

**Guillem Rigail**

PhD. in Applied Mathematics

Computational Statistics for Biology

IPS2

Bat 630 Rue Noetzlin

91190 Gif-sur-Yvette

firstname.lastname@inrae.fr

LAMME

23 bd de France

91037 Évry Cedex

## CURRENT POSITION

---

Avr. 2017

Researcher at INRAE (CR) working in two laboratories

Institut des Sciences des Plantes de Paris-Saclay (IPS2), Genomics Networks Team

Laboratoire de Mathématiques et Modélisation d'Évry, Stat. & Genome Team

## RESEARCH THEMES

---

My research focuses on the development of biostatistical models, statistical methods and algorithms for the analysis and interpretation of high throughput biological data.

### **Modeling and inferring dependence in big data settings**

— Models : multiple changepoints detection, classification, regression

— Computational statistics, Dynamic Programming, Convex relaxation, Penalized Likelihood

### **Applications in molecular biology and bioinformatics**

— DNA copy number analysis, RNA-seq, DNA-seq, Chip-seq, Gene networks inference

— Practical evaluation of biostatistical tools, R packages

### **Publications**

— 42 in statistical, bioinformatical and biology peer reviewed journals

— 2 in the proceedings of ICML

— H-index of 21 according to Google Scholar

**Grant** My research is funded by an ATIGE Grant from Genopole (2017-2021)

## EDUCATION

---

18 sept 2020

**Habilitation à diriger des recherches.** UNI. PARIS-SACLAY.

In french : « Quelques développements statistiques et algorithmiques pour l'analyse de données génomiques. »

*Rapporteurs*

A-L BOULESTEIX (IBE), D. CAUSEUR (AgroCampus Ouest), S. SCHBATH (INRAE)

*Examineurs*

S. ARLOT (UPSAY), A. BAR-HEN (CNAM), J-P. VERT (Google Brain)

2007-2010

**PhD. in applied mathematics.** AGROPARISTECH.

Statistical and algorithmic developments for the analysis of Triple Negative Breast Cancers.

*Rapporteurs*

G. MILLS (MD Anderson Cancer Center, TX US), N. ZHANG (Stanford Uni., CA US)

*Examineurs*

A. ANTONIADIS (Uni. Fourier), T. SCHIEX (INRA), L. WESSELS (NKI-AVL, Pays-Bas)

2005-2007

**Bioinformatics Master,** Université de Rouen

2002-2005

**Ingénieur Agronome,** AGROPARISTECH

## SCIENTIFIC EXPERIENCE

---

- 2012-2017     **Assistant Professor**  
Evry University (UEVE), Excellence Research Chair from INRA.  
Institut des Sciences des Plantes de Paris-Saclay (IPS2)  
Statistical and algorithmic developments for the analysis of high throughput biological data.
- 2011           **Post-doc**, NKI-AVL, Amsterdam, Pays-Bas  
Bioinformatics and Statistics group, supervised by L. WESSELS  
Statistical and algorithmic developments for the analysis of capture DNA-seq data.
- 2007-2010     **PhD. Student**  
Institut Curie, Translational Department  
U900 INSERM/Institut Curie/Mines ParisTech  
Unité MIA Paris AGROPARISTECH/INRA.  
Statistical and algorithmic developments for the analysis of Triple Negative Breast Cancers.

## SOME RECENT PUBLICATIONS

---

- 2020           Fast tree aggregation for consensus hierarchical clustering. A. Hulot, J. Chiquet, F. Jaffrézic, and G. Rigai.  
**BMC bioinformatics**
- 2020           A log-linear time algorithm for constrained changepoint detection. T. Hocking, G. Rigai, P. Fearnhead, and G. Bourque.  
**Journal of Machine Learning Research**
- 2019           New efficient algorithms for multiple change-point detection with reproducing kernels. A. Celisse, G. Marot, M. Pierre-Jean, and G. Rigai.  
**Computational Statistics and Data Analysis**
- 2018           Changepoint detection in the presence of outliers. P. Fearnhead and G. Rigai.  
**Journal of the American Statistical Association**
- 2016           Synthetic data sets for the identification of key ingredients for rna-seq differential analysis. G. Rigai, S. Balzergue, V. Brunaud, E. Blondet, A. Rau, O. Rogier, J. Caius, C. Maugis-Rabusseau, L. Soubigou-Taconnat, S. Aubourg, L. Claire, M.L. Martin-Magniette, and E. Delannoy.  
**Briefings in Bioinformatics**
- 2015           Fast tree inference with weighted fusion penalties. J. Chiquet, P. Gutierrez, and G. Rigai.  
**Journal of Computational and Graphical Statistics**

## SCIENTIFIC EVENTS

---

In 2019, I co-organized a workshop on **Changepoint and anomaly detection in big data settings** in Paris, with Idris Eckley, Paul Fearnhead and Stéphane Robin : <https://www.statscale.org/changepoint-workshop-2019>.

I am part of the organisation comitee of **NetBIO** : <https://mia.toulouse.inra.fr/NETBIO>. The goal of this work group is to create a space to discuss statistical and bioinformatic methods for the construction and analysis of graphs and discuss the underlying mathematical or computationnal models and their adequacy to adress biological questions.

I am co-organizing the Math For Genomics seminar with C. Ambroise and V. Runge : <https://mathforgenomics.github.io/projet/>

## SUPERVISIONS

---

- 2017-2019 Vincent RUNGE, **Post-doc.**  
Computational Statistics for changepoints detection.
- 2017-today Martina SUNDQUIST, **PhD.**  
Classification of Triple Negative Breast Cancers.  
25% with J. CHIQUET and T. DUBOIS
- 2013-2016 Trung HA, **PhD.**  
Multivariate analysis for robust regulatory network inference.  
25% with M-L. MARTIN-MAGNETTE and J. CHIQUET
- 2016 Mengliang YE, **IE**  
Transcriptomic analysis of breast cancers.  
25% with T. DUBOIS
- 2018 Arnaud LIEHRMANN, M1 internship (3 months)  
Segmentation of RNA-seq profiles, taking into account the length of segments  
75% with V. RUNGE
- 2017 Pauline FOURGOUX, M2 intership (6 months)  
Analysis of PPR footprinting data.  
50% with V. BRUNAUD
- 2017 Martina SUNDQUIST, M2 intership (6 months)  
Classification of Triple Negative Breast Cancers.  
25% with J. CHIQUET, L. DE KONING and T. DUBOIS
- 2016 Audrey HULOT, M2 intership (6 months)  
Clustering and multivariate analysis of metagenomic data  
25% with J. CHIQUET and F. JAFFRÉZIC
- 2015 Valentin DERVIEUX, M1 internship (4 months)  
Clustering and multivariate analysis of metagenomic data.  
50% with J. CHIQUET
- 2013 Pierre GUTIERREZ, M2 intership (6 months)  
Multi-class differential analysis with fused-Anova.  
50% with J. CHIQUET

## TEACHINGS

---

From 2012 to 2018, I taught approximately 64 hours per year at Université d'Évry, Université Paris-Saclay and École Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise (ENSIIE). I mostly taught statistics (hypothesis testing, regression and graphical models) to students in biology, bioinformatics and data-science.

In 2016, I was a member of the scientific committee of the Summer School : "From gene expression to genomic network" organised by Saclay Plant Sciences ([click here to go to the webpage](#)).

In 2018, I taught statistics at the GDRBim Summer School (JC2BIM'2018).

Since 2016, I am teaching statistics at the Genopole Summer School with Christophe Ambroise. ([click here to go to the webpage](#))

Since 2014, I am in charge of a module untitled System Biology 1 for Master 1 Students in Biology at Université Paris-Saclay.