

Andrea Rau

INRAE
GABI (GiBBS team)
Domaine de Vilvert
78350 Jouy-en-Josas, France

phone : (+33) (0) 1 34 65 22 82
email : andrea.rau@inrae.fr
Twitter : @andreamrau
<https://andrea-rau.com>

INRAE
BioEcoAgro
2 Chaussée Brunehaut
50136 Estrées-Mons, France

phone : (+33) (0) 3 22 85 76 90

Current position

- Research Scientist (*Chargée de Recherche*) at INRAE

Education

- 2017 **HDR* in Applied Mathematics**, Université d'Évry-Val-d'Essonne (Évry, France)
obtained September 26, 2017.
Title : Statistical methods and software for the analysis of transcriptomic data.
* An HDR is the French accreditation to supervise research and represents the highest French academic qualification level based on independent scholarship. It is reviewed by and defended before an academic committee.
- 2007–2010 **Ph.D. in Statistics**, Purdue University (West Lafayette, Indiana, USA)
obtained August 6, 2010.
Title : Reverse engineering gene regulatory networks using genomic time-course data.
Advisors : Rebecca W. Doerge, Jean-Louis Foulley, and Florence Jaffrézic
- 2005–2007 **M.S. (Master of Science) in Applied Statistics**, Purdue University (West Lafayette, Indiana, USA).
Internship : Using time series and intervention models to analyze the effect of advertising changes on store sales for the pharmacy chain Walgreens (Deerfield, Illinois, USA)
- 2001–2005 **B.A. (Bachelor of Arts) in French and B.A. in Mathematics**,
concentration in Statistics (with distinction), Saint Olaf College (Northfield, Minnesota, USA).
Internship : Application of pharmacokinetic analysis literature to data from a Phase I clinical trial and determination of a limited-sample model, under the direction of Alfred Furth at the Mayo Clinic (Rochester, Minnesota, USA)

Work experience

2016–present	Research Scientist (<i>Chargée de Recherche</i>), INRA (Jouy-en-Josas, France)
2019	Adjunct Assistant Professor (4 months), Medical College of Wisconsin Milwaukee, Wisconsin, USA
2017–2019	AgreenSkills+ Visiting Scholar (20 months), Zilber School of Public Health, University of Wisconsin-Milwaukee
2016	Visiting Scholar (6 weeks), Zilber School of Public Health, University of Wisconsin-Milwaukee
2012–2017	Adjunct Assistant Professor , Ensai (Rennes, France)
2011–2016	Junior Research Scientist (<i>Chargée de Recherche</i> 2nd class, CR2), INRA (Jouy-en-Josas, France)
2010–2011	Post-doctoral researcher , Inria - Île-de-France (Orsay, France) <i>Subject</i> : Clustering RNA sequencing data. <i>Advisors</i> : Gilles Celeux and Marie-Laure Martin-Magniette
2008–2010	Research assistant for R. W. Doerge, Department of Statistics at Purdue University (West Lafayette, Indiana, USA)
2007	Consultant in the Statistical Consulting Service, Department of Statistics at Purdue University (West Lafayette, Indiana, USA)

Research subjects

Research themes : Analysis of high-throughput sequencing data (RNA-seq, ATAC-seq), mixture models, supervised classification methods, multi-omic integration, gene regulatory networks

Areas of application : Genomic and transcriptomic data

Skills

<i>Computing</i>	Operating systems	Linux, Windows
	Languages	C, HTML
	Programs	R, \LaTeX
<i>Languages</i>	English	Maternal language
	French	Fluent (read, written, spoken)

Awards

• Graduate Women in Science Programs travel award	2010
• Student travel award, Conference on Applied Statistics in Agriculture at Kansas State University	2010
• VIGRE Fellowship, Purdue University	2005–2010
• Honorable mention, Gertrude M. Cox Scholarship	2009
• Graduate student scholarship, Gordon Conference on Quantitative Genetics	2009

and Genomics

- BioC Conference graduate student scholarship 2009
- A.H. Ismail Interdisciplinary Program doctoral research travel award 2009
- Buntrock academic scholarship, Saint Olaf College 2001
- National Merit Scholar 2001
- Presidential Scholar 2001

Professional organizations

- Société Française de Statistique (SFdS, 2011-present)
- American Statistical Association (ASA, 2005-present)

Books

1. Albert, I., Ancelet, S., David, O., Denis, J.-B., Makowski, D., Parent, É., **Rau, A.**, and Soubeyrand, S. (2015). *Initiation à la statistique bayésienne : Bases théoriques et applications en alimentation, environnmenet, épidémiologie et génétique* : Éditions Ellipses, collection références sciences.

Peer-reviewed articles

2020

1. Sellem, E., Marthey, S., **Rau, A.**, Jouneau, L., Bonnet, A., Perrier, J.-P., Fritz, S., Le Danvic, C. Boussaha, M., Kiefer, H., Jammes, H., Schiblier, L. (2020) A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity in six breeds. *Epigenetics and Chromatin* 13 :19. doi : 10.1186/s13072-020-00340-0.
 2. **Rau, A.**, Manansala, R., Flister, M. J., Rui, H., Jaffrézic, F., Laloë, D.*, and Auer, P. L.* (2020) Individualized multi-omic pathway deviation scores using multiple factor analysis. *Biostatistics* (accepted), <https://doi.org/10.1101/827022>.
- * These authors contributed equally to this work.
3. Godichon-Baggioni, A., Maugis-Rabusseau, C. and **Rau, A.** (2020) Multi-view cluster aggregation and splitting, with an application to multi-omic breast cancer data. *Annals of Applied Statistics*, 14 :2, 752-767.

2019

4. Jehl, F., Désert, C., Klopp, C., Brenet, M., **Rau, A.**, Leroux, S., Boutin, M., Muret, K., Blum, Y., Esquerré, D., Gourichon, D., Burlot, T., Collin, A., Pitel, F., Benani, A., Zerjal, T., Lagarrigue, S. (2019) Chicken adaptive response to low energy diet : main role of the hypothalamic lipid metabolism revealed by a phenotypic and multi-tissue transcriptomic approach. *BMC Genomics*, 20. doi : 10.1186/s12864-019-6384-8.
5. Foissac, S., Djebali, S., Munyard, K., Villa-Vialaneix, N., **Rau, A.**, Muret, K., Esquerre, D., Zytnicki, M., Derrien, T., Bardou, P., Blanc, F., Cabau, C., Crisci, E., Dhorne-Pollet, S., Drouet, F., Gonzales, I., Goubil, A., Lacroix-Lamande, S., Laurent, F., Marthey, S., Marti-Marimon, M., Momal-Leisenring, R., Mompert, F., Quere, P., Robelin, D., San Cristobal, M., Tosser-Klopp, G., Vincent-Naulleau, S., Fabre, S., Pinard-Van der Laan, M.-H., Klopp, C., Tixier-Boichard, M., Acloque, H., Lagarrigue, S., Giuffra, E. (2019) Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biology* 17 : 108.

6. Dhara, S., **Rau, A.**, Flister, M., Recka, N., Laiosa, M., Auer, P., and Udvardia, A. (2019) Cellular reprogramming for successful CNS axon regeneration is driven by a temporally changing cast of transcription factors. *Scientific Reports* 9 :14198, doi : 10.1038/s41598-019-50485-6.
7. **Rau, A.**, Dhara, S., Udvardia, A., and Auer, P. (2019) Regeneration Rosetta : An interactive web application to explore regeneration-associated gene expression and chromatin accessibility. *G3 : Genes | Genomes | Genetics*, doi : 10.1534/g3.119.400729.
8. Plasterer, C., Tsaih, S.-W., Lemke, A., Schilling, R., Dwinell, M., **Rau, A.**, Auer, P., Rui, H., Flister, M.J. (2019) Identification of a rat mammary tumor risk locus that is syntenic with the commonly amplified 8q12.1 and 8q22.1 regions in human breast cancer patients. *G3 : Genes | Genomes | Genetics* 9(5) :1739-1743. doi : 10.1534/g3.118.200873.
9. Ramayo-Caldas, Y., Zingaretti, L., Bernard, A., EstellÃ©, J., Popova, M., Pons, N., Bellot, P., Mach, N., **Rau, A.**, Roume, H., Perez-Encisco, M., Faverdin, P., Edouard, N., Dusko, S., Morgavi, D.P. and Renand, G. (2019) Identification of rumen microbial biomarkers linked to methane emission in Holstein dairy cows. *Journal of Animal Breeding and Genetics*, doi : 10.1111/jbg.12427.
10. **Rau, A.**, Flister, M. J., Rui, H. and Livermore Auer, P. (2019) Exploring drivers of gene expression in The Cancer Genome Atlas. *Bioinformatics*, 35(1) :62-68.

2018

11. Godichon-Baggioni, A., Maugis-Rabusseau, C. and **Rau, A.** (2018) Clustering transformed compositional data using K-means, with applications in gene expression and bi-cycle sharing system data. *Journal of Applied Statistics*, <https://doi.org/10.1080/02664763.2018.1454894>.
12. **Rau, A.** and Maugis-Rabusseau, C. (2018) Transformation and model choice for RNA-seq co-expression analysis. *Briefings in Bioinformatics*, bbw128, <https://doi.org/10.1093/bib/bbw128>.
13. Verrier, E., Genet, C., Laloë, D., Jaffrézic, J., **Rau, A.**, Esquerré, D., Dechamp, N., Ciobartu, C., Hervet, C., Krieg, F., Quillet, E., Boudinot, P. (2018) Genetic and transcriptomic analyses provide new insights on the early antiviral response to VHSV in resistant and susceptible rainbow trout. *BMC Genomics*, 19 :482.
14. Maroille, T., Berri, M., Lemonnier, G., Esquerré, D., Chevalé, C., Mélo, S., Meurens, F., Coville, J.L., Leplat, J.J., **Rau, A.**, Bed'hom, B., Vincent-Naulleau, S., Mercat, M.J., Billon, Y., Lepage, P., Rogel-Gaillard, C., and Estellé, J. (2018). Immunome differences between porcine ileal and jejunal Peyer's patches revealed by global transcriptome sequencing of gut-associated lymphoid tissues. *Scientific Reports*, 8 :9077.
15. Mondet, F., **Rau, A.**, Klopp, C., Rohmer, M., Severac, D., Le Conte, Y., and Alaux, C. (2018). Transcriptome profiling of the honeybee parasite *Varroa destructor* provides new biological insights into the mite adult life cycle. *BMC Genomics*, 19 :328.
16. He, B., Tjhung, K., Bennett, N., Chou, Y., **Rau, A.**, Huang, J., and Derda, R. (2018). Compositional bias in naïve and chemically-modified phage-displayed libraries uncovered by paired-end deep sequencing. *Scientific Reports*, 8 :1214.

2017

17. Monneret, G., Jaffrézic, F., **Rau, A.**, Zerjal, T. and Nuel, G. (2017) Identification of marginal causal relationships in gene networks from observational and interventional expression data. *PLoS One* 12(3) : e0171142.

18. Sauvage, C., **Rau, A.**, Aichholz, C., Chadoeuf, J., Sarah, G., Ruiz, M., Santoni, S., Causse, M., David, J., Glémin, S. (2017) Domestication rewired gene expression and nucleotide diversity patterns in tomato. *The Plant Journal* 91(4) :631-645.

2016

19. Rigauill, G., Balzergue, S., Brunaud, V., Blondet, E., **Rau, A.**, Rogier, O., Caius, J., Maugis-Rabusseau, C., Soubigou-Taconnat, L., Aubourg, S., Lurin, C., Martin-Magniette, M.-L., and Delannoy, E. (2016) Synthetic datasets for the identification of key ingredients for RNA-seq differential analysis. *Briefings in Bioinformatics*, doi : <https://doi.org/10.1093/bib/bbw092>.

2015

20. Gallopin, M., Celeux, G., Jaffrézic, F., **Rau, A.** (2015) A model selection criterion for model-based clustering of annotated gene expression data. *Statistical Applications in Genetics and Molecular Biology*, 14(5) : 413-428.
21. Monneret, G., Jaffrézic, F., **Rau, A.**, Nuel, G. (2015). Estimation d'effets causaux dans les réseaux de régulation génique : vers la grande dimension. *Revue d'intelligence artificielle*, 29(2) : 205-227.
22. **Rau, A.**, Maugis-Rabusseau, C., Martin-Magniette, M.-L., Celeux, G. (2015) Co-expression analysis of high-throughput transcriptome sequencing data with Poisson mixture models. *Bioinformatics*, 31(9) : 1420-1427.

2014

23. **Rau, A.**, Marot, G., and Jaffrézic, F. (2014). Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinformatics* 15:91.
24. Nuel, G., **Rau, A.**, and Jaffrézic, F. (2014) Using pairwise ordering preferences to estimate causal effects in gene expression from a mixture of observational and intervention experiments. *Quality Technology and Quantitative Management* 11(1):23-37.
25. Endale Ahanda, M.-L., Zerjal, T., Dhorne-Pollet, S., **Rau, A.**, Cooksey, A., and Giuffra, E. (2014) Impact of the genetic background on the composition of the chicken plasma miRNome in response to a stress. *PLoS One*, 9(12) : e114598.
26. Brenault, P., Lefevre, L. **Rau, A.**, Laloë, D., Pisoni, G., Moroni, P., Bevilacqua, C. and Martin, P. (2014) Contribution of mammary epithelial cells to the immune response during early stages of a bacterial infection to *Staphylococcus aureus*. *Veterinary Research* 45:16.

2013

27. **Rau, A.**, Jaffrézic, F., and Nuel, G. (2013) Joint estimation of causal effects from observational and intervention gene expression data. *BMC Systems Biology* 7:111.
28. Gallopin, M. **Rau, A.**, and Jaffrézic, F. (2013). A hierarchical Poisson log-normal model for network inference from RNA sequencing data. *PLoS One* 8(10) : e77503.
29. **Rau, A.**, Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics* 29(17) : 2146-2152.
30. Dillies, M.-A.*, **Rau, A.***, Aubert, J.*, Hennequet-Antier, C.*, Jeanmougin, M.*, Servant, N.*, Keime, C.*, Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Charif, D., Le Crom, S.*, Guedj, M.*, and Jaffrézic, F*. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics* 14(6) : 671-683.

* These authors contributed equally to this work.

2012 and earlier

31. **Rau, A.**, Jaffrézic, F., Foulley, J.-L., and Doerge, R. W. (2012). Reverse engineering gene regulatory networks using approximate Bayesian computation. *Statistics and Computing*, 22 : 1257-1271.
32. **Rau, A.**, Jaffrézic, F., Foulley, J.-L., and Doerge, R. W. (2010). An empirical Bayesian method for estimating biological networks from temporal microarray data. *Statistical Applications in Genetics and Molecular Biology* : Vol. 9 : Iss. 1, Article 9.
33. Furth, A., Mandrekar, S., Tan, A. **Rau, A.**, Felten, S., Ames, M. Adjei, A. Erlichman, C. and Reid, J. (2008). A limited sample model to predict area under the drug concentration curve for 17-(allylamino)-17-demethoxygeldanamycin and its active metabolite 17-(amino)-17-demethoxygeldanamycin. *Cancer Chemotherapy Pharmacology* 61(1) : 39-45.

Peer-reviewed conference proceedings

1. **Rau, A.**, Jaffrézic, F., Foulley, J.-L., and Doerge, R. W. (2010). Approximate Bayesian approaches for reverse engineering biological networks. *Proceedings of the Kansas State University Conference on Applied Statistics in Agriculture*. Manhattan, Kansas.

Book chapters

1. Martin-Magniette, M.-L., Maugis-Rabusseau, C., **Rau, A.** (2016) Clustering of co-expressed genes. In : Choix et agrégation de modèles : Journée d'Etudes Statistiques. Ed. F. Bertrand, J.-J. Dreesbeke, G. Saporta, C. Thomas-Agnan.

Pre-prints, technical reports, and submitted articles

1. Devogel, N., Auer, P. L., Manansala, R., **Rau, A.**, and Wang, T. (2020) A unified linear mixed model for familial relatedness and population structure in genetic association studies. Submitted.
2. Cho, Y., **Rau, A.**, Reiner, A., Auer, P. L. (2020) Mendelian randomization analysis with survival outcomes. Submitted.
3. Cazals, A., Estellé, J., Bruneau, N., Coville, J.-L., Menanteau, P., Rossignol, M.-N., Jarret, D., Bevilacqua, C., **Rau, A.**, Bed'Hom, B., Velge, P., and Calenge, F. (2020) Impact of host genetics on caecal microbiota composition and on Salmonella carriage in chicken. Submitted.
4. Nuel, G., **Rau, A.**, and Jaffrézic, F. (2013). Joint likelihood calculation for intervention and observational data from a Gaussian Bayesian network. *arXiv preprint* arXiv:1305.0709.
5. **Rau, A.**, Celeux, G., Martin-Magniette, M.-L., and Maugis-Rabusseau, C. (2011). Clustering high-throughput sequencing data with Poisson mixture models. *Inria Research Report* 7786.

Other publications

1. **Rau, A.** (2008). Success of Volunteer Statistical Consulting Service Leads to Expanded Network. *The Statistical Consultant*, 25(1).
2. **Rau, A.** (2008). STATCOM Network Engages Growing Number of Student Volunteers. *Newsletter for the Section on Statistical Education*, 13(1).
3. **Rau, A.** (2008). Success of Statistical Service Leads to Expanded Network. *Amstat News*, April 2008.

Conference presentations

- July 2020 JOBIM 2020, virtual presentation
Invited keynote : *Integrative and interactive analyses of multi-omics data*
- Dec. 2019 EuroBioc 2019, Brussels, Belgium
Invited talk : *Individualized multi-omic pathway deviation scores using multiple factor analysis*
- Apr. 2019 AgreenSkills+ annual meeting, Brussels, Belgium
Poster : *Integrative methods for multi-omic data reveal multi-level gene and pathway regulation*
- Jan. 2018 Plant and Animal Genomes (PAG) XXVI, San Diego, California, USA
coseq : *An R/Bioconductor package for co-expression analyses of RNA-seq data*
- Jan. 2018 Plant and Animal Genomes (PAG) XXVI, San Diego, California, USA
Invited talk (FAANG workshop) : *An update on the FAANG pilot project FR-AgENCODE*
- July 2017 Working Group on Model-Based Clustering, Perugia, Italy
Invited talk : *Model-based clustering to identify co-expressed genes from high-throughput sequencing data*
- July 2017 useR!2017, Brussels, Belgium
Clustering transformed compositional data using coseq
- Nov. 2016 INRA RNA-seq day, Avignon
Invited talk : *Statistical tools to identify and visualize co-expression clusters from RNA-seq data*
- July 2016 Joint Statistical Meetings of the American Statistical Association, Chicago
Identifying marginal causal relationships in gene networks from observational and interventional expression data
- May 2016 2nd International Symposium on Microgenomics, Technical Workshop, Jouy-en-Josas
Invited talk : *Experimental design in 'omics studies*
- Feb. 2015 15th Workshop : Statistical Methods for Post-Genomic Data, Munich
HTSCluster : *a mixture-based approach for co-expression analyses of RNA-seq data*
- Nov. 2013 Statistical analysis of RNA-seq data : Advances and challenges, Paris
HTSDiff : *More sensitive differential analysis of RNA-seq data*
- June 2013 Deuxièmes rencontres R, Lyon
HTSFilter : *Data-based filtering for replicated high-throughput sequencing experiments*
- May 2013 Journée de la transcriptome végétale de l'URGV-Genopole, Evry
Invited round table : *Statistics applied to RNA-seq*
- Mar. 2013 StatSeq meeting on genetical genomics, Paris
Joint estimation of causal effects from observational and intervention gene expression data
- Oct. 2012 Journée APLIBIO (Alliance des PLates-formes Île-de-France de BIOinformatique), Paris

Invited talk : *A comprehensive evaluation of normalization methods for high-throughput RNA sequencing data analysis*

- July 2012 Joint Statistical Meetings of the American Statistical Association, San Diego, California
Clustering high-throughput sequencing data using Poisson mixture models
- Jan. 2012 12th Workshop : Statistical Methods for Post-Genomic Data, Lyon
Clustering high-throughput sequencing data using Poisson mixture models
- Apr. 2011 StatSeq Workshop, Toulouse
Invited talk : *Model-based cluster analysis for transcriptomic data* (G. Celeux, speaker)
- Jan. 2011 11th Workshop : Statistical Methods for Post-Genomic Data, Paris
Reverse Engineering Gene Networks Using Approximate Bayesian Computation
- Apr. 2010 Conference on Applied Statistics in Agriculture, Manhattan, Kansas
Approximate Bayesian methods for reverse engineering biological networks
- Sep. 2009 GENESYS Satellite Meeting at the European Conference on Complex Systems, Warwick, UK
Reverse-Engineering Gene Networks from Microarray Data with Dynamic Bayesian Networks
- Aug. 2009 Joint Statistical Meetings of the American Statistical Association, Washington, DC
Using Dynamic Bayesian Networks with Hidden States to Infer Gene Regulatory Networks
- May 2009 2nd Biennial Workshop on Statistical Bioinformatics and Stochastic Systems Biology, Newcastle, UK
Poster : *Reverse-Engineering Genetic Regulatory Interactions from Transcriptomic Data using Dynamic Bayesian Networks*
- Apr. 2009 Conference on Applied Statistics in Agriculture, Manhattan, Kansas
Poster : *An Empirical Bayes Approach to Inferring Genetic Regulatory Interactions with Dynamic Bayesian Networks*
- Feb. 2009 Gordon Conference on Quantitative Genetics and Genomics, Galveston, Texas
Poster : *An Empirical Bayes Approach to Inferring Genetic Regulatory Interactions with Dynamic Bayesian Networks*
- Aug. 2008 Joint Statistical Meetings of the American Statistical Association, Denver, Colorado
Poster : *Seven Years of StatCom at Purdue : Managing a Growing Number of Student Volunteers*

Seminar and working group presentations

- Jan. 2020 AgroParisTech statistics seminar, Paris, France
Integrative methods for multi-omic data reveal multi-level gene regulation
- Nov. 2019 INRA MaIAGE research seminar, Jouy en Josas, France
Integrative methods for multi-omic data reveal multi-level gene regulation
- Oct. 2019 Journée régionale Genotoul, Toulouse, France
Integrative methods for multi-omic data reveal multi-level gene regulation
- Sep. 2019 EpiFun workshop, Orléans, France
Integrative methods for multi-omic data reveal multi-level gene regulation
- Dec. 2018 Division of Biostatistics Seminar at MCW, Milwaukee, Wisconsin
Exploring drivers of gene expression in The Cancer Genome Atlas
- June 2018 MixStatSeq Workshop on Mixture Models : Theory and Application, Paris

- Co-expression analyses of RNA-seq data in practice with the R/Bioconductor package coseq*
- Apr. 2018 Research seminar series, Joseph J. Zilber School of Public Health, Milwaukee, WI
Exploring drivers of gene expression in The Cancer Genome Atlas
- Mar. 2018 Physiology Department Seminar at MCW, Milwaukee, Wisconsin
Exploring drivers of gene expression in The Cancer Genome Atlas
- June 2017 INRA national bioinformatics workshop, Dijon
Easy interactivity in R with (gg)plotly and Shiny
- Mar. 2017 SAPS doctoral school : Experimental animal biology and predictive modelisation
Challenges in data integration
- Sep 2016 Seminar at the Human and Molecular Genetics Center, Milwaukee College of Medicine
Transformation, model choice, and visualization for RNA-seq co-expression
- Sep 2016 Seminar at the Zilber School of Public Health, Milwaukee, Wisconsin
Transformation, model choice, and visualization for RNA-seq co-expression
- May 2016 Groupe de travail de statistiques du LMRS de Rouen
Poisson mixtures with slope heuristics and visualization tools for RNA-seq co-expression
- Mar. 2016 INRA national bioinformatics workshop, Toulouse
From genotype to phenotype : what statistical methods to integrate heterogeneous data ?
- Mar. 2016 SAPS doctoral school : Experimental animal biology and predictive modelisation
Integration of heterogeneous 'omics data
- Mar. 2016 INRA NGS club, Jouy-en-Josas
Poisson mixture models and visualization tools for RNA-seq co-expression
- Jan. 2016 MAP5 seminar at Université Paris-Descartes
Poisson mixtures with slope heuristics and visualization tools for RNA-seq co-expression
- Jan. 2016 4th Annual SFdS Young Statisticians and Probabilists Day, Paris
Model selection in mixture model based classification : Applications in biostatistics
- Nov. 2015 Seminar at Toulouse Mathematics Institute (IMT), Toulouse
Statistical analysis of microarray and RNA-seq data
- Nov. 2015 Statistique seminar, Paris
Integrative clustering and classification in multiple heterogeneous data
- Sep. 2015 NETBIO working group, Paris
RNA-seq co-expression analysis using mixture models
- Sep. 2015 Cirad seminar, Montpellier
HTScluster : a mixture-based approach for co-expression analyses of RNA-seq data
- Oct. 2014 SELECT seminar, Orsay
Slope heuristics : the missing ingredient for identifying co-expressed genes from RNA-seq data
- Nov. 2013 Statistics for Systems Biology (SSB) seminar, Evry
HTSFilter : filtering replicated RNA-seq data using a data-driven approach
- Nov. 2013 Seminar at INRA-GABI, Jouy-en-Josas
Reinforcing the biology-statistics feedback loop with tools for genomic data analysis
- Oct. 2013 Statistics for Integrative Biology (SIB) seminar, Rennes
HTSAnalysis : a suite of R/Bioconductor packages for the analysis of RNA-seq data
- Sep. 2013 NETBIO working group, Paris
Joint estimation of causal effects from observational and intervention gene expression data

- June 2013 Statistique et Santé working group, Paris
Joint estimation of causal effects from observational and intervention gene expression data
- June 2013 AppliBUGS Workshop, Paris
Joint estimation of causal effects from observational and intervention gene expression data
- June 2013 Statistics seminar, Toulouse
Joint estimation of causal effects from observational and intervention gene expression data
- Feb. 2013 Statistics for Integrative Biology seminar, Rennes
Joint estimation of causal effects from observational and intervention gene expression data
- Dec. 2012 Assemblée générale PEPI IBIS, Toulouse
Differential analysis of RNA-seq data by unsupervised classification
- Nov. 2012 Statomique seminar, Lyon
Independent data-based filtering for replicated high-throughput sequencing experiments
- June 2012 LGC and SAGA seminar at INRA, Toulouse
Clustering high-throughput sequencing data using Poisson mixture models
- June 2012 SSB working group seminar, Jouy en Josas
Clustering high-throughput sequencing data using Poisson mixture models
- Feb. 2012 MIA “Biological network inference” methodological working group meeting, Paris
Inferring gene regulatory networks with hidden variables using state space models
- Dec. 2011 AppliBugs Workshop, Paris
Exploring the identifiability of gene regulatory networks with approximate Bayesian computation
- May 2011 Seminar at the Institut de Recherche Mathématique Avancée, Strasbourg
Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)
- Apr. 2011 Seminar at the Unité de Recherche en Génomique Végétale, Évry
Reverse Engineering Gene Networks : A Statistician’s Perspective
- Apr. 2011 Seminar at the Institut de Mathématiques de Luminy, Marseille
Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)
- Mar. 2011 Seminar at the Laboratoire Statistique et Génome, Évry
Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)
- Mar. 2011 Seminar at the équipe Génétique et Génomique Statistique, Le Kremlin Bicêtre
Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)
- Jan. 2011 Rencontre de statistique autour des modèles hiérarchiques, Strasbourg
Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)
- Oct. 2010 INA P-G, Paris Descartes, and SELECT working group, Paris
Reverse Engineering Gene Networks Using Approximate Bayesian Computation (ABC)
- Apr. 2010 Bioinformatics seminar at Purdue University, West Lafayette, Indiana
Approximate Bayesian methods for reverse engineering biological networks
- June 2009 Seminar at the AgroParisTech, Paris
Inférence sur les réseaux génomiques par des modèles espace-état
- June 2009 Seminar at the UMR GABI-INRA, Jouy-en-Josas
Inférence sur les réseaux génomiques par des modèles espace-état
- Mar. 2009 Ph.D. student seminar, INRA Département de Génétique Animale, Jouy-en-Josas
Reverse Engineering Gene Regulatory Networks
- Mar. 2008 Ph.D. student seminar, INRA Département de Génétique Animale, Toulouse

Participation in working groups and seminars

- Statomique working group (quarterly, 2009-present)

Software and R packages

- **padma** : Pathway deviation scores using multiple factor analysis
Available on Bioconductor at <https://www.bioconductor.org/packages/padma>.
- **Regeneration Rosetta** : An R/Shiny interactive web application to explore regeneration-associated gene expression and chromatin accessibility
Available at <http://ls-shiny-prod.uwm.edu/rosetta>.
- **maskmeans** : Multi-view aggregation/splitting K-means clustering algorithm
Available on GitHub at <https://github.com/andreamrau/maskmeans>.
- **EDGE in TCGA** : An R/Shiny interactive web application for the exploration of drivers of gene expression in The Cancer Genome Atlas
Available at http://ls-shiny-prod.uwm.edu/edge_in_tcg/.
- **coseq** : Co-expression analysis of sequencing data
Available on Bioconductor at <https://bioconductor.org/packages/coseq/>.
- **ICAL** : Model selection for model based clustering of annotated data
Available on Github at <https://github.com/Gallop/ICAL>.
- **metaRNASeq** : Meta-analysis of RNA-seq data
Available on CRAN at <http://cran.r-project.org/web/packages/metaRNASeq>.
- **HTSFilter** : Filter for replicated high-throughput sequencing data
Available on Bioconductor at www.bioconductor.org/packages/HTSFilter/.
- **HTScluster** : Clustering high throughput sequencing data
Available on CRAN at <http://cran.r-project.org/web/packages/HTScluster>.
- **ebdbNet** : Empirical Bayes estimation for Dynamic Bayesian Networks
Available on CRAN at <http://cran.r-project.org/web/packages/ebdbNet>.

Administrative activities and service

Reviewer	<i>Revue d'Intelligence Artificielle, IEEE Proceedings, Statistical Applications in Genetics and Molecular Biology, BMC Bioinformatics, BMC Genomics, RNA, Nucleic Acids Research, Molecular Genetics and Genomics, F1000 Research, Bioinformatics, Computational and Structural Biotechnology Journal, GigaScience, BMC Medical Genetics, Briefings in Bioinformatics, Annals of Applied Statistics, Genome Biology, The Plant Journal, Nature Computational Science</i>
2019	Scientific committee member, User!2019 international conference
2016	Scientific committee president, Rencontres R national conference
2015–2019	Member of the Conseil Scientifique des Utilisateurs (CSU) of the MIGALE Bioinformatics platform (INRA, Jouy en Josas)
2010	Chair, Committee on Student Pro Bono Statistics of the ASA

- 2009–2010 Member, Committee on Student Pro Bono Statistics of the American Statistical Association (ASA)
- 2009 Organizer of an invited round table (*The Pros of Pro Bono Statistics*) at the Joint Statistical Meetings of the American Statistical Association (ASA) in Washington, DC, USA
- 2006–2010 Member, Statistics in the Community (StatCom) at Purdue University : StatCom is a volunteer organization of graduate students that provides free professional statistical consulting services to government and nonprofit groups

Funding

Current Research Support

- **GENE-SWitCH** : The regulatory genome of swine and chicken : functional annotation during development
Principal investigators : Elisabetta Giuffra and Herve Acloque
Role : Co-investigator and task leader
Funding period : 2019 - 2023
Agency : H2020 RIA grant
- **EpiFun** : Systems biology for genomic selection
Principal investigators : Nathalie Vialaneix and Thomas Faraut
Role : Co-investigator
Funding period : 2018 - 2020
Agency : INRA internal metaprogram grant
- **LIPOMEC** : Towards a better understanding of ruminant milk lipolysis through an integrative biology approach in milk and mammary epithelial cells
Principal investigators : Christelle Cebo
Role : Co-investigator
Funding period : 2018 - 2022
Agency : French National Research Agency (ANR) grant

Past Research Support

- **Microficient** : Relationships between digestive microbiota and feed efficiency in cattle
Principal investigators : Yuliaxis Ramayo and Gilles Renand
Role : Co-investigator
Funding period : 2016 - 2019
Agency : AP-2016-007
- **CARISTO-PF** : Characterization and management of health and environmental risks linked to the development of ciguatera in phytobenthos in French Polynesia
Principal investigator : Gregory Nuel
Role : Co-investigator
Funding period : 2016 - 2019
Agency : French Polynesia Territory grant
- **MixStatSeq** : Mixture-based procedures for statistical analysis of RNA-seq data

Principal investigator : Cathy Maugis-Rabusseau

Role : Co-investigator

Funding period : March 2014 - February 2018

Agency : French National Research Agency (ANR) grant (ANR-13-JS01-0001-01)

• **SalmoCar** : Genetic and microbiotal control of *Salmonella* carriage in chicken and mice

Principal investigator : Xavier Montagutelli

Role : Co-investigator

Funding period : January 2015 - December 2017

Agency : Institut Carnot Pasteur Maladies Infectieuses (PMI) / Institut Carnot Santé Animale (ICSA) grant

• **COSI-net** : Using COmbinatorial gene Silencing & Inactivation to infer gene NETworks

Principal investigator : Andrea Rau

Funding period : August 2015 - December 2016

Agency : INRA Animal Genetics Department internal grant

• **Causality** : Causal network inference

Principal investigator : Florence Jaffrézic

Role : Co-investigator

Funding period : January 2014 - December 2014

Agency : INRA Animal Genetics Department internal grant

Advising

• Fanny Mollandin (2019-2022 Ph.D., co-supervision with P. Croiseau)

Incorporating known functional annotations into Bayesian genomic prediction models

Alumni

• Dr. Gilles Monneret (2014-2017, Ph.D., co-supervision with F. Jaffrézic and G. Nuel)

Estimation of causal effects in gene networks from observational and intervention data

• Raphaëlle Momal-Leisenring (2017, M2 intern)

Integrative statistical analysis of multi-omics data

• Frédéric Jehl (2017, M2 intern, co-supervision with T. Zerjal)

Impact of heat stress on liver and blood transcriptomes of laying hens

• Dr. Manuel Revilla Sanchez (2016 3-month Ph.D. Erasmus+ Learning Mobility, co-supervision with Jordi Estelle and Yuliaxis Ramayo Caldas and Ph.D. manuscript *rapporteur*)

An integrative gene network analysis of the genetic determination of pig fatty acid composition

• Babacar Ciss (2016, M2 intern, co-supervision with E. Sellem)

Constructing predictive models for ovine production data

• Dr. Méлина Gallopin (2012-2015, Ph.D., co-supervision with G. Celeux and F. Jaffrézic)

Classification and network inference for RNA-seq data

• Audrey Hulot (2015, M1 intern)

Incorporating a priori biological knowledge into gene network inference from observational and intervention gene expression data

• Meriem Benabbas (2015, M1 intern)

Identifying differentially expressed genes from RNA-seq data using mixtures of generalized

linear models

- Marc Teissier and Chaoyu Dong (2014, Master STV/EM-ABG)
Power to detect significantly differential gene expression using RNA-seq data
- Mélina Gallopin (2012, M2 intern)
Gene network inference from RNA-seq data
- Rémi Bancal (2012, M2 intern)
Gene network estimation by adaptive knockout experiments

Membership in advisory and evaluation committees

- Antoine Leduc (2019-2022, member of Ph.D. advisory committee)
- Dr. Alyssa Imbert (2018, member of Ph.D. evaluation committee)
- Frédéric Jehl (2017-2020, member of Ph.D. advisory committee)
- Gabriel Guillocheau (2016-2018, member of Ph.D. advisory committee)
- Dr. Valentin Voillet (2016, member of Ph.D. evaluation committee)
- Member of evaluation committee for assistant professor position at Université Rennes I, UMR Inserm IRSET 1085 (2016)
- Member of evaluation committee for assistant professor position at Université Rennes I, IGDR (Institute of Genetics and Developmental biology of Rennes), CNRS UMR 6290 (2013)

Teaching

- 2020 **Agrocampus Researcher School** (6-7 February)
Statistical analysis of RNA-seq data
Type : Instructor (with S. Lagarrigue and Y. Blum)
• Rennes
- 2019 **Researcher training session : From gene expression to genomic networks** (17-22 July)
Differential analysis of RNA-seq data
Type : Instructor (with M.-L. Martin-Magniette and E. Delannoy)
• INRAE AgroImpact
- 2019 **Analysis of livestock metagenomics datasets** (13-17 May)
Type : Instructor (with J. Estellé)
• INRA URZ, Guadeloupe
- 2019 **Physiological genomics** (10 hours)
R Bootcamp
Type : Instructor
• Medical College of Wisconsin
- 2018 **Data management and visualization in R** (3 course units)
Type : Instructor
• University of Wisconsin-Milwaukee
- 2017 **PiGutNet Training School** (3 hours)
Differential abundance analysis for microbial marker-gene surveys with `metagenomeSeq`
Type : Instructor
• Jouy en Josas

- 2017 **Mathematics for the Life Sciences : Statistical Engineering and Genomics, M2**
 Bayesian statistics for genomics course (18 hours)
Type : Instructor (coursework and labs)
Program : Bayesian statistics for population genetics, numeric methods for Bayesian computation, empirical Bayes, approximate Bayesian computation, Bayesian integrative analysis of multi-omics data
Supervisor : Marie-Luce Taupin
 • Université d'Évry Val d'Essonne
- 2017 **Agrocampus Researcher School** (1-2 February)
 Statistical analysis of RNA-seq data
Type : Instructor (with S. Lagarrigue and Y. Blum)
 • Rennes
- 2017 **Biostatistics Master, M2** : Genomics course (12 hours)
Type : Instructor (coursework and labs)
Program : Differential analysis for RNA sequencing data, systems biology approaches for gene network inference
Supervisor : Brigitte Gelein
 • Ensai, Rennes
- 2016 **SPS Summer School : From gene expression to genomic networks** (17-22 July)
 Co-expression analysis of RNA-seq data (3 hours)
Type : Instructor (coursework and labs)
Program : Co-expression, K-means, hierarchical clustering, finite mixture models
Supervisor : Marie-Laure Martin-Magniette
 • Institute of Plant Sciences Paris-Saclay
- 2016 **Mathematics for the Life Sciences : Statistical Engineering and Genomics, M2**
 Bayesian statistics for genomics course (18 hours)
Type : Instructor (coursework and labs)
Program : Bayesian statistics for population genetics, numeric methods for Bayesian computation, empirical Bayes, approximate Bayesian computation, Bayesian integrative analysis of multi-omics data
Supervisor : Marie-Luce Taupin
 • Université d'Évry Val d'Essonne
- 2016 **Agrocampus Researcher School** (10-11 February)
 Statistical analysis of RNA-seq data
Type : Instructor (with S. Lagarrigue and Y. Blum)
 • Rennes
- 2016 **Biostatistics Master, M2** : Genomics course (33 hours)
Type : Instructor (coursework and labs), and genomics project supervision
Program : Differential analysis for RNA sequencing data, systems biology approaches for gene network inference, and analysis of multi-omic TCGA data
Supervisor : Brigitte Gelein
 • Ensai, Rennes
- 2015 **Mathematical Engineering for Life Sciences Master, M1** : Case study (10 hours)
Type : Instructor (coursework and labs)
Program : Differential analysis for RNA sequencing data, systems biology approaches
Supervisor : Etienne Birmelé

- Université Paris Descartes, Paris
- 2015 **Biostatistics Master, M2** : Genomics course (33 hours)¹

Type : Instructor (coursework and labs), and genomics project supervision
Program : Differential analysis for RNA sequencing data, systems biology approaches for gene network inference, and analysis of the Sequencing Quality Control Consortium RNA-seq and qPCR data
Supervisor : Brigitte Gelein

 - Ensai, Rennes
- 2013 **BioBayes Researcher School** (7-11 October)

Bayesian statistical methods : Introduction to theory and applications in food, environment, epidemiology, and genetics
Type : Scientific committee member and Instructor (coursework and labs)

 - Cannes Mandelieu
- 2013 **Biostatistics Master, M2** : Genomics course (6 hours)

Type : Instructor (coursework and labs) and genomics project jury member
Program : Differential analysis for RNA sequencing data
Supervisor : Brigitte Gelein

 - Ensai, Rennes
- 2012 **Training school on rabbit and pig genome analysis** (6 hours)

Type : Instructor
Program : Differential analysis for microarray and RNA sequencing data

 - COST action research school, Norwich, United Kingdom
- 2012 **Biostatistics Master, M2** : Genomics course (6 hours)

Type : Instructor (coursework and labs)
Program : Differential analysis for RNA sequencing data
Supervisor : Marian Hristache

 - Ensai, Rennes
- 2012 **Statistical Methods for Genome Enabled Prediction** (2 hours)

Type : Intervenante (cours)
Programme : Approximate Bayesian methods : Application to gene regulatory networks
Enseignant principal : Daniel Gianola

 - European Graduate School in Animal Breeding and Genetics, Paris
- 2012 **Next generation sequencing school for INRA researchers** (2 hours)

Type : Instructor
Program : Differential analysis for RNA sequencing data

 - Research school (INRA), Écully
- 2012 **Bioinformatics and Biostatistics / Mathematical engineering / Probability and Statistics Master, M2** : Computational biostatistics (6 hours)

Type : Instructor (coursework and labs)
Program : Differential analysis for RNA sequencing data
Supervisor : Marc Lavielle

 - UFR de Sciences, Université Paris-Sud 11
- 2011 **Bioinformatics and Biostatistics Master, M1** : Statistical modeling (24 hours)

1. The genomics course at Ensai was significantly expanded and re-developed in 2015 by myself and Mickaël Guedj.

Type : Lab assistant (R)

Program : Maximum likelihood estimation of non-linear models, Wald and likelihood ratio tests, logistic regression, bootstrap

Supervisor : Marie-Anne Poursat

• UFR de Sciences, Université Paris-Sud 11

2011 **Bioinformatics and Biostatistics / Mathematical engineering / Probability and Statistics Master, M2** : Computational biostatistics (12 hours)

Type : Instructor (coursework and labs)

Program : Introduction to Bayesian analysis for population genetics, introduction to statistical analysis for transcriptome data

Supervisor : Marc Lavielle

• UFR de Sciences, Université Paris-Sud 11

2006 **1st and 2nd year undergraduate** : Elementary statistical methods (60 hours)

Type : Lab assistant (SPSS)

Program : Descriptive statistics, confidence intervals, simple linear regression, probability calculations, test statistics for means and proportions, correlation

Supervisor : Ellen Gundlach

• Department of Statistics, Purdue University (West Lafayette, Indiana, USA)