate the interaction between alcohol intake and a polymorphism in mitochondrial

ND3. The 3 polymorphisms (rs4880, rs1050450 and rs2853826) were genotyped using TaqMan assays. Alcohol consumption was ascertained by way of epidemiological questionnaires. Logistic regression models were used to evaluate disease risk, while proportional hazard models were used to study survival between groups defined by genotype and alcohol consumption.
Results Interactions between polymorphisms in MnSOD and GPX-1, or between mitochondrial polymorphisms and alcohol intake were not observed to have a significant effect on breast (p-interaction of 0.34 and 0.98 respectively) or prostate cancer risk (p-interaction of 0.49 and 0.50 respectively). We observe a weak inverse association between prostate cancer and GPX-1 Leu198Leu carriers (OR 0.87, 95% CI 0.79 – 0.97, p = 0.01). We also observe changes in breast cancer overall survival with respect to genotype at A10398G and alcohol consumption (HR 0.66 95% CI 0.49 – 0.88 for G10398 carries who consume alcohol).
Conclusion Given the high power (> 80% for each analysis) we dispose of in these two studies, it is unlikely that interactions tested have more than moderate effects on breast or prostate cancer risk. The weak associations observed between prostate cancer risk and the GPX-1 polymorphism rs1050450 could be due to chance, as well as the putative interaction between alcohol consumption and the ND3 rs2853826 polymorphism on breast cancer survival need both further epidemiological and biological confirmation.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

CHAMARD Clémence
Université de Lorraine
Etudiante
Vandoeuvre lès Nancy
clemence.chamard@laposte.net

From here to ERalpha36, a new predictive marker for breast tumor therapeutic response ?

Clémence Chamard1, Helene Dumond1, Alain Jung2, Amand Chesnel1, Christine Macabre2, Stéphane Flament1, Sonia Ledrappier2, Joseph Abecassis2and Taha Boukhobza1.

1 CNRS-Université de Lorraine, UMR 7039, Centre de Recherches en Automatique de Nancy, BP70239, Vandoeuvre lès Nancy, F-54506, France
2 EA 3430. Centre Paul Strauss.3, rue Porte de l'Hôpital. 67000. Strasbourg.

Breast cancer is the main cause of cancer-induced morbidity and mortality in women. Breast tumors are usually classified according to their nuclear estrogen receptor status. Such a classification led to the use of endocrine therapeutic agents against ER-positive tumors [ER+]. Nevertheless, numerous therapeutic failures are observed due to unclear resistance mechanism.
In this context, the objective of our study is to evaluate the efficiency to take into account the estrogen membrane receptors (ERalpha36 and GPER) in the diagnostic classification of breast tumors and thus to construct a gene network useful for therapeutic predictions. More precisely, we are interested in improving the classification of so-called "ER-positive" breast tumors by taking into account the key role of ERalpha36 and GPER in the control of breast tumors non genomic estrogen response and metastatic potential.The aim is to better define the field of acceptance of [ER+] versus [ER-] classification and to help the clinicians choosing the best therapeutic strategy for each patient. To do so, we set up a retrospective study, performed on hundreds of breast tumor samples: ERalpha36, GPER and metastatic marker expressions were measured by real-time PCR in almost 100 [ER+] as well as 60 triple-negative [ER-] tumor samples. Then, we performed statistical analyses between gene expression levels and clinical parameters (grade, survival, treatment). The proposed approach for the analysis consists in performing the classification pertinence of a gene in three steps: after choosing a target gene and a classification threshold to separate the sample into two categories, we identify, using a Bayesian inference technique, two networks which are involved the metastatic markers. Then, we compute the distance between the two networks, which is defined to take into account both the structural differences between the networks (existence or not of relations between the markers) and the compartmental differences (behavioral differences in the relations). The distances between the networks

represent then the classification performance of the target gene and allow us to find the more pertinent classifiers. Taken together, the results from this project should lead to (i) a better understanding of the breast tumor hormone sensitive status, (ii) a validation of new predictive markers of response in order to improve therapeutic orientation, (iii) the potential discovery of new therapeutic targets in triple-negative tumors and (iv) the development of a new methodology to exhibit pertinent classifiers for transcriptome analysis based studies.

ABSTRACT

CIANFERANI Sarah
CNRS
Chercheur
Strasbourg
sarah.cianferani@unistra.fr
0368852679

Proteomics for cancer biomarker discovery : where do we stay in finding a needle in a haystack?

Biomarker discovery for clinical purposes is one of the major areas in which proteomics has been the subject of intense interest and activity. Proteomic-scale capabilities enable thousands of proteins to be identified from complex mixtures within one run. However, despite considerable efforts, development of innovative technologies and multiple approaches, the contribution of proteomic methods has been disappointing. This is partly due to the lack of coherent pipeline that connects biomarker discovery methods and validation tools. In this perspective we will highlight and analyze main causes for this limited success. Issues ranging from sample preparation, analytical developments, proteomic data interpretation and bioinformatic validation will be addressed. Global proteomic strategies for biomarker discovery phase as well as targeted approaches for biomarker validation will be presented.

ABSTRACT

HAZEEB Zubair
Aligarh Muslim University
Aligarh, India
zubairhaseeb02@gmail.com
+919897957998

A PROOXIDANT MECHANISM OF CANCER CHEMOPREVENTIVE PROPERTIES OF PLANT POLYPHENOLS

H Zubair, HY Khan, MF Ullah, A Ahmad, SM Hadi
Department of Biochemistry, Faculty of Life Sciences, Aligarh Muslim University, Aligarh-202002, India
Zubairhaseeb02@gmail.com

To account for the observed anticancer properties of plant polyphenols, we have earlier proposed a mechanism which involves the mobilization of endogenous copper ions by polyphenols leading to the generation of reactive oxygen species (ROS) that serve as proximal DNA cleaving agents. As a further confirmation we show that oral administration of copper to rats leads to elevated copper levels in lymphocytes. When such lymphocytes were isolated and treated with different polyphenols, an increased level of DNA breakage in comet assay was observed. Preincubation of lymphocytes having elevated copper levels with the membrane permeable copper chelator neocuproine, resulted in inhibition of polyphenol induced DNA degradation. However, membrane impermeable chelator of copper, as well as iron and zinc chelators were ineffective in causing such inhibition. We have further shown that normal breast epithelial cells cultured in a medium supplemented with copper, become sensitized to polyphenol-induced growth inhibition and are found to have an increased expression of copper transporter ctr1. We have also shown that polyphenols were able to inhibit cell proliferation and induce apoptosis in different cancer cell lines and that such cell death is prevented to a significant extent by copper chelator neocuproine and ROS scavengers. It is well established that serum and tissue concentrations of copper are greatly increased in various malignancies. In view of this fact, the present results further confirm our earlier findings and strengthen our hypothesis that an important anticancer mechanism of plant polyphenols could be the mobilization of intracellular copper leading to ROS-mediated cellular DNA breakage.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

KAOMA Tony
CRP Santé
Bioinformaticien
Luxembourg
tony.kaoma@crp-sante.lu

Minimizing the Verhaak's 840-gene signature of Glioblastoma multiform subtypes.

Tony KAOMA1, Daniel STIEBER2, Petr NAZAROV1, Nathalie NICOT1 and Laurent VALLAR1

1 Genomic research unit, CRP-Santé, Luxembourg
2 Neuro-Oncology Laboratory, CRP-Santé, Luxembourg

Background
Several gene signatures have been proposed to classify gene expression profiling of Glioblastoma multiform (GBM) patients in different GBM subtypes. One of the most popular GBM gene signatures was introduced by Verhaak et al and includes 840 genes. This signature is able to stratify GBM into four subtypes: Proneural, Neural, Classical and Mesenchymal which are associated with particular phenotypes, expression of specific genes, chromosomal aberration, age of patients and more interestingly with response to treatment and life expectancy. Although powerful, this signature cannot be used as such for diagnostic or pronostic purposes since the number of genes it includes is too high. There is a need to reduce the number of genes in the Verhaak 840-gene signature without affecting its performance.

Method
To generate a compact gene signature, we used ClaNC (CLassifying microarrays to Nearest Centroids) to select set of genes (Dabney AR, Bioinformatics 2005) and three difference methods to evaluate the ability of the selected genes to classify GBM samples (centroid approach, partial least squares regression and nearest template prediction). Our approach was tested on the 202 microarrays previously used by Verhaak.

Result
By an iterative decrease of the number of genes, we generated a 60-gene signature where each four GBM subtypes are represented by 15 genes. Our data indicated that the proposed signature performs as well as the Verhaak's 840-gene signature. Using this compact signature, we were able to classify an independent set of 21 GBM patients based on microarray expression analysis.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

LENNON Sarah
Laboratoire de spectrométrie de masse bio-organique
Doctorante
STRASBOURG
sarah.lennon@etu.unistra.fr
0368852624

Membrane proteomics to search for specific glioblastoma cancer stem cells biomarkers.

Sarah Lennon (1), Jean-Michel Saliou (1), Jihu Dong (2), Hervé Chneiweiss (3), Alain Van Dorsselaer (1), Christine Carapito (1), Marie-Claude Kilhoffer (2), Jacques Haiech (2), Sarah Cianférani (1).

Glioblastomas (GBMs) are the most common type of malignant tumors in central nervous system in adults. GBMs remain one the most lethal and least successfully treated primary brain tumors, the median survival of treated patients being less than 16 months. Poor GBMs prognosis can be explained by difficulties in cancer detection and treatment resistance of these aggressive tumors which are highly infiltrative (in brain and spinal cord) and extremely resistant to conventional treatments (1). Today bevacizumab (Avastin, anti VEGF) is the only therapeutic recombinant protein approved by the FDA to treat GBM (2). Cancer stem cells (CSCs) might be responsible for treatment resistance and tumor recurrence. According to the CSC hypothesis, current therapies that are extremely cytotoxic to the bulk of highly proliferative tumor cells fail to obliterate the relatively quiescent and resistant CSCs, thereby allowing these cells to survive and drive tumor recurrence. Thus molecular targeting of CSCs in GBMs (named GSCs) may directly improve the efficacy of current therapies, by reducing tumor recurrence. Thus molecular targeting of CSCs in GBMs (named GSCs) may directly improve the efficacy of current therapies, by reducing tumor recurrence.

Proteomic studies have already provided important insights in the pathophysiology of GBMs but scarce efforts have been made on glioblastoma CSCs. Considering that membrane proteins (MPs) play key roles in the development and progression of cancer and that a number of drug targets are based on MPs, the aim of the present work is to develop an original and specific membrane proteomic methodology to search for new GSCs biomarkers, that could be further developed for active immunotherapy (cancer vaccines).

We present here a 1D-gel membrane proteomic strategy to search for specific glioblastoma CSCs biomarkers. Briefly ghost membrane protein preparations from different GSCs and control (NSC, HEK and HA) cell lines were produced. Further proteomic analysis consists in 1D-gel separation, followed by optimized nanoLC-MS/MS (Ion Trap Amazon, Bruker) and protein identifications in databases. To ensure maximum confidence, the data were validated by two search engines: Mascot and Omssa. Difficulties related to proteomics will be particularly discussed here.

ABSTRACT

MAGLOTT-ROTH Anne
IGBMC-Plateforme de criblage
Illkirch
maglottr@igbmc.fr

The High Throughput Cell-based Screening Facility of IGBMC, Illkirch, France
Benoît Fischer, Anne Maglott-Roth, Amélie Weiss, Florence Gross and Laurent Brino
IGBMC, UMR7104 CNRS UdS Inserm U964, 1 rue Laurent Fries, 67404 Illkirch cedex
Contact : www.igbmc.fr

Founded in 2005 by the Cancéropôle Grand Est (CGE) for consolidating the potential of the Alsacian research, this creation has been made possible by initial financial supports from local governments and from the CGE. Besides, the facility takes revenues from contributions of granted research projects. Located at the CEBGS building of IGBMC, the platform has a 100 sqm laboratory with two P2 laboratories equipped for high throughput mammalian cell culture. The platform's core team consists of four engineers and one technician, with strong experience in RNA interference, gene expression, mammalian cell culture and cellular biology. A transfected cell array is a biotechnological tool to speed up the characterization of the role of genes suspected to be responsible for a particular cellular phenotype. It is based on a high throughput screening combined with single cell high content phenotypic analysis. The platform objective is to identify genes and pathways implied in human diseases, with high therapeutic potential for treatment, using molecule transfection in cells (siRNA, …). High throughput cell transfection and phenotypic analysis are operated on a TECAN robotic station in Class II cabinet and on two automated high content imaging and analysis systems, respectively. A high speed data transfer network has been implemented to fulfil all user needs in term of data storage and access, distance image analysis. Because of its powerful potential in Drug Discovery and in the understanding of the biology of the cell functions, the development of this facility is well suited to give support to diverse research projects covering Cancer, Stem cell & Development, Virology fields. In 2012, the facility joined the RNAi Global Intitiative consortium and acquired human and mouse genome siRNA libraries. It also joined the Ingestem national infrastructure to develop novel cell models using stem cell differentiation technology.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

MICHAUT Magali
Netherlands Cancer Institute
Postdoc
Amsterdam
m.michaut@nki.nl
0031 20 512 9055

Pathway mutation status predicts chemotherapy response in triple negative breast cancer

Esther H. Lips1,*, Magali Michaut2,*, Lennart Mulder1, Marlous Hoogstraat3, Marco J. Koudijs3, René Bernards2, Jelle Wesseling4, Sjoerd Rodenhuis5,†, Lodewyk F. A. Wessels2, †

1. Department of Molecular Pathology, the Netherlands Cancer Institute, Amsterdam, The Netherlands
2. Department of Molecular Carcinogenesis, the Netherlands Cancer Institute, Amsterdam, The Netherlands
3. Center for Personalized Cancer Treatment, Department of Medical Oncology UMC Utrecht, Utrecht, The Netherlands
4. Department of Pathology, Amsterdam, the Netherlands Cancer Institute, Amsterdam, The Netherlands
5. Department of Clinical Oncology, the Netherlands Cancer Institute, Amsterdam, The Netherlands
* equally contributed
† corresponding authors

Background: No targeted treatments exist for triple negative breast cancer, leaving chemotherapy as the only treatment option. Even though initial response to chemotherapy is often good, many patients relapse and develop resistance to chemotherapy.

Methods: To identify biomarkers of chemotherapy resistance and putative directed treatment targets, we performed next generation sequencing for 2000 genes involved in cancer and related processes. DNA from 31 pretreatment biopsies and matched normal blood was sequenced. Biopsies were derived from patients scheduled to receive neoadjuvant chemotherapy with doxorubicin/cyclophosphamide. Tumors were divided in responders (n=18) and non-responders (n=13), depending on whether or not a pathological complete remission (pCR) was achieved.

Results: We first checked that all patients with a confirmed germline BRCA1 mutation (n=8) were detected. We then focused on somatic mutations. Only a small number of genes were mutated in several samples and none of them were predictive of response. However, a pathway analysis showed that mutations in phosphatidylinositol signaling (PI3K) were significantly more frequent in the non-responders, with mutations present in 10/13 non-responders and 2/18 responders (adjusted-p=0.013). Similar levels of enrichment were found in the chemokine signaling and integrin signaling pathways.

Conclusion: Mutations in genes of the PI3K pathway occur frequently in triple negative breast cancers that do not achieve a pCR on neoadjuvant chemotherapy with doxorubicin/cyclophosphamide. After validation, treatment regimens that combine chemotherapy with blockage of this pathway should be investigated for these tumors.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

MULLER Arnaud
CRP-Sante
Engineer Bioinformatician
Luxembourg
arnaud.muller@crp-sante.lu
0035226970283

Title: Integrative Analysis Of Non-Small Cell Lung Cancer Combining RNA-seq and Microarrays Experiments.

Authors: Arnaud Muller1, Petr Nazarov1, Tony Kaoma1, Nathalie Nicot1, Francois Bernardin1, Laurent Vallar1

Abstract: mRNA sequencing (RNA-Seq) using high-throughput next-generation sequencing (NGS) technologies is increasingly supplanting DNA microarrays for genome-wide gene expression measurements, and is expected to overcome some of the latter's limitations. However, contrary to the already mature technology of microarray, NGS poses new challenges as its strengths and weaknesses have yet to be fully identified and the nature of RNA-seq data is not fully understood.

Here we compared the performance of these two techniques by assessing overlapping features, data reliability and confidence in the context of a functional analysis of a large expression dataset generated from non-small cell lung (NSCL) tissue specimens. We used Illumina® paired-end RNA-seq sequencing and Affymetrix® Human Exon 1.0 ST arrays to profile gene expression in 20 matched normal adjacent tissue and tumor pairs of both adenocarcinoma and squamous cell carcinoma, the 2 most frequent subtypes of NSCL tumors. Microarray data quality was assessed using GC-RMA and significance analysis was performed with limma. RNA-seq data quality was evaluated using the TopHat-Cufflinks suite, while significance analysis was performed by processing sequencing data with EdgeR. Finally, functional analysis was performed by analyzing the datasets using Ingenuity®, DAVID software and a dedicated proprietary knowledge base focused on oncogenes.

Our study showed that although RNA-seq undeniably offers some benefits compared to microarrays, NGS still have several limitations and constraints. Our data indicated however that the two techniques are more complementary than competitive. Comparison of microarray and NGS expression data revealed a high level of concordance, and combining the two datasets significantly improved the robustness of functional analysis.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

NAZAROV Petr
CRP-Sante
Chercheur biostatisticien
Luxembourg
petr.nazarov@crp-sante.lu
(+352) 26 970-283

Pipeline for an integrative analysis of interplay between microRNAs, transcription factors, and target genes using microarray data

Petr V. Nazarov1, Susanne E. Reinsbach2, Arnaud Muller1, Nathalie Nicot1, Demetra Philippidou2, Stephanie Kreis2, Laurent Vallar1
1 Genomics Research Unit, Centre de Recherche Public de la Santé, L-1526 Luxembourg
2 Signal Transduction Laboratory, Life Sciences Research Unit, University of Luxembourg, L-1511 Luxembourg

MicroRNAs (miRNAs) are small non-coding RNAs that negatively regulate gene expression at the post-transcriptional level. It is hypothesised that miRNAs are involved in fine-tuning fundamental cellular processes and confer robustness to biological responses. Here, we investigated simultaneously the transcriptional changes of miRNA and mRNA expression levels over time following activation of the Jak/STAT pathway by IFN-γ stimulation of melanoma cells. To examine global miRNA and mRNA expression patterns, time-series microarray data were analysed.
We established and validated a pipeline for the analysis and integration of such data including: differential expression analysis, co-expression analysis, functional annotation followed by analysis of dynamics of biological functions, prediction of up-stream regulators and, finally, integration of mRNA, miRNA and predicted transcriptional regulators into a network in order to detect regulation motifs. Our benchmarking of 3 statistical methods for detection genes regulated in time ('betr', 'timecourse' and 'limma') suggested that for one-class analysis 'limma' outperforms other two.
We observed a delayed response of miRNAs (after 24 – 48 hours) with respect to mRNAs (12 – 24 hours) and identified biological functions involved at each step of the cellular response to IFN-γ treatment. By linking expression profiles of transcriptional regulators and miRNAs with their annotated functions, we demonstrated the dynamic interplay of miRNAs, upstream regulators and genes with biological functions. Finally, our data revealed network motifs in form of feed-forward loops involving transcriptional regulators, mRNAs and miRNAs. Knowledge from integrated data on miRNAs together with dynamic and matched transcriptome data will contribute to a more comprehensive view of biological systems, their regulation and their behaviour over time.

ABSTRACT

PAUL Nicodème
Institut d'Immunologie et d'Hématologie
Strasbourg
npaul@unistra.fr

Automated pipeline for mutation discovery using family-based exome sequencing on the SOLiD platform
Nicodème PAUL, Head of Bioinformatics, GENOMAX NGS platform, INSERM UMR_S 1109, School of Medicine, Strasbourg

Advances in high-throughput DNA sequencing technologies have revolutionized the study of genetic variation in the human genome. Whole-genome sequencing currently represents the most comprehensive strategy for genome-wide variant detection. But it is rather costly for large sample sizes and variants detected in noncoding region remain largely uninterpretable. By contrast, whole exome sequencing has been widely applied in the identification of germline mutations underlying Mendelian disorders [1], somatic mutations in various cancers [2] and de novo mutation in neurodevelopmental disorders [3]. Various computational algorithms have been developed to be integrated and combined for disease causing variant characterization. Here, we present an integrated computational pipeline to fully automate the process of variant detection or mutation discovery from exome sequencing SOLiD raw data using pedigree information.

[1] Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. Genome Biol. 12, 228 (2011).
[2] Zang,Z.J., Cutcutache,I., Poon,S.L., Zhang,S.L., McPherson,J.R., Tao,J., Rajasegaran,V., Heng,H.L., Deng,N., Gan,A. et al. (2012) Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat. Genet., 44, 570–574

[3] O'Roak, B. J. et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature Genet. 43, 585–589 (2011)

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

RAFFELSBERGER Wolfgang
IGBMC
Illkirch
wolfgang.raffelsberger@igbmc.fr
0388 65 3300

Title : Multilevel Transcriptome Analysis using the automated platform GxDb
Wolfgang Raffelsberger, Laetitia Podevin, Raymond Ripp, Hélène Polveche and Olivier Poch
Laboratoire de Bioinformatique et Génomique Intégratives, IGBMC, Illkirch, France

The web-based platform GxDb (http://gx.igbmc.fr/gx/) allows automatic upload, analysis, storage of microarray-experiments as optimized data-base, and contains tools for data visualization and mining of the resultant data. In particular, this platform was designed to address multiple questions that can be formulated in parallel and to search for consistent high quality results i) in a full automatic mode or ii) in a user-interactive session over the GxDb www-portal. The data-base allows handling proprietary data via https-login into work-groups. Numerous diagnostic reports help the expert in quickly understanding the nature of a given experiment. Recent developments address finding genes found consistently over multiple parallel complementary data-treatment methods.

ABSTRACT

SHAFQUAT Azim
Aligarh Muslim University
Aligarh, India
shafquatazim@gmail.com
+919897705545

Alternatively spliced three novel transcripts of gria1 in the cerebellum and cortex of mouse brain

Shafquat Azim, AR Banday, M Tabish
Department of Biochemistry, Faculty of Life Sciences, A.M.U., Aligarh, India 202002
email: shafquatazim@gmail.com

Glutamate receptor type 1 (GluR1) subunit of α-amino-3-hydroxy-5-methyl-4-isoxazole propionate receptors plays an important role in the expression of long-term potentiation and memory formation. GluR1 is encoded by gria1 gene containing 16 exons and 15 introns in mouse. Previous studies have reported two alternatively spliced variants of this subunit. These flip and flop variants differ enormously in their properties as well as expression. In our studies, we report the presence of three new transcripts of this gene present in the cerebellum and cortex of mouse brain produced by alternative splicing at 5' end. Four new exons are reported; N1 is located in 5' untranslated region, N2 is located in the 1st intronic region while N3 and N4 are located in the 2nd intronic region. The properties of these new exons encoding N-terminal variants are highly diverse. N1, N3 and 4 are coding while N2 is a non-coding exon and results in a truncated transcript. The existence of N2 exon containing transcript is further supported by the presence of an Expressed Sequence Tag from the database. The translated amino acid sequences of these transcripts differ in the presence of signal peptide as well as in their phosphorylation and acetylation pattern. The differences in their properties might be involved in receptor modulation.

6th Intercanceropole Bioinformatics Workshop
Cancer Genomics and Data Integration
8-9th April 2013 – Strasbourg

ABSTRACT

WALTER Vincent
IGBMC
Strasbourg
walterv@igbmc.fr

Vincent Walter, Julie Thompson, Olivier Poch and Hoan Nguyen

Laboratoire de Bioinformatique et Génomique Intégratives, IGBMC

Recently, various methods for gene prioritization have emerged to identify the most promising genes among larger pools of candidates using integrative computational analysis of public and private biomedical and genomic "big data"1. Initially developed to help identify disease-causing genes, these methods are becoming cornerstone whenever genes or networks are to be selected on the basis of heterogeneous functional data. The need for the continuous collection, integration and robust exploitation of large volumes of data from multiple sources, is leading to novel developments in all aspects of information technologies.
Here, we present GEPETTO for GEne PrioriTization Tool devoted to the prioritization of genes involved in human diseases. GEPETTO takes advantage of our SM2PH Knowledegbase2-4 (decrypthon.igbmc.fr/sm2ph/) which currently incorporates in a single architecture, tools and data related to all human genes (roughly 22 000), their evolution, their tissue expressions, the effects of mutations and associated phenotypes. The modular architecture of GEPETTO allows the integration of additional data sources, ontologies or algorithms (PubMed, IDGP, EvoluCode, Reactome…), as well as user-defined scoring methods or training sets dedicated to specific applications. It currently incorporates six prioritization modules, based on gene sequence, protein-protein interactions, gene expression, disease-causing probabilities, protein evolution and genomic context). GEPETTO is an original open-source framework, written in Java/Python, distributed under the LGPL license, for gene selection and prioritization on a desktop computer that ensures confidentiality of personal data. The GEPETTO software is available at sourceforge.net/projects/gepetto/files/ or decrypthon.igbmc.fr/sm2ph/cgi-bin/gepetto.

1) Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nat Rev Genet. 2012 Jul 3;13(8):523-36.
2) Friedrich A, Garnier N, Gagnière N, Nguyen H, Albou LP, Biancalana V, Bettler E, Deléage G, Lecompte O, Muller J, Moras D, Mandel JL, Toursel T, Moulinier L, Poch O. SM2PH-db: an interactive system for the integrated analysis of phenotypic consequences of missense mutations in proteins involved in human genetic diseases. Hum Mutat. 31:127-35 (2010).
3) Luu TD, Rusu AM, Walter V, Ripp R, Moulinier L, Muller J, Toursel T, Thompson JD, Poch O, Nguyen H. MSV3d: database of human MisSense Variants mapped to 3D protein structure. Database (Oxford). 2012:bas018 (2012).
4) Luu TD, Rusu A, Walter V, Linard B, Poidevin L, Ripp R, Moulinier L, Muller J, Raffelsberger W, Wicker N, Lecompte

O, Thompson JD, Poch O, Nguyen H. KD4v: Comprehensible Knowledge Discovery System for Missense Variant. Nucleic Acids Res. 40:W71-5 (2012).

ABSTRACT

AUPET Jean-Baptiste
Cancéropôle Grand Est
Chargé de mission – Projet Microscopie Virtuelle
Besançon
jbaupet@chu-besancon.fr
+33 (0)6 62 00 10 91

An example of the CGE MiViP@GE project use: DNA sequencing data coupled with digitized microscopy slides for detection of premature genomic anomalies in primitive sclerosing cholangitis

J.B. Aupet (1), H.J. Delecluse (2), O. Klinke (2), M. Guenneugues (1), S. Valmary-Degano (3).

(1) Cancéropôle Grand-Est, (2) German Cancer Research Center DKFZ, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany, (3) CHRU Besançon, 3 Boulevard Flemming, 25000 Besançon, France.

Introduction: In order to demonstrate the added value of sharing digitized pathology slides in multicentric studies as permitted by the MiVIP@GE platform currently implemented by the Cancéropôle Grand-Est (CGE). We will take profit of a genomic pilot research project to probe the capabilities of the system. Digitized slides are assigned to research projects of CGE, accessible only by authorized persons. This project builds up on collaboration between Besançon University Hospital and the DKFZ in Heidelberg, associates microscopic slides, extracted DNA qualification data as well deep sequencing results for colonic biopsies in patients with primary sclerosing cholangitis (PSC) associated to inflammatory bowel disease.

Material and Methods: MiViP@GE allows researchers share all clinical and experimental data useful for the research project run in common. In the pilot project, these data are, for each patient:

Anonymized identifier, age, initial diagnosis with date and following diagnosis on a time line,slides for morphologic control (organ, lesion, percentage of tumor cells), DNA extraction data (concentration, Optical Density, quantity of DNA), sequencing results.

DNA samples are submitted to high throughput sequencing on an Hiseq sequencer. The obtained sequences are aligned to reference human genomes and the mutations that are present in the tumour or in the inflamed tissue but not in the controls are listed. The most important of these mutations are then confirmed by conventional Sanger sequencing. Their frequency of these mutations in a pool of 50 PSC samples is determined by Sanger sequencing. These results will be shared between the groups involved in the project. Authorized researchers must be able to access and add all the data required for the project, text file, excel, pdf, dicom imaging, MRI, scanner...

Results: Macrodissection was performed on liver and colonic samples to separately analyze the abnormal epithelium and the normal underlying conjunctive tissue (Altogether 10 samples). Morphological controls of the macrodissection were scanned and discussed between the partners of the project using the virtual microscopy

platform. We have identified multiple mutations in genes that are involved in the control of cell proliferation and of genomic integrity. Some of these mutations were already present in the inflamed tissue and could possibly be used as predictive markers of neoplastic transformation.

Conclusion: Digital pathology is now possible thanks to the new generation scanners and sharing software supporting virtual slides streaming over Internet. The CGE initiative aims at giving access to this new generation of tools through the implementation of a portal opened to researchers within the inter-region and their worldwide collaborators. Particular attention is brought to the fast developments in this field, with the concern of chosing solutions ensuring the highest compatibility level. Such opportunities and issues are discussed within a nation-wide working group that gathers the digital pathology promoters within each canceropole.