IN PARTNERSHIP WITH:
INSERM

Université de Bordeaux

Activity Report 2015

Project-Team SISTM

Statistics In System biology and Translational Medicine

RESEARCH CENTER
Bordeaux - Sud-Ouest

THEME
Computational Neuroscience and Medecine

Table of contents

Project-Team SISTM

Creation of the Team: 2013 April 02, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

3.3.2. - Data mining
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.4. - Optimization and learning
3.4.5. - Bayesian methods
6.1.1. - Continuous Modeling (PDE, ODE)
6.2.4. - Statistical methods
6.3.4. - Model reduction

Other Research Topics and Application Domains:

1.1. - Biology
1.1.11. - Systems biology
1.1.6. - Genomics
1.1.7. - Immunology
1.1.9. - Bioinformatics
1.4. - Pathologies
2.2.4. - Infectious diseases
2.2.5. - Immune system diseases
2.3. - Epidemiology
2.4.1. - Pharmaco kinetics and dynamics
2.4.2. - Drug resistance
2.8. - Sports, performance, motor skills

SISTM team is localized at the Carreire site of Bordeaux University.

1. Members

Research Scientists

Daniel Commenges [INSERM, Senior Researcher, HdR]
Laura Richert [Univ. Bordeaux II, Assistant Professor (AHU)]

Faculty Members

Rodolphe Thiebaut [Team leader, Univ. Bordeaux II, Professor, HdR]
Marta Avalos Fernandez [Univ. Bordeaux II, Associate Professor]
Robin Genuer [Univ. Bordeaux, Associate Professor]
Benoit Liquet [Univ. Bordeaux II, Associate Professor, until Aug 2015, HdR]

Engineers

Chariff Alkhassim [INSERM, from Mar 2015]
Henri Bonnabau [Univ. Bordeaux II]
Hadrien Lorenzo [INSERM, from Oct 2015]

Aaron Ayllon Benitez [INSERM, from Oct 2015]

PhD Students

Boris Hejblum [Univ. Bordeaux, until Jun 2015]
Ana Jarne Munoz [INSERM, until Dec 2015]
Chloe Pasin [Univ. Bordeaux, from Sep 2015]
Marie-Quitterie Picat [Univ. Bordeaux II, until Nov 2015]
Perrine Soret [INSERM, from Sep 2014]
Laura Villain [Univ. Bordeaux, from Feb 2015]
Mélanie Née [INSERM, from Oct 2015]

Administrative Assistants

Catherine Cattaert Megrat [Inria]
Anton Ottavi [INSERM, European Projects coordinator]
Destandau Eugénie [INSERM, Communication]

Others

Gaëlle Lefort [Inria, Master Internship, from Apr 2015 until Sep 2015]
Marina Travanca [Inria, Master Internship, from Mar 2015 until Sep 2015]
Maeva Kyheng [Univ. Bordeaux II, Master Internship, from Apr 2015 until Jun 2015]
Emilie Chanfreau [Univ. Bordeaux II, Master Internship, from Apr 2015 until Jun 2015]

2. Overall Objectives

2.1. Overall Objectives

The overall objective of SISTM is to develop statistical methods for the integrative analysis of health data, especially those related to clinical immunology to answer specific questions risen in the application field.

To reach this objective we are developing statistical methods belonging to two main research areas:

Statistical and mechanistic modeling, especially based on ordinary differential equation systems, fitted to population and sparse data

Statistical learning methods in the context of high-dimensional data

These two approaches are used for addressing different types of questions. Statistical learning methods are developed and applied to deal with the high dimensional characteristics of the data. The outcome of this research leads to hypotheses linked to a restricted number of markers. Mechanistic models are then developed and used for modeling the dynamics of a few markers. For example, regularized methods can be used to select relevant genes among 20000 measured with microarray technology, whereas differential equations can be used to capture the dynamics and relationship between several genes followed over time by a q-PCR assay or RNA-seq.

Data are generated in clinical trials or biological experimentations. Our main application of interest is the immune response to vaccine or other immune interventions (such as exogenous cytokines), mainly in the context of HIV infection. The methods developed in this context can be applied in other circumstances but the focus of the team on immunology is important for the relevance of the results and their translation into practice, thanks to a longstanding collaboration with several immunologists and the implication of the team in the Labex Vaccine Research Institute. Exemples of objectives related to this application field are:

To understand how immune response is generated with immune interventions (vaccines or interleukines)

To predict what would be the immune response to a given immune intervention for designing next studies and adapting interventions to individual patients

3. Research Program

3.1. Mecanistic modelling

When studying the dynamics of a given marker, say the HIV concentration in the blood (HIV viral load), one can for instance use descriptive models summarizing the dynamics over time in term of slopes of the trajectories [37]. These slopes can be compared between treatment groups or according to patients' characteristics. Another way for analyzing these data is to define a mathematical model based on the biological knowledge of what drives HIV dynamics. In this case, it is mainly the availability of target cells (the CD4+ T lymphocytes), the production and death rates of infected cells and the clearance of the viral particles that impact the dynamics. Then, a mathematical model most often based on ordinary differential equations (ODE) can be written [30]. Estimating the parameters of this model to fit observed HIV viral load gave a crucial insight in HIV pathogenesis as it revealed the very short half-life of the virions and infected cells and therefore a very high turnover of the virus, making mutations a very frequent event [29].

Having a good mechanistic model in a biomedical context such as HIV infection opens doors to various applications beyond a good understanding of the data. Global and individual predictions can be excellent because of the external validity of a model based on main biological mechanisms. Control theory may serve for defining optimal interventions or optimal designs to evaluate new interventions [22]. Finally, these models can capture explicitly the complex relationship between several processes that change over time and may therefore challenge other proposed approaches such as marginal structural models to deal with causal associations in epidemiology [21].

Therefore, we postulate that this type of model could be very useful in the context of our research that is in complex biological systems. The definition of the model needs to identify the parameter values that fit the data. In clinical research this is challenging because data are sparse, and often unbalanced, coming from populations of subjects. A substantial inter-individual variability is always present and needs to be accounted as this is the main source of information. Although many approaches have been developed to estimate the parameters of non-linear mixed models [33], [40], [25], [31], [26], [39], the difficulty associated with the complexity of ODE models and the sparsity of the data leading to identifiability issues need further research.

3.2. High dimensional data

With the availability of omics data such as genomics (DNA), transcriptomics (RNA) or proteomics (proteins), but also other types of data, such as those arising from the combination of large observational databases (e.g. in pharmacoepidemiology or environmental epidemiology), high-dimensional data have became increasingly common. Use of molecular biological technics such as Polymerase Chain Reaction (PCR) allows for amplification of DNA or RNA sequences. Nowadays, microarray and Next Generation Sequencing (NGS) techniques give the possibility to explore very large portions of the genome. Furthermore, other assays have also evolved, and traditional measures such as cytometry or imaging have became new sources of big data. Therefore, in the context of HIV research, the dimension of the datasets has much grown in term of number of variables per individual than in term of number of included patients although this latter is also growing thanks to the multi-cohort collaborations such as CASCADE or COHERE organized in the EuroCoord network [1]. As an exemple, in a recent phase 1/2 clinical trial evaluating the safety and the immunological response to a dendritic cell-based HIV vaccine, 19 infected patients were included. Bringing together data on cell count, cytokine production, gene expression and viral genome change led to a 20 Go database [36]. This is far from big databases faced in other areas but constitutes a revolution in clinical research where clinical trials of hundred of patients sized few hundred of Ko at most. Therefore, more than the storage and calculation capacities, the challenge is the comprehensive analysis of these datasets.

[1] see online at http://www.eurocoord.net

The objective is either to select the relevant information or to summarize it for understanding or prediction purposes. When dealing with high dimensional data, the methodological challenge arises from the fact that datasets typically contain many variables, much more than observations. Hence, multiple testing is an obvious issue that needs to be taken into account [34]. Furthermore, conventional methods, such as linear models, are inefficient and most of the time even inapplicable. Specific methods have been developed, often derived from the machine learning field, such as regularization methods [38]. The integrative analysis of large datasets is challenging. For instance, one may want to look at the correlation between two large scale matrices composed by the transcriptome in the one hand and the proteome on the other hand [27]. The comprehensive analysis of these large datasets concerning several levels from molecular pathways to clinical response of a population of patients needs specific approaches and a very close collaboration with the providers of data that is the immunologists, the virologists, the clinicians...

4. Application Domains

4.1. Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from "bench side to bedside") and vice versa, a process often called "Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [35], "Systems vaccinology" [32], "Systems medicine" [24]. From the statistician point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

4.2. The case of HIV immunology

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [28]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [23]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data

A work in collaboration with J. Skinner has been published in *PLoS Computational Biology* : [10]

Gene set analysis methods, which consider predefined groups of genes in the analysis of genomic data, have been successfully applied for analyzing gene expression data in cross- sectional studies. The time-course gene set analysis (TcGSA) introduced here is an exten- sion of gene set analysis to longitudinal data. The proposed method relies on random effects modeling with maximum likelihood estimates. It allows to use all available repeated measurements while dealing with unbalanced data due to missing at random (MAR) mea- surements. TcGSA is a hypothesis driven method that identifies a priori defined gene sets with significant expression variations over time, taking into account the potential heteroge- neity of expression within gene sets. When biological conditions are compared, the method indicates if the time patterns of gene sets significantly differ according to these conditions. The interest of the method is illustrated by its application to two real life datasets: an HIV therapeutic vaccine trial (DALIA-1 trial), and data from a recent study on influenza and pneumococcal vaccines. In the DALIA-1 trial TcGSA revealed a significant change in gene expression over time within 69 gene sets during vaccination, while a standard univariate individual gene analysis corrected for multiple testing as well as a standard a Gene Set Enrichment Analysis (GSEA) for time series both failed to detect any significant pattern change over time. When applied to the second illustrative data set, TcGSA allowed the identification of 4 gene sets finally found to be linked with the influenza vaccine too although they were found to be associated to the pneumococcal vaccine only in previous analyses. In our simulation study TcGSA exhibits good statistical properties, and an increased power compared to other approaches for analyzing time-course expression patterns of gene sets. The method is made available for the community through an R package.

5.1.2. Two new books

DC is co-editor and RT is co-author of the two following books :

- Daniel Commenges and Héléne Jacqmin-Gadda (2015), Dynamical Biostatistical Models, Chapman & Hall.
- Daniel Commenges and Hélène Jacqmin-Gadda (2015), Modèles biostatistiques pour l'épidémiologie. De Boeck.

6. New Software and Platforms

6.1. New Software

6.1.1. sgPLS

Sparse Group Partial Least Square
KEYWORD: Bioinformatics
FUNCTIONAL DESCRIPTION

The Sparse Group Partial Least Square package (sgPLS) provides sparse, group, and sparse group versions of partial least square regression models.

- Contact: Benoît Liquet
- URL: https://cran.r-project.org/web/packages/sgPLS/index.html

6.2. Upgraded Software

6.2.1. TcGSA

Time-course Gene Set Analysis
KEYWORDS: Bioinformatics - Genomics
FUNCTIONAL DESCRIPTION

An R package for the gene set analysis of longitudinal gene expression data sets. Available on the CRAN website, this package implements a Time-course Gene Set Analysis method and provides useful plotting functions facilitating the interpretation of the results.

- Contact: Boris Hejblum
- URL: https://cran.r-project.org/web/packages/TcGSA/index.html

6.2.2. VSURF

Variable Selection Using Random Forests
KEYWORD: Bioinformatics
FUNCTIONAL DESCRIPTION

An R package for Variable Selection Using Random Forests. Available on CRAN, this package performs an automatic (meaning completely data-driven) variable selection procedure. Originally designed to deal with high dimensional data, it can also be applied to standard datasets.

- Contact: Robin Genuer
- URL: http://cran.r-project.org/web/packages/VSURF/index.html

7. New Results

7.1. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data

The application of TcGSA methodology has revealed the commitment of inflammatory pathways and T-cell pathway in response of DC-based vaccine.

8. Partnerships and Cooperations

8.1. Regional Initiatives

The team have strong links with Bordeaux CHU ("Centre Hospitalier Universitaire").

8.2. National Initiatives

8.2.1. Labex Vaccine Research Institute (VRI)

There are strong collaborations with immunologists involved in the Labex Vaccine Research Institute (VRI) as RT is leading the Biostatistics/Bioinformatics division.

8.2.2. Expert Appraisals

Expertise of a project for Institut de Recherche en Santé Publique IRESP (MA)

RT is a member of the scientific advisory board of the Ebola VSV ring trial (published in New England Journal of Medicine in August 2015) and of the Pierre Louis Institute of Epidemiology and Public Health (Paris)

RT is a member of the Systems biology and cancer comity (Plan Cancer)

8.2.3. Partnership with the french swimming federation

Convention between the "Fédération francaise de natation" and Inria (18950 euros) for the R&D project "Quels schémas de périodisation pour la préparation des Jeux Olympiques à Rio ?"

8.2.4. Partnership with ANSM

DRUGS-SAFE platform funded by ANSM.

8.3. European Initiatives

8.3.1. Collaborations in European Programs, except FP7 & H2020

Program: The EBOVAC2 project is one of 8 projects funded under IMI Ebola+ programme that was launched in response to the Ebola virus disease outbreak. The project aims to assess the safety and efficacy of a novel prime boost preventive vaccine regimen against Ebola Virus Disease (EVD).

Project acronym: EBOVAC2

Project title: EBOVAC2

Coordinator: Rdolphe Thiébaut

Other partners: Inserm (France), Labex VRI (France), Janssen Pharmaceutical Companies of Johnson & Johnson, London School of Hygiene & Tropical Medicine (United Kingdom), The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom), Le Centre Muraz (Burknia Faso), Inserm Transfert (France)

Abstract: Given the urgent need for an preventive Ebola vaccine strategy in the context of the current epidemic, the clinical development plan follows an expedited scheme, aiming at starting a Phase 2B large scale safety and immunogenicity study as soon as possible while assuring the safety of the trial participants.

Phase 1 trials to assess the safety and immunogenicity data of the candidate prime-boost regimen in healthy volunteers are ongoing in the UK, the US and Kenya and Uganda. A further study site has been approved to start in Tanzania. Both prime-boost combinations (Ad26.ZEBOV prime + MVA-BN-Filo boost; and MVA-BN-Filo prime + Ad26.ZEBOV boost) administered at different intervals are being tested in these trials.

Phase 2 trials (this project) are planned to start as soon as the post-prime safety and immunogenicity data from the UK Phase I are available. Phase 2 trials will be conducted in healthy volunteers in Europe (France and UK) and non-epidemic African countries (to be determined). HIV positive adults will also be vaccinated in African countries. The rationale for inclusion of European volunteers in Phase 2, in addition to the trials in Africa, is to allow for higher sensitivity in safety signal detection in populations with low incidence of febrile illnesses, to generate negative control specimens for assay development, to allow for inclusion of health care workers or military personnel that may be deployed to Ebola-endemic regions.

8.4. International Initiatives

8.4.1. Participation In International Programs

RT is participating to the EUROCOORD network on HIV cohort collaborations as :

a member of the scientific committee of IWHOD International Workshop on HIV Observational Databases from 2013,

a project leader on defining references for the CD4 count response to antiretrovirals.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

David Conesa (Associate Professor of Biostatistics, "Spatial and Temporal Statistics in Epidemiology and Environment" Research Group, Universitat de Val encia, Spain) visited the team through the Erasmus+ program.

Following the RHOMEO project (ANR-BBSRC Systems biology 2007 call, 2007-2011) steered by RT, a strong collaboration has been established with Pr Robin Callard (UCL Immunology) who is visiting the team in Bordeaux one month each year, Andy Yates (Physicists, Glasgow Univ) and Ben Seddon (NIMR, UCL Immunology).

Also, several other international collaboration have been initiated through the Labex:

Raphael Gottardo, Zoe Moodie, Steve Self in Seattle (HVTN HIV vaccine Trial Network, Fred Hutchinson cancer centre)

Marcus Altfeld (Immunologists, Hambourg & Harvard).

8.5.2. Visits to International Teams

8.5.2.1. Sabbatical programme

BL was on sabbatical in Queensland University, Australia until Sep 2015.

8.5.2.2. Research stays abroad

Chloé Pasin visited (from 11/10/14 to 10/04/15) Steve Self at HVTN, Seattle.

Chariff Alkhassim (from 07/04/15 to 10/04/15) visited François Caron at Oxford University, United-Kingdom.

MA (from 24/02/15 to 10/03/15) and Perrine Soret (from 26/12/15 to 28/01/16) visited Cristian Meza and Karine Bertin (Inria Chili) at CIMFAV (Centre for Research and Modeling of Random Phenomena – Valparaíso), Univ Valparaiso, Chili, concerning the project "New challenges in mixed-effects models".

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

The practical part of "Atelier INSERM", *Big Data in clinical research*, was organized at Bordeaux by the team in Oct 2015.

9.1.1.1. Member of the organizing committee

BMW (Bordeaux Modeling Workshop), a two days workshop was organized (with 30 participants).

8th French Clinical Epidemiology Conference EPICLIN

Colloque Francophone International sur l'Enseignement de la Statistique - CFIES'2015, Janv 2015 Bordeaux

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committee

RT is a member of the scientific committee of IWHOD International Workshop on HIV Observational Databases from 2013,

9.1.3. Journal

9.1.3.1. Member of the editorial board

Lifetime Data Analysis (DC)

Stat Surveys (DC)

Journal de la Société Française de Statistique (DC)

9.1.3.2. Reviewer

The members of the team reviewed numerous papers for the following international journals :

AIDS (RT)

Biometrical (BL)

Biometrics (DC)

Health Services and Outcome Methodology (DC)

International Jounral of Epidemiology (DC)

Journal of Applied Statistics (MA)

Journal of Multivariate Analysis (RG)

Journal of the Royal Statistical Society: Series A (DC)

Statistical Methods and Applications (MA)

Statistics in Medicine (DC, RT)

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master : MA teaches in the two years of the Master of Public Health at ISPED, Univ. Bordeaux, France. Furthermore, she is head of the first year of the master.

Master : DC, teaches occasionnaly in the Biostatistics specialty of the second year of the Master of Public Health.

Master : RG, teaches in the two years of the Master of Public Health.

BL teaches at the School of Mathematics and Physics (The University of Queensland, Australia).

Master : RT, teaches in the two years of the Master of Public Health, and he is head of the Epidemiology specialty of the second year of the Master of Public Health.

E-learning

MA is head of the first year of the e-learning program of the Master of Public Health, and teaches in it.

RG teaches in the first year of the e-learning program of the Master of Public Health.

RT is head of the Epidemiology specialty of the second year of the e-learning program of the Master of Public Health, and teaches in it.

RG and Perrine Soret participate to the IdEx Bordeaux University "Défi numérique" project "Begin'R".

9.2.2. Supervision

PhD defended in Dec 2015 : Ana Jarne, *Modélisation de la réponse à l'IL-7*, co-directed by Daniel Commenges & Rodolphe Thiébaut

PhD defended in Mar 2015 : Boris Hejblum, *Analyse integrative de données de grande dimension appliquée à la recherche vaccinale*, co-directed by Rodolphe Thiébaut & François Caron

PhD defended in Oct 2015 : Marie-Quitterie Picat, *Méthodes pour l'analyse intégrative des marqueurs immunologiques*, directed by Rodolphe Thiébaut

PhD in progress : Perrine Soret, *Modélisation de données longitudinales en grande dimension*, from Oct 2014, directed by Marta Avalos

PhD in progress : Chloé Pasin, *Modelling the immune response to HIV vaccine*, from Sep 2015, co-directed by Rodolphe Thiébaut and François Dufour

PhD in progress : Laura Villain, *Analysing and modeling the effect of interleukin 7 in HIV-infected patients*, from Sep 2015, co-directed by Rodolphe Thiébaut and Daniel Commenges

PhD in progress : Mélanie Née *Recherche et caractérisation de profils attentionnels : mieux comprendre la place de l'attention dans la survenue des accidents de la vie courante*, from Oct 2015, co-directed by Emmanuel Lagarde (60%), Cédric Galera (20%), Marta Avalos (20%)

Master internship : Chariff Alkhassim, *Reconnaissance automatique de populations cellulaires à l'aide de processus de Dirichlet*, from Feb 2015 to Sep 2015, co-directed by Rodolphe Thiébaut & François Caron

Master internship : Edouard Lhomme, *Analyse de la réponse immunologique au vaccin Ad5 dans un essai americain (HVTN 068)*, from Feb 2015 to Sep 2015, directed by Rodolphe Thiébaut

Master internship : Marina Travanca *Prédiction des accidents de la vie courante à partir de facteurs environnementaux et comportementaux : comparaison de méthodes d'apprentissage statistique adaptées aux données de l'observatoire MAVIE* from Apr 2015 to Jun 2015, co-directed by Marta Avalos and Ludivine Orriols

Master internship : Gaëlle Lefort *Développement d'un outil statistique d'aide à la décision pour l'organisation de l'entraînement chez des sportifs de haut niveau*, directed by Marta Avalos

Master internship : Emilie Chanfreau, *Etude de l'élagage dans la méthode des forêts aléatoires*, from Apr 2015 to Jun 2015, directed by Robin Genuer

Master internship : Maëva Kyeng, *ETUDE DE L'EVOLUTION DE LA CHARGE VIRALE CHEZ LES PATIENTS ATTEINTS Du VIH : COMPARAISON DE DEUX METHODES STATISTIQUES*, from Apr 2015 to Jun 2015, directed by Perrine Soret

9.2.3. Juries

Members of the team were involved in 6 PhD juries, 2 professorships and 2 HDR.

9.3. Popularization

MA and Perrine Soret animate "Les maths sont bonnes pour la santé" for high school students through the "Fête de la Science" organized at Inria, Oct 2015.

10. Bibliography

Major publications by the team in recent years

[1] M. AVALOS, N. D. ADROHER, E. LAGARDE, F. THIESSARD, Y. GRANDVALET, B. CONTRAND, L. ORRIOLS. *Prescription-Drug-Related Risk in Driving: Comparing Conventional and Lasso Shrinkage Logistic Regressions*, in "Epidemiology", 2012, vol. 23, n^{o} 5, pp. 706–712

[2] D. FURMAN, P. HEJBLUM, N. SIMON, V. JOJIC, L. DEKKER, R. THIÉBAUT, J. TIBSHIRANI, M. DAVIS. *Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination*, in "Proceedings of the National Academy of Sciences", January 2014, vol. 111, n^{o} 2, pp. 869-74, http://dx.doi.org/10.1073/pnas.1321060111

[3] B. LIQUET, K.-A. LE CAO, H. HOCINI, R. THIÉBAUT. *A novel approach for biomarker selection and the integration of repeated measures experiments from two assays*, in "BMC bioinformatics", 2012, vol. 13, n^{o} 1, 14 p. , http://dx.doi.org/10.1186/1471-2105-13-325

[4] M. PRAGUE, D. COMMENGES, J. DRYLEWICZ, R. THIÉBAUT. *Treatment monitoring of HIV infected patients based on mechanistic models*, in "Biometrics", 2012, vol. 68, n^{o} 3, pp. 902–911

[5] R. THIÉBAUT, J. DRYLEWICZ, M. PRAGUE, C. LACABARATZ, S. BEQ, A. JARNE, T. CROUGHS, R.-P. SEKALY, M. M. LEDERMAN, I. SERETI. *Quantifying and Predicting the Effect of Exogenous Interleukin-7 on CD4+ T Cells in HIV-1 Infection*, in "PLoS computational biology", 2014, vol. 10, n^{o} 5, e1003630

Publications of the year

Articles in International Peer-Reviewed Journals

[6] M. AVALOS, H. POUYES, Y. GRANDVALET, L. ORRIOLS, E. LAGARDE. *Sparse conditional logistic regression for analyzing large-scale matched data from epidemiological studies: a simple algorithm*, in "BMC Bioinformatics", 2015, vol. 16, n^{o} Suppl 6, S1 p. [*DOI :* 10.1186/1471-2105-16-S6-S1], https://hal.inria.fr/hal-01217312

[7] H. AYOUB, B. AINSEBA, M. LANGLAIS, R. THIÉBAUT. *Parameters identification for a model of T cell homeostasis*, in "Mathematical Biosciences and Engineering", 2015, vol. 12, n^{o} 5, pp. 917–936 [*DOI :* 10.3934/MBE.2015.12.917], https://hal.archives-ouvertes.fr/hal-01164668

[8] D. COMMENGES, A. GÉGOUT-PETIT. *The stochastic system approach for estimating dynamic treatments effect*, in "Lifetime Data Analysis", 2015, 18 p. [*DOI :* 10.1007/s10985-015-9322-3], https://hal.inria.fr/hal-01205328

[9] R. GENUER, J.-M. POGGI, C. TULEAU-MALOT. *VSURF: An R Package for Variable Selection Using Random Forests*, in "The R Journal", December 2015, vol. 7, n° 2, pp. 19-33, https://hal.archives-ouvertes.fr/hal-01251924

[10] B. P. HEJBLUM, J. SKINNER, R. THIÉBAUT. *Time-Course Gene Set Analysis for Longitudinal Gene Expression Data*, in "PLoS Computational Biology", 2015, vol. 11, n° 6, e1004310 [*DOI :* 10.1371/JOURNAL.PCBI.1004310], https://hal.inria.fr/hal-01203446

[11] P. HELLARD, M. AVALOS, F. GUIMARAES, J. F. TOUSSAINT, P. DAVID. *Training-Related Risk of Common Illnesses in Elite Swimmers over a Four-Year Period*, in "Medicine and Science in Sports and Exercise", 2015, vol. 47, n° 4, pp. 698-707 [*DOI :* 10.1249/MSS.0000000000000461], https://hal.archives-ouvertes.fr/hal-01099379

[12] H. KAMINSKI, I. GARRIGUE, L. COUZI, B. TATON, T. BACHELET, J.-F. MOREAU, J. DECHANET-MERVILLE, R. THIÉBAUT, P. MERVILLE. *Surveillance of $\gamma\delta$ T Cells Predicts Cytomegalovirus Infection Resolution in Kidney Transplants*, in "Journal of the American Society of Nephrology : JASN", 2015, n° 8, 00 p. [*DOI :* 10.1681/ASN.2014100985], https://hal.archives-ouvertes.fr/hal-01164667

[13] L. RICHERT, E. LHOMME, C. FAGARD, Y. LEVY, G. CHÊNE, R. THIÉBAUT. *Recent developments in clinical trial designs for HIV vaccine research*, in "Human vaccines & immunotherapeutics", 2015, vol. 11, n° 4, pp. 1022–9 [*DOI :* 10.1080/21645515.2015.1011974], https://hal.archives-ouvertes.fr/hal-01164644

International Conferences with Proceedings

[14] M. LE GOFF, M. F. AVALOS, P. JOLY, M.-A. JUTAND. *Evolution des stratégies pédagogiques d'un DU*, in "Colloque Francophone International sur l'Enseignement de la Statistique – CFIES'2015", Bordeaux, France, January 2015, https://hal.inria.fr/hal-01253156

[15] M. NÉE, M. F. AVALOS, L. ORRIOLS, E. LAGARDE. *Impact of unmeasured covariates on bias and statistical power in health administrative databases: a simulation study*, in "XVth Spanish Biometric Conference and the Vth Ibéro-American Biometric Meeting 2015", Bilbao, Spain, 2015, https://hal.inria.fr/hal-01253141

[16] P. SORET, C. MEZA, K. BERTIN, M. F. AVALOS, P. HELLARD. *Function selection in mixed models using L1-penalization*, in "XVth Spanish Biometric Conference and the Vth Ibéro-American Biometric Meeting 2015", Bilbao, Spain, 2015, https://hal.inria.fr/hal-01252267

National Conferences with Proceedings

[17] A. TODESCHINI, R. GENUER. *Compétitions d'apprentissage automatique avec le package R rchallenge*, in "47èmes Journées de Statistique de la SFdS", Lille, France, Société Française de Statistique, June 2015, https://hal.inria.fr/hal-01157147

Conferences without Proceedings

[18] R. GENUER, J.-M. POGGI, C. TULEAU-MALOT, N. VILLA-VIALANEIX. *Random forests and big data*, in "47ème Journées de Statistique de la SFdS", Lille, France, Société Française de Statistique, June 2015, https://hal.archives-ouvertes.fr/hal-01160643

Other Publications

[19] R. GENUER, J.-M. POGGI, C. TULEAU-MALOT, N. VILLA-VIALANEIX. *Random Forests for Big Data*, November 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01233923

[20] P. SORET, M. F. AVALOS, R. THIÉBAUT. *High-dimensional longitudinal genomic data: a survey and evaluation of publicly available implementations of machine learning methods*, November 2015, Statistical Analysis of Massive Genomic Data, Poster, https://hal.inria.fr/hal-01253151

References in notes

[21] O. AALEN, K. KJETIL ROYSLAND, J. GRAN, B. LEDERGERBER. *Causality, mediation and time: a dynamic viewpoint*, in "Journal of the Royal Statistical Society: Series A (Statistics in Society)", 2007, vol. 175, n^{o} 4, pp. 831–861

[22] F. CASTIGLIONE, B. PICCOLI. *Cancer immunotherapy, mathematical modeling and optimal control*, in "Biometrical Journal", 2007, vol. 247, n^{o} 4, pp. 723-32

[23] R. M. GRANICH, C. F. GILKS, C. DYE, K. M. DE COCK, B. G. WILLIAMS. *Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model*, in "Lancet", 2009, vol. 373, n^{o} 9657, pp. 48-57, 0140 6736 English

[24] L. HOOD, Q. TIAN. *Systems approaches to biology and disease enable translational systems medicine*, in "Genomics Proteomics Bioinformatics", 2012, vol. 10, n^{o} 4, pp. 181–5

[25] Y. HUANG, D. LIU, H. WU. *Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system*, in "Biometrics", 2006, vol. 62, n^{o} 2, pp. 413–423

[26] E. KUHN, M. LAVIELLE. *Maximum likelihood estimation in nonlinear mixed effects models*, in "Computational Statistics & Data Analysis", 2005, vol. 49, n^{o} 4, pp. 1020–1038

[27] K.-A. LE CAO, P. MARTIN, C. ROBERT-GRANIÉ, P. BESSE. *Sparse canonical methods for biological data integration: application to a cross-platform study*, in "BMC bioinformatics", 2009, vol. 10, 34 p.

[28] C. LEWDEN, D. SALMON, P. MORLAT, S. BEVILACQUA, E. JOUGLA, F. BONNET, L. HERIPRET, D. COSTAGLIOLA, T. MAY, G. CHÊNE. *Causes of death among human immunodeficiency virus (HIV)-infected adults in the era of potent antiretroviral therapy: emerging role of hepatitis and cancers, persistent role of AIDS*, in "International Journal of Epidemiology", 2005, vol. 34, n^{o} 1, pp. 121-130, 0300 5771 English

[29] A. S. PERELSON, A. U. NEUMANN, M. MARKOWITZ, J. M. LEONARD, D. D. HO. *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time*, in "Science", 1996, vol. 271, n^{o} 5255, pp. 1582-6

[30] A. S. PERELSON. *Modelling viral and immune system dynamics*, in "Nature Reviews Immunology", 2002, vol. 2, n^{o} 1, pp. 28-36

[31] J. PINHEIRO, D. BATES. *Approximations to the log-likelihood function in the nonlinear mixed-effects model*, in "Journal of Computational and Graphical Statistics", 1995, vol. 4, n^{o} 1, pp. 12–35

[32] B. PULENDRAN. *Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology*, in "Nature Reviews Immunology", 2009, vol. 9, n^{o} 10, pp. 741-7

[33] H. PUTTER, S. HEISTERKAMP, J. LANGE, F. DE WOLF. *A Bayesian approach to parameter estimation in HIV dynamical models*, in "Statistics in Medicine", 2002, vol. 21, n^{o} 15, pp. 2199–2214

[34] A. REINER, D. YEKUTIELI, Y. BENJAMINI. *Identifying differentially expressed genes using false discovery rate controlling procedures*, in "Bioinformatics", 2003, vol. 19, n^{o} 3, pp. 368–375

[35] C. SCHUBERT. *Systems immunology: complexity captured*, in "Nature", 2011, vol. 473, n^{o} 7345, pp. 113-4

[36] R. THIÉBAUT, B. HEJBLUM, L. RICHERT. *[The analysis of "Big Data" in clinical research.]*, in "Epidemiology and Public Health / Revue d'Epidémiologie et de Santé Publique", January 2014, vol. 62, n^{o} 1, pp. 1–4 [*DOI :* 10.1016/J.RESPE.2013.12.021], http://www.hal.inserm.fr/inserm-00933691

[37] R. THIÉBAUT, H. JACQMIN-GADDA, A. BABIKER, D. COMMENGES. *Joint modelling of bivariate longitudinal data with informative dropout and left-censoring, with application to the evolution of CD4+cell count and HIV RNA viral load in response to treatment of HIV infection*, in "Statistics in Medicine", 2005, vol. 24, n^{o} 1, pp. 65-82

[38] R. TIBSHIRANI. *Regression shrinkage and selection via the lasso*, in "Journal of the Royal Statistical Society: Series B (Statistical Methodology)", 1996, vol. 58, pp. 267–288

[39] Y. WANG. *Derivation of various NONMEM estimation methods*, in "Journal of Pharmacokinetics and pharmacodynamics", 2007, vol. 34, n^{o} 5, pp. 575–593

[40] H. WU. *Statistical methods for HIV dynamic studies in AIDS clinical trials*, in "Statistical Methods in Medical Research", 2005, vol. 14, n^{o} 2, pp. 171–192