Bordeaux Population Health – *Inserm UMR 1219*
Université de Bordeaux
146 rue Léo Saignat
33076 Bordeaux cedex
FRANCE

French, 36 years old
boris.hejblum@inserm.fr
boris.hejblum.science

Boris Hejblum

Research faculty in Biostatistics (*Chargé de Recherche*)

Research experience

2021–present **Faculty Researcher (*Chargé de Recherche*)** in Biostatistics, tenured, Inserm U1219 *Bordeaux Population Health* research center, *SISTM team*, Bordeaux (France).

2016–2021 **Associate Professor (*Maître de Conférences*)** in Biostatistics, tenured, ISPED *Bordeaux School of Public Health*, Bordeaux University, Bordeaux (France).

2016 **Postdoctoral Research Associate**, Department of Biostatistics, Harvard School of Public Health, Boston (USA).

2015–2016 **Postdoctoral Research Fellow**, Department of Biostatistics, Harvard School of Public Health, Boston (USA).

2011–2015 **Research Assistant** (Ph.D. student), Inserm U897 *Biostatistics team*, Bordeaux (France).

Apr.–Sept. 2011 **Research Assistant** (Masters intern), Inserm U897 *Biostatistics team*, Bordeaux (France).
Development of dynamic statistical models applied to the epidemiology of myocardial infarction.

May–Jul. 2011 **Statistician Assistant** (Masters intern), *AltraBio* (start-up in biotechnologies), Lyon (France).
Analysis of transcriptomics data of preclinical trials.

Education

2024 **Habilitation to supervize research** (H.D.R. *Habilitation à Diriger des Recherches*), Bordeaux University.
Statistical methods for leveraging high-dimensional data from high-throughput measurements in vaccine clinical development.

2011–2015 **Ph.D. in Biostatistics**, ISPED *Bordeaux School of Public Health*, Bordeaux University.
Integrative analysis of high-dimensional data applied to vaccine research.
Advisors: Pr. Rodolphe Thiébaut (rodolphe.thiebaut@u-bordeaux.fr),
François Caron (caron@stats.ox.ac.uk)

2008–2011 **Master of Science (M.Sc.) in Statistics** (*diplôme d'ingénieur*), ENSAI, National School for Statistics and Information Analysis (*École Nationale de la Statistique et de l'Analyse de l'Information*), Rennes (France). Specialization in biostatistics, with high honors.

2011 **Master of Science (M.Sc.) in Statistics and Econometrics**, Department of Mathematics, University of Rennes 1, Rennes (France). Dual degree partnership in conjunction with studies at ENSAI (additional education focused on scientific research).

2009 **Bachelor of Science (B.Sc.) in Mathematics** (*licence de mathématiques*), Pierre and Marie Curie University – Paris 6 (UPMC), Paris (France). In conjunction with studies at ENSAI (dual curriculum, remote learning).

2006–2008 **Post-Secondary Preparatory Classes** (*Classes Préparatoires aux Grandes Écoles – CPGE*), Lycée Hoche, Versailles (France). University-level courses required in preparation for competitive exams into top universities, engineering, and graduate schools (France's *Grandes Écoles*). Major in Mathematics and Physics.

2006 **High school diploma**, Lycée Richelieu, Rueil-Malmaison (France). With high honors.

Teaching experience

2019 - present **International Ph.D. course**, Graduate School of Health and Medical Sciences, University of Copenhaghen, (Danemark)

- Bayesian methods in biomedical research (graduate class, 3.5 days per year)

2018 - present **Ph.D. courses**, Bordeaux University (France)

- R for development & performance (graduate class, 18h per year)
- Basics for data science using R (graduate class, 12h per year)

2019 - present **Master in Public Health**, ISPED, Bordeaux University (France)

- omics data analysis (graduate class, 20h per year)
- data visualization (undergraduate class, 4h)

2021 - present **Master in numerical sciences & bio-health**, École Centrale Nantes (France)

- Statistical learning in high-dimension (graduate class, 2h per year)

2016 - 2021 **Associate Professor**, Bordeaux University, France

Ph.D. courses:

- Introduction to Bayesian analysis for biometric research (graduate class, 18h per year)

Master in Public Health Data Science & Master in Biostatistics courses:

- likelihood estimation and multivariate regression (graduate class, 30h per year)
- factor methods for multivariate data analysis (graduate class, 30h per year)
- Bayesian analysis and sampling methods (graduate class, 30h per year)
- omics data analysis (graduate class, 20h per year)
- sparse Partial Least Squares methods (graduate class, 7h per year)
- ANOVA regression (gratuate class, 7.5h per year)
- hypothesis testing (graduate class, 30h per year)
- advanced R (undergraduate class, 15h per year)

2012 - 2014 **Teaching Assistant**, Bordeaux University, France

Master in Public Health and Master in Biostatistics courses:

- MCMC methods for Bayesian analysis (graduate class, 12h)
- sparse Partial Least Squares methods (graduate class, 5h)
- basic statistics (undergraduate class 16h)
- logistic regression (undegratuate class, 12h)
- R software (undergraduate class 9h)

Scientific supervision

Postdoctoral researchers

- Laura Villain (2019 – 2021: 100%)
- Hung Van Tran (2019: 50%)

Ph.D. students

- Sara Fallet (2024 – ...: 50%)
- Annesh Pal (2023 – ...: 100%)
- Arthur Hughes (2023 – ...: 50%)
- Kalidou Ba (2022 – ...: 50%)
- Benjamin Hivert (2020 – 2024: 50%)
- Paul Freulon (2019 – 2022: 50%)
- Marine Gauthier (2018 – 2021: 50%)
- Soufiane Ajana (2017 – 2019: 15%)
- Stephanie Chan (2016: 15%)

Engineers

- Quentin Laval (2024 – ...: 100%)
- Mélanie Huchon (2021 – 2025: 100%)
- Clément Nerestan (2019 – 2021: 100%)
- Mélanie Durand (2019 – 2022: 100%)
- Solène Delhaye (2017 – 2019: 100%)

Interns

- Theodora Georgakopoulou Vigneras (M2 internship 2025: 50%)
- Alarig Vigneras (M1 internship 2024: 50%)
- Mathéo Le Floch (M1 internship 2024: 50%)
- Arthur Hughes (M2 internship 2023: 100%)
- Maud Perpere (M1 internship 2023: 100%)
- Emma Avisou (M1 internship 2021: 100%)
- Clément Bonnet (M1 internship 2021: 100%)
- Benjamin Hivert (M2 master thesis 2020: 100%)
- Anthony Devaux (M2 master thesis 2019: 100%)
- Aaron Sonabend (PhD research visit 2019: 100%)
- Victor Gasque (M1 internship 2019: 50%)
- Thomas Ferté (M2 master thesis 2019: 100%)
- Marine Gauthier (M2 master thesis 2018: 100%)
- Roxane Coueron (M2 master thesis 2018: 50%)
- Paul Tauzia (M2 master thesis 2017: 50%)
- Chariff Alkhassim (M2 master thesis 2015: 50%)
- Damien Chimits (M2 master thesis 2014: 50%)
- Lise Cahuzac (M1 internship 2013: 50%)

Grants & funding

2022-2026 **Work-Package leader** *Réseau de Recherche Impulsion* "Public Health Data Science", *Université de Bordeaux*.

2016-2026 **Participant** ANRS LabEx Programme "Vaccine Research Institite" (VRI).

2023-2027 **Task leader** PEPR *Santé Numérique*, axis "Statistical and AI based Methods for Advanced clinical Trials Challenges in digital Health" (funding 1 PhD student).

2023-2027 **Task leader** PEPR *Santé Numérique*, axis "multiScale AI for SingleCell-based precision MEDicine" (funding 50% of 1 PhD student).

2018-2020 **Principal Investigator** of the Inria associate-team DESTRIER: "DEfining Surrogacy of early Transcriptomics foR vaccInE Response" (36K€ over 3 years)

2020-2024 **Participant** (genomics-statistics referent) in the EU H2020 Framework Programme "IP-cure-B" (*Immune profiling to guide host-directed interventions to cure HBV infections*).

2018-2020 **Principal Investigator** of the Inria associate-team SWAGR: "Statistical Workforce for Advanced Genomics using RNA-seq" (36K€ over 3 years)

2019-2022 **Participant** (computational statistics referent) in the ANR-18-CE36-0004 "DyMES" (*Dynamic Models for Epidemiological Longitudinal Studies of Chronic Diseases*).

2019-2021 **Principal Investigator** of the Technology Development Action from Inria Bordeaux Sud-Ouest "VASI" (*Visualization and Analysis Solutions for Immunologists*): 2 year support for a software engineer.

2018-2020 **Teaching discharge** for research at Inria Bordeaux Sud-Ouest: 96h per year.

2017-2020 **Participant** (RNA-seq anlaysis referent) in the Transcan-2 ERA-NET " GLIOMA-PRD" (*Multiparametric analysis of the evolution and progression of low grade glioma*): support for a post-doctoral researcher for 2 years.

2016-2019 **Participant** (référent statistique en grande dimension) au *Research and Innovation Programme* n°634479 de EU H2020 EYE-RISK (*Systems medicine for identifying risk factors, molecular mechanisms and therapeutic approaches for age-related macular degeneration*).

2016 **Recipient** of a travel grant from the Harvard Program in Quantitative Genomics (PQG) to attend the ENAR conference.

2011 **Recipient** of a Ph.D. grant from the EHESP (*École des Hautes Études en Santé Publique*, Rennes, France) – ranked 1st.

Patents

2021 Invention patent EP20306527/WO2022122959A1 (inventor $1/5^{th}$)
Use of cd177 as biomarker of worsening in patients suffering from covid-19

2020 Invention patent WO2021058914A1/FR1910515 (inventor $1/7^{th}$)
Prediction of the content of omega-3 polyunsaturated fatty acids in the retina by measuring 7 cholesterol ester molecules

Research expertise

Statistical genomics & high dimensional data: I have a strong interest in models for high dimensional data. I am familiar with the multiple testing issue and potential strategies to face it. I have worked on sparse Partial Least Squares methods, and with other dimension reduction approaches such as the random forests or the LASSO. I have analyzed gene expression data in a clinical trial context and I am familiar with the specificities of this kind of data, such as preprocessing.

Artificial Intelligence for health: I have developed various artificial intelligence approaches to solve biomedical data analysis bottlenecks. In particular, I am working on machine learning approaches to automate the processing of flow and mass cytometry measurements, and also on automated medical diagnosis from both structured data and free text medical notes in English, French and Chinese through language agnostic algorithms.

Electronic Health Records: I am currently developing models to perform probabilistic record linkage to match electronic health records without using identifier variables, and to predict disease phenotype from electronic health record data, with application in infection and rheumatoïd arthritis.

Bayesian nonparametric models: I am interested in statistical learning methods such as nonparametric Bayesian mixture of skew distributions for the clustering of large cell populations.

Evidence synthesis causal analysis: I studied stochastic modeling of life-course health data. The developed idea was to explore potential causal factors of myocardial infarction by relating the drift of a degradation process with metadata from the literature.

Software development & maintenance

2023 **citcdf**: an R package for performing Conditional Independence Testing Through Conditional Cumulative Distribution Function Estimation. Available on GitHub. *Co-creator & maintainer*.

2022 **CytOpT**: an R package for automatic gating transfer in cytometry data using optimal transport with domain adaptation. Uses Python code.. Available on CRAN, development version on GitHub. *Co-creator & maintainer*.

2020 **dearseq**: an R package for Differential Expression Analysis for RNA-seq data through a robust variance component test. Available on Bioconductor, development version on GitHub. *Co-creator & maintainer*.

2019 **vici**: an interactive R Shiny application for accurate estimation of vaccine induced cellular immunogenicity with bivariate linear modeling. Available online or locally from the CRAN, development version on GitHub. *Creator & maintainer*.

2019 **marqLevAlg**: an R package for (parallelized) optimization of convex multiparametric functions. Available on CRAN, development version on GitHub. *Contributor*.

2019 **foodingraph**: an R package for displaying weighted undirected food networks from adjacency matrices. Available on CRAN, development version on GitHub. *Co-creator*.

2019 **phenotypr**: an R package for probabilistic phenotyping patients from electronic health records using both diagnosis codes and natural language processed medical notes. Available on CRAN, development version on GitHub. *Creator & maintainer*.

2017 **ludic**: an R package for probabilistic record linkage using diagnosis codes. Available on CRAN, development version on GitHub. *Co-creator & maintainer*.

2017 **cytometree**: an R package for automatic gating and annotation of flow-cytometry data. Available on CRAN, development version on GitHub. *Co-creator & maintainer*.

2017 **sslcov**: an R package for covariance semi-supervised learning. Available on GitHub. *Co-creator*.

2016 **tcgsaseq**: an R package for longitudinal RNA-seq data analysis at the gene set level. Available on GitHub. *Co-creator & maintainer*.

2017 **kernscr**: an R package for survival analysis by gene sets in presence of competing risks. Available on CRAN, development version on GitHub. *Co-creator & maintainer*.

2015 **NPflow**: an R package for clustering of large cell populations with Dirichlet process mixture of skew-Normal and skew-t distributions. Uses C++ code to speed up computation.. Available on CRAN, development version on GitHub. *Co-creator & maintainer*.

2014 **TcGSA**: an R package for longitudinal gene-expression data from microarrays at the gene set level. Available on CRAN, development version on GitHub. *Creator & maintainer*.

Outreach activities

2022-present *Chiche ! 1 Scientifique, 1 Classe* Program by Inria
1h presentation & open discussion about scientific research with high-school students.

2024 *Nuit des chercheurs 2024* with the Inserm exhibition "Des virus émergents et des épidémies".

2018 Outreach stand "*Is there more data in a drop of blood than in my smartphone?*" at the 10 year anniversary of Inria Bordeaux Sud-Ouest

2012 Poster presentation at the Summer University of Sidaction on longitudinal analysis applied to HIV vaccine research

Research visits abroad

2018-2019 (2×3 weeks) **MRC Biostatistics Unit, Cambridge University**, Cambridge (United-Kingdom)
invited by Sylvia Richardson, Professor.

2018 (1 week) **Rand Corporation, Statistics group**, Santa Monica (CA, USA)
invited by Denis Agniel, Associate Statistician.

2016-2017 (2×1 week) **Harvard University, Department of Biostatistics**, Cambridge (MA, USA)
invited by Tianxi Cai, Professor.

2013-2014 (3×1 week) **University of Oxford, Department of Statistics**, Oxford (United-Kingdom)
invited by François Caron, Research Fellow.

2012 (1 month) **Benaroya Research Institute**, Chaussabel Laboratory, Seattle (WA, USA)
invited by Damien Chaussabel, Director of Systems Immunology.

2011 (1 month) **Baylor Institute for Immunology Research**, Dallas (TX, USA).

Scientific evaluation

2024-2025 **Member of the ANRS-MIE CSS13 ("*Recherches cliniques*")**

2025 **Member of the Scientific Comittee for the 10th Channel Network Conference in Liège (Belgium)**

Statistique, exposome et santé environnementale

2025 **Member of the Strategic Comittee for the Inserm Cl**
et de suivi de la trajectoire

2025 **Member of the Scientific Comittee for the "Mathematics of Single-Cell Data-Analysis" reserch school at C.I.R.M.**

2025 **Member of the Scientific Comittee for the conference on "Statistics, exposome and environmental health" in Rennes**

2024-2025 **Expert reviewer for the ANR in CE45 ("*Interfaces : mathématiques, sciences du numérique – biologie, santé*") and CE23 ("*Intelligence artificielle et science des données*")**

2025 **Expert reviewer for the "DIM1HEALTH 2.0" call**

2024 **Reviewer of the PhD thesis of Nicolas Enjalbert-Courrech, *Université de Toulouse***

2023-2024 **Reviewer for the MESSIDORE project call from Inserm IReSP "Méthodologie des ESSais cliniques Innovants, Dispositifs, Outils et Recherches Exploitant les données de santé et biobanques"**

2023 **Member of the Scientific Comittee for the 9th Channel Network Conference in Wageningen (Netherland)**

2021 **Member of the PhD defense committee of Shaima Belhechmi, *Université Paris-Saclay***

2021 **Reviewer for the *ANRT***, *(Association Nationale de la Recherche Technologique)*

2021 **Member of the Scientific Committee for the 42nd ISCB conference**

2021 **Member of the Pharm. D. defense committee of Blandine Malbos, *Université d'Angers***

2019 **Invited member of the PhD defense committee of Soufiane Ajana, *Université de Bordeaux***

Editorial activities

2025 - present **Reproducible Research Editor for the *Biometrical Journal***

2025 - present **Associate Editor for *Biometrics***

Reviewer for international peer-reviewed scientific journals

Annals of Applied Statistics, *Bayesian Analysis*, *BioData Mining*, *Bioinformatics*, *Biometrics*, *Cell Reports Methods*, *Cancer Reports*, *Computational Statistics Data Analysis*, *Journal of Open Source Software*, *Journal of Statistical Computation and Simulation*, *PLOS Computational Biology*, *Scientific Reports*, *STAT*, *Statistics in Medicine*, *Statistical Applications in Genetics and Molecular Biology*, *WIREs Applications in Genetics and Molecular Biology*

Academic responsibilities

2022–2024 **Organizer of the yearly datathon for the "Public Health Data Science" reserach network from the University of Bordeaux**

2023–2024 **Member of the Organizing Comittee for the 55th annual conference *"Journées de Statistique"* of the French Statistical Society (SFdS) in 2024**

2021–present **French Biometric Society correspondant to the Channel Network region of the International Biometrics Society**

2019–present **Member of the Bureau of the French Biometric Society** *(Société Française de Biométrie)* **– webmaster**

2017–present **Organizer of the Biostatistics Seminar series from the Public Health Department of Bordeaux University** (biweekly)

2019 **Co-organizer of the Bordeaux Statistics Seminar series** (quarterly)

2018 **Co-organizer of the workshop in honor of Daniel Commenges' 70th birthday**

2012–2014 **Founder of the ISPED Ph.D. students** (weekly) **seminar**

2009–2010 **President** (formerly Secretary General) **of the ENSAI Business Networking Forum**
Responsible for organizing the yearly networking event between companies and ENSAI students

2009 **Vice President of the ENSAI Student Council**
Organize and coordinate associative activities and social life at the school

Publications

▷ Books:

Desquilbet L, Granger S, Hejblum BP, Legrand A, Pernot P & Rougier N. *Vers une recherche reproductible : Faire évoluer ses pratiques*. Urfist de Bordeaux, 2019.

▷ Preprints:

Zhou D, Tong H, Wang L, Liu S, Xiong X, Gan Z, Griffier R, Hejblum B, Liu Y, Hong C, Bonzel C, Cai T, Pan K, Ho Y, Costa L, Panickan VA, Gaziano JM, Mandl K, Jouhet V, Thiebaut R, Xia Z, Cho K, Liao K & Cai T. Representation learning to advance multi-institutional studies with electronic health record data. *arXiv* 2502.08547, 2025. DOI: 10.48550/arXiv.2502.08547

Hivert B, Agniel D, Thiébaut R & Hejblum BP. Running in circles: Practical limitations for real-life application of data fission and data thinning in post-clustering differential analysis. *arXiv* 2405.13591, 2024. DOI: 10.48550/arXiv.2405.13591

Gauthier M, Agniel D, Thiébaut R & Hejblum BP. Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis. *bioRxiv* 2021.05.21.445165, 2021. DOI: 10.1101/2021.05.21.445165

Villain L, Ferté T, Thiébaut R & Hejblum BP. Gene set analysis for time-to-event outcome with the generalized berk-jones statistic. *bioRxiv* 2021.09.07.459329, 2021. DOI: 10.1101/2021.09.07.459329

Hejblum BP, Kunzmann K, Lavagnini E, Hutchinson A, Robertson DS, Jones SC & Eckes-Shephard AH. Realistic and robust reproducible research for biostatistics. *Preprints* 2020060002, 2020. DOI: 10.20944/preprints202006.0002.v1

▷ Published or in press: (* indicates equal contibution)

Ferté T, Ba K, Dutartre D, Legrand P, Jouhet V, Thiébaut R, Hinaut X & Hejblum BP. Reservoir Computing in R: A Tutorial for Using reservoirnet to Predict Complex Time-Series. *Computo*, 2025. DOI: 10.57750/arxn-6z34

Bigot J, Freulon P, Hejblum BP & Leclaire A. On the potential benefits of entropic regularization for smoothing wasserstein estimators. *Electronic Journal of Statistics* 19(2):3867-3894, 2025. DOI: 10.1214/25-EJS2430

Chouleur T, Etchegaray C, Villain L, Lesur A, Ferté T, Rossi M, Andrique L, Simoncini C, Giacobbi A, Gambaretti M, Lopci E, Fernades B, Dittmar G, Bjerkvig R, Hejblum B, Thiébaut R, Saut O, Bello L & Bikfalvi A. A strategy for multimodal integration of transcriptomics, proteomics, and radiomics data for the prediction of recurrence in patients with IDH-mutant gliomas. *International Journal of Cancer* 157(3):573-587, 2025. DOI: 10.1002/ijc.35441

Hughes A, Parast L, Thiébaut R & Hejblum BP. RISE: Two-stage rank-based identification of high-dimensional surrogate markers applied to vaccinology. *Statistics in Medicine* 44(20-22):e70241, 2025. DOI: 10.1002/sim.70241

Hivert B, Agniel D, Thiébaut R & Hejblum BP. Post-clustering difference testing: Valid inference and practical considerations with applications to ecological and biological data. *Computational Statistics & Data Analysis* 107916, 2024. DOI: 10.1016/j.csda.2023.107916

Hejblum BP, Ba K, Thiébaut R & Agniel D. Neglecting normalization impact in semi-synthetic RNA-Seq data simulation generates artificial false positives. *Genome Biology* 25:281, 2024. DOI: 10.1186/s13059-024-03231-9

Collin A, Hejblum BP, Vignals C, Lehot L, Thiébaut R, Moireau P & Prague M. Using population based kalman estimator to model COVID-19 epidemic in france: Estimating the effects of non-pharmaceutical interventions on the dynamics of epidemic. *International Journal of Biostatistics* 20(1):13-41, 2024. DOI: 10.1515/ijb-2022-0087

White E, Papagno L, Samri A, Sugata K, Hejblum B, Henry AR, Rogan DC, Darko S, Recordon-Pinson P, Dudoit Y, Llewellyn-Lacey LA, Buseyne F, Migueles SA, Price DA, Andreola M, Satou Y, Thiébaut R, Katlama C, Autran B, C DD & Appay V. Clonal succession after prolonged antiretroviral therapy rejuvenates CD8+ t cell responses against HIV-1. *Nature Immunology* 25:1555-1564, 2024. DOI: 10.1038/s41590-024-01931-9

Freulon P, Bigot J & Hejblum BP. CytOpT: Optimal Transport with Domain Adaptation for Interpreting Flow Cytometry data. *Annals of Applied Statistics* 17(2):1086-1104, 2023. DOI: 10.1214/22-AOAS1660

Agniel D, Hejblum BP, Thiébaut R & Parast L. Doubly-robust evaluation of high-dimensional surrogate markers. *Biostatistics* 24(4):985-999, 2023. DOI: 10.1093/biostatistics/kxac020

Blengio F, Hocini H, Richert L, Lefebvre C, Durand M, Hejblum B, Tisserand P, McLean C, Luhn K, Thiebaut R & Lévy Y. Identification of early gene expression profiles associated with long-lasting antibody responses to the Ebola vaccine Ad26. ZEBOV/MVA-BN-Filo. *Cell Reports* 42(9):113101, 2023. DOI: 10.1016/j.celrep.2023.113101

Vignals C, Hejblum BP & Prague M. Modéliser la COVID-19: De la population à l'individu. *Interstices*, 2023.

Thiébaut R, Hejblum B, Mougin F, Tzourio C & Richert L. ChatGPT and beyond with artificial intelligence (AI) in health: Lessons to be learned. *Joint Bone Spine* 90(5):105607, 2023. DOI: 10.1016/j.jbspin.2023.105607

Ferté T, Jouhet V, Greffier R, Hejblum BP & Thiébaut R. The benefit of augmenting open data with clinical data-warehouse EHR for forecasting SARS-CoV-2 hospitalizations in Bordeaux area, France. *JAMIA open* ooac086, 2022. DOI: 10.1093/jamiaopen/ooac086

Richert L, Lelièvre J, Lacabaratz C, Hardel L, Hocini H, Wiedemann A, Lucht F, Poizot-Martin I, Bauduin C, Diallo A, Rieux V, Durand M, Hejblum BP, Launay O, Thiébaut R, Lévy Y, on behalf of the ANRS VRI01 Study group. T-cell immunogenicity, gene expression profile and safety of four heterologous prime-boost combinations of HIV vaccine candidates in healthy volunteers - results of the randomized multi-arm phase i/II ANRS VRI01 trial. *Journal of Immunology* 208(12):2663-2674, 2022. DOI: 10.4049/jimmunol.2101076

Rinchai D, Deola S, Zoppoli G, Ahamed Kabeer BS, Taleb S, Pavlovski I, Maacha S, Gentilcore G, Toufiq M, Mathew L, Liu L, Vempalli FR, Mubarak G, Lorenz S, Sivieri I, Cirmena G, Dentone C, Cuccarolo P, Giacobbe D, Baldi F, Garbarino A, Cigolini B, Cremonesi P, Bedognetti M, Ballestrero A, Bassetti M, Hejblum BP, Augustine T, Van Panhuys N, Thiébaut R, Branco R, Chew T, Shojaei M, Short K, Feng C, PREDICT-19 consortium, Zughaier SM, De Maria A, Tang B, Ait Hssain A, Bedognetti D, Grivel J, Chaussabel D. High–temporal resolution profiling reveals distinct immune trajectories following the first and second doses of COVID-19 mRNA vaccines. *Science Advances* 8(45):eabp9961, 2022. DOI: 10.1126/sciadv.abp9961

Acar N, Merle BMJ, Ajana S, He Z, Grégoire S, Hejblum BP, Martine L, Buaud B, Bron AM, Creuzot-Garcher CP, Korobelnik J, Berdeaux O, Jacqmin-Gadda H, Bretillon L, Delcourt C, for the Biomarkers of Lipid Status And metabolism in Retinal ageing (BLISAR) Study Group. Predicting the retinal content in omega-3 fatty acids for age-related macular-degeneration. *Clinical and Translational Medicine* 11(7):e404, 2021. DOI: 10.1002/ctm2.404

Lefèvre-Arbogast S, Hejblum BP, Helmer C, Klose C, Manach C, Low DY, Urpi-Sarda M, Andres-Lacueva C, González-Domínguez R, Aigner L, Altendorfer B, Lucassen PJ, Ruigrok SR, De Lucia C, Du Preez A, Proust-Lima C, Thuret S, Korosi A & Samieri C. Early signature in the blood lipidome associated with subsequent cognitive decline in the elderly: A case-control analysis nested within the three-city cohort study. *EBioMedicine* 64:103216, 2021. DOI: 10.1016/j.ebiom.2021.103216

Colas C, Hejblum B, Rouillon S, Thiébaut R, Oudeyer P, Moulin-Frier C & Prague M. EpidemiOptim: A toolbox for the optimization of control policies in epidemiological models. *Journal of Artificial Intelligence Research* 71:479-519, 2021. DOI: 10.1613/jair.1.12588

Zhang HG*, Hejblum BP*, Weber G, Palmer N, Churchill S, Szolovits P, Murphy S, Liao K, Kohane I & Cai T. ATLAS: An automated association test using probabilistically linked health records with application to genetic studies. *Journal of the American Medical Informatics Association* 28(12):2582-2592, 2021. DOI: 10.1093/jamia/ocab187

Ferté T, Cossin S, Schaeverbeke T, Barnetche T, Jouhet V & Hejblum BP. Automatic phenotyping of electronical health record: PheVis algorithm. *Journal of Biomedical Informatics* 117:103746, 2021. DOI: 10.1016/j.jbi.2021.103746

Ajana S, Cougnard-Grégoire A, Colijn J, Merle BM, Verzijden T, Jong P, Hofman A, EYE-RISK Consortium, Vingerling J, Hejblum BP, Korobelnik J, Meester-Smoor M, Jacqmin-Gadda H, Klaver C, Delcourt C. Predicting progression to advanced age-related macular degeneration from clinical, genetic and lifestyle factors using machine learning. *Ophthalmology* 128(4):587-597, 2021. DOI: 10.1016/j.ophtha.2020.08.031

Philipps V, Hejblum BP, Prague M, Commenges D & Proust-Lima C. Robust and efficient optimization using a marquardt- levenberg algorithm with r package marqLevAlg. *The R Journal* 13(2):365-379, 2021. DOI: 10.32614/RJ-2021-089

Lin L & Hejblum BP. Bayesian mixture models for cytometry data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 13:e1535, 2021. DOI: 10.1002/wics.1535

Lévy Y, Wiedemann A*, Hejblum BP*, Durand M, Lefebvre C, Surénaud M, Lacabaratz C, Perreau M, Foucat E, Déchenaud M, Tisserand P, Blengio F, Hivert B, Gauthier M, Cervantes-Gonzalez M, Bachelet D, Laouénan C, Bouadma L, Timsit J, Yazdanpanah Y, Pantaleo G, Hocini H & Thiébaut R. CD177, a specific marker of neutrophil activation, is associated with coronavirus disease 2019 severity and death. *iScience* 24(7):102711, 2021. DOI: 10.1016/j.isci.2021.102711

Bouadma L, Wiedemann A, Patrier J, Surenaud M, Wicky P, Foucat E, Diehl J, Hejblum BP, Sinnah F, Montmollin E, Lacabaratz C, Thiébaut R, Timsit J & Lévy Y. Immune alterations during SARS-CoV-2-related acute respiratory distress syndrome. *Journal of Clinical Immunology* 40:1082-1092, 2020. DOI: 10.1007/s10875-020-00839-x

Lhomme E, Hejblum BP, Lacabaratz C, Wiedemann A, Lelièvre J, Lévy Y, Thiébaut R & Richert L. Analyzing cellular immunogenicity in vaccine clinical trials: A new statistical method including non-specific responses for accurate estimation of vaccine effect. *Journal of Immunological Methods* 477:112711, 2020. DOI: 10.1016/j.jim.2019.112711

Gauthier M, Agniel D, Thiébaut R & Hejblum BP. Dearseq: A variance component score test for RNA-seq differential analysis that effectively controls the false discovery rate. *NAR Genomics and Bioinformatics* 2(4):lqaa093, 2020. DOI: 10.1093/nargab/lqaa093

Wiedemann A, Foucat E, Hocini H, Lefebvre C, Hejblum BP, Durand M, Krüger M, Keita AK, Ayouba A, Mély S, Fernandez J, Touré A, Fourati S, Lévy-Marchal C, Raoul H, Delaporte E, Koivogui L, Thiébaut R, Lacabaratz C, Lévy Y, PostEboGui Study Group. Long-lasting severe immune dysfunction in ebola virus disease survivors. *Nature Communications* 11:3730, 2020. DOI: 10.1038/s41467-020-17489-7

Chan SF, Hejblum BP, Chakrabortty A & Cai T. Semi-supervised estimation of covariance with application to phenome-wide association studies with electronic medical records data. *Statistical Methods in Medical Research* 29:455-465, 2020. DOI: 10.1177/0962280219837676

Hejblum BP, Alkhassim C, Gottardo R, Caron F & Thiébaut R. Sequential dirichlet process mixture of skew t-distributions for model-based clustering of flow cytometry data. *Annals of Applied Statistics* 13(1):638-660, 2019. DOI: 10.1214/18-AOAS1209

Ajana S, Niyazi A, Bretillon L, Hejblum BP, Jacqmin-Gadda H & Cécile D. Benefits of dimension reduction in penalized regression methods for high dimensional grouped data: A case study in low sample size. *Bioinformatics* 35:3628-3634, 2019. DOI: 10.1093/bioinformatics/btz135

Thiébaut R, Hejblum BP, Hocini H, Bonnabau H, Skinner J, Montes M, Lacabaratz C, Richert L, Palucka K, Banchereau J & Levy Y. Gene expression signatures associated with immune and virological responses to therapeutic vaccination with dendritic cells in HIV-infected individuals. *Frontiers in Immunology* 10:874, 2019. DOI: 10.3389/fimmu.2019.00874

Low DY, Lefèvre-Arbogast G, Urpi-Sarda M, Micheau P, Petera M, Centeno D, Durand S, Estelle P, Korosi A, Lucassen PJ, Aigner L, Proust-Lima C, Hejblum BP, Helmer C, Andres-Lacueva C, Thuret S, Samieri C & Manach C. Diet-related metabolites associated with cognitive decline revealed by untargeted metabolomics in a prospective cohort. *Molecular Nutrition & Food Research* 63:1900177, 2019. DOI: 10.1002/mnfr.201900177

Hejblum BP, Cui J, Lahey LJ, Cagan A, Sparks JA, Sokolove J, Cai T & Liao KP. Association between anti-citrullinated fibrinogen antibodies and coronary artery disease in rheumatoid arthritis. *Arthritis Care & Research* 70:1113-1117, 2018. DOI: 10.1002/acr.23444

Commenges D, Alkhassim C, Gottardo R, Hejblum BP & Thiébaut R. Cytometree: A binary tree algorithm for automatic gating in cytometry analysis. *Cytometry: Part A* 93(11):1132-1140, 2018. DOI: 10.1002/cyto.a.23601

Sinnott JA, Cai F, Yu S, Hejblum BP, Hong C, Kohane IS & Liao KP. PheProb: probabilistic phenotyping using diagnosis codes to improve power for genetic association studies. *Journal of the American Medical Informatics Association* 25(10):1359-1365, 2018. DOI: 10.1093/jamia/ocy056

Lefèvre-Arbogast S, Gaudout D, Bensalem J, Letenneur L, Dartigues J, Hejblum BP, Féart C, Delcourt C & Samieri C. Pattern of polyphenol intake and the long-term risk of dementia in older persons. *Neurology*, 2018. DOI: 10.1212/WNL.0000000000005607

Hejblum BP, Weber GM, Liao KP, Palmer NP, Churchill S, Shadick NA, Szolovits P, Murphy SN, Kohane IS & Cai T. Probabilistic record linkage of de-identified research datasets with discrepancies using diagnosis codes. *Scientific Data* 6:180298, 2018. DOI: 10.1038/sdata.2018.298

Neykov M, Hejblum BP & Sinnott JA. Kernel machine score test for pathway analysis in the presence of semi-competing risks. *Statistical Methods in Medical Research* 27(4):1099-1114, 2018. DOI: 10.1177/0962280216653427

Agniel D & Hejblum BP. Variance component score test for time-course gene set analysis of longitudinal RNA-seq data. *Biostatistics* 18(4):589-604, 2017. DOI: 10.1093/biostatistics/kxx005

Rechtien A, Richert L, Lorenzo H, Martrus G, Hejblum B, Dahlke C, Kasonta R, Zinser M, Stubbe H, Matschl U, Lohse A, Krähling V, Eickmann M, Becker S, Agnandji ST, Krishna S, Kremsner PG, Brosnahan JS, Bejon P, Njuguna P, Addo MM, Becker S, Krähling V, Siegrist CA, Huttner A, Kieny MP, Moorthy V, Fast P, Savarese B, Lapujade O, Thiébaut R, Altfeld M & Addo M. Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV. *Cell Reports* 20(9):2251-2261, 2017. DOI: 10.1016/j.celrep.2017.08.023

Liao KP*, Sparks JA*, Hejblum BP, Kuo I, Cui J, Lahey LJ, Cagan A, Gainer VS, Liu W, Cai TT, Sokolove J & Cai T. Phenome-wide association study of autoantibodies to citrullinated and noncitrullinated epitopes in rheumatoid arthritis. *Arthritis & Rheumatology* 69(4):742-749, 2017. DOI: 10.1002/art.39974

Liquet B, De Micheaux PL, Hejblum BP & Thiébaut R. Group and sparse group partial least square approaches applied in genomics context. *Bioinformatics* 32(1):35-42, 2016. DOI: 10.1093/bioinformatics/btv535

Hejblum BP, Skinner J & Thiébaut R. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data. *PLOS Computational Biology* 11(6):e1004310, 2015. DOI: 10.1371/journal.pcbi.1004310

Furman D*, Hejblum BP*, Simon N, Jojic V, Dekker CL, Thiébaut R, Tibshirani RJ & Davis MM. Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. *Proceedings of the National Academy of Sciences* 111(2):869-874, 2014. DOI: 10.1073/pnas.1321060111

Thiébaut R, Hejblum BP & Richert L. The analysis of "Big Data" in clinical research. *Revue d'Épidemiologie et de Santé Publique* 62(1):1-4, 2014. DOI: 10.1016/j.respe.2013.12.021

Commenges D & Hejblum BP. Evidence synthesis through a degradation model applied to myocardial infarction. *Lifetime Data Analysis* 19(1):1-18, 2013. DOI: 10.1007/s10985-012-9227-3

Selected communications

▷ Oral communications: (* indicates invited talks)

Portugal 2024* Hejblum B, Gauthier M, Fallet S, Thiébaut R, Agniel D, Conditional independence testing by comparing empirical conditional cumulative distribution functions, *International Symposium on Non-parametric Statistics (ISNPS) 2024*, Braga.

Mexico 2022* Hejblum B, Parast L, Agniel D, Transcriptomics: a potential early surrogate for vaccine response ?, *BIRS-CMO 22w5184*, Oaxaca.

Latvia 2022 Hejblum B, Gauthier M, Ba K, Thiébaut R, Agniel D, Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis, *31st International Biometric Conference*, Riga.

France 2022* Hejblum B, Machine learning approaches for the analysis of bulk and single-cell RNA-seq data, *4th GenMed workshop on Medical Genomics*, Paris.

Germany 2022* Hejblum B, Teaching Bayesian statistics during a pandemic, *German Association for Medical Informatics, Biometry and Epidemiology (GMDS) Teaching & Didactics workshop*, Saarbrücken.

France 2021* Prague M, Collin A, Wittkop L, Dutartre D, Clairon Q, Moireau P, Thiébaut R, Hejblum B, Leveraging random effects to estimate the impact of NPIs on epidemic dynamics across French regions, *8th Channel Network Conference of the International Biometric Society*, Paris.

France 2021* Hejblum B, Clustering of flow cytometry data using non parametric Bayesian modeling, *Séminaire LMBA*, Vannes.

France 2021* Hejblum B, Distribution-free complex hypothesis testing for single-cell RNA-seq differential expression analysis, *Statistical Methods for Post-Genomic Data (SMPGD) – 2021*, online.

France 2020 Hejblum B, Gauthier M, Thiébaut R, Agniel D, A variance component score test for flexible RNA-Seq data differential analysis, *Statistical Methods for Post-Genomic Data (SMPGD) – 2020*, Paris.

France 2019* Hejblum B, Montani I, Leffondré K, Diallo G, Mougin F, Pariente A, Richert L, Thiessard F, Joly P, Alioum A, Tzourio C, Thiébaut R, Enseigner la science des données en santé publique, *Colloque Francophone International sur l'Enseignement de le Statistique (CFIES)*, Strasbourg.

France 2019* Hejblum B, Gauthier M, Thiébaut R, Agniel D, Controlling Type-I error in RNA-seq differential analyses through a variance component score test with an application to tuberculosis infection, *Séminaire de l'équipe de Statistique de l'Institut de Recherche MAthématique de Rennes (IRMAR)*, Rennes.

UK 2019* Hejblum B, Kirk PDW, Scaling up nonparametric Bayesian clustering with MCMC for big data applications, *12th International Conference of the ERCIM WG on Computational and Methodological Statistics*, Londres.

Taiwan 2019* Hejblum B, Gauthier M, Thiébaut R, Agniel D, A variance component score test applied to RNA-Seq differential analysis, *3rd EcoSta Conference*, Taichung.

France 2019 Hejblum B, Lhomme E, Thiébaut R, Richert L, VICI: a Shiny app for accurate estimation of Vaccine Induced Cellular Immunogenicity with bivariate modeling, *UseR! 2019*, Toulouse.

France 2018* Hejblum B, Gauthier M, Thiébaut R, Agniel D, Controlling type-I error and false discoveries in RNA-seq differential analyses through a variance component score test, *Bioinfo-Biostat GenoToul Annual Day*, Toulouse.

Spain 2018 Hejblum B, Agniel D, A variance component score test for RNA-seq differential analysis in vaccine trials, *29th International Biometric Conference*, Barcelona.

UK 2017* Hejblum, Alkhassim, Gottardo, Caron, Thiébaut, Dirichlet Process Mixtures of Multivariate Skew t-distributions for Unsupervised Clustering of Cell Populations from Flow-Cytometry Data, *BSU invited Seminar*, Cambridge.

Spain 2017 Hejblum B, Agniel D, Type I error and false discovery rate control in RNA-seq differential analyses through a variance component score test, *38th Annual Conference of the International Society for Clinical Biostatistics*, Vigo.

USA 2016 Hejblum B, Agniel D, Time-course Gene Set Analysis of longitudinal RNA-seq data, *ENAR 2016 Spring Meeting*, Austin (TX).

Italy 2014 Hejblum B, Caron F, Thiébaut R, Bayesian analysis of time-course flow cytometry data with Dirichlet process mixture modeling, *27th International Biometric Conference*, Florence.

France 2014 Hejblum B, Genuer R, Thiébaut R, Variable selection in high-dimensional dataset: comparison of sPLS with other approaches in an HIV vaccine trial, *8th International Conference on Partial Least Squares and Related Methods*, Paris.

France 2014* Hejblum B, Caron F, Thiébaut R, Bayesian nonparametric modeling of flow cytometry data with Dirichlet process mixtures, *Ph.D. students working group of the LSTA (Laboratoire de Statistique Théorique et Appliquée) in Paris 6 University*, Paris.

Spain 2013 Thiébaut R, Hejblum B, Skinner J, Montes M, Chêne G, Palucka K, Banchereau J, Lévy Y, Integrative Analysis of Responses to Dendritic-Cell Vaccination Identifies Signatures Correlated with Control of HIV Replication: The DALIA Trial, *AIDS Vaccine 2013, AIDS Research and Human Retroviruses*, Barcelone.

Norway 2012 Hejblum B, Skinner J, Thiébaut R, Application of Gene Set Analysis of Time-Course gene expression in a HIV vaccine trial, *33rd Annual Conference of the International Society for Clinical Biostatistics*, Bergen.

▷ Written communications

USA 2015 Hejblum B, Cai T, Weber G, Probabilistic Patient Linkage Algorithms for PIC-SURE, *BD2K all Hands Meeting 2015*, Bethesda (MD).

UK 2014 Hejblum B, Caron F, Thiébaut R, Hierarchical Analysis of Time-Course Flow Cytometry Data with Dirichlet Process Mixture Modeling, *Medical Research Council Conference on Biostatistics in celebration of the MRC Biostatistics Unit's centenary year*, Cambridge.