Summary

Dr. Zuguang Gu

Staff Scientist in National Center of Tumor Diseases, German Cancer Research Center, Heidelberg
Email: zuguang.gu@gmail.com
Website: https://jokergoo.github.io/
ORCiD: https://orcid.org/0000-0002-7395-8709
Google Scholar: https://scholar.google.de/citations?user=zheH1qkAAAAJ
GitHub: https://github.com/jokergoo/

I obtained my bachelor's, master's and doctoral degrees from Nanjing University, China in 2005, 2008 and 2012. From 2013, I have been engaged in precision medicine research at the German Cancer Research Center (DKFZ) for more than 10 years. I have rich experience and in-depth understanding of the analysis of various types of omics data. I have developed statistical methods and popular biological software, which are widely used in the bioinformatics field with strong international influence.

During the 11 years of working in DKFZ, I mainly use the R programming language and statistics as tools, take cancers or specific diseases as the research objects, and have the following achievements:

1) As an active R package developer, I have independently developed many biological software packages for data analysis and visualization (18 have been published. The complete list can be found from https://jokergoo.github.io/software/). Many of them are widely used in the bioinformatics field worldwide. The total downloads of my software have exceeded 5.8 million. Two successful software are *ComplexHeatmap* and *circlize* which have become standard visualization tools in the field.
2) In the Heidelberg Institute of Personized Oncology (HIPO) of DKFZ, as a "core bioinformatician", I participated in more than 20 projects on specific cancers or diseases. I have rich experience in the methods and tools for various types of omics data. At the same time, I also developed new methods and workflows for customized analysis. I am involved in the clinical practice of precision medicine at the National Center of Tumor Diseases (NCT Heidelberg), providing data analysis and software development services.

By February 2024, I have published 46 academic papers, in 21 of which I am the first author or corresponding author. My first-author paper journals include *Bioinformatics* (8 papers), *Nucleic Acid Research*, *Briefings in Bioinformatics*, *Journal of Systems and Software*, *Nature Immunology*, and *Nature Communications*. My Google Scholar citations exceed 12200 of which the total citations of my first-author papers exceed 9600. My *h*-index is 29, and I am among the top 2% of highly cited scientists in the world in 2021 and 2022. I have reported 7 times on the Bioconductor Conference, and 2 times on the useR! Conference, the largest international conference of the R language.

I have strong influence in the bioinformatics community. I am enthusiastic on sharing knowledge of software development and data analysis skills on my blog (https://jokergoo.github.io/post/) and on Twitter (https://twitter.com/jokergoo_gu, >3K followers). From internet (Google, YouTube or on the

Chinese social media) it is also easy to find a large number of articles and tutorials introducing my software, made and shared by the community.

Besides that, I have teaching and training experience in Heidelberg University and the global bioinformatics community. I am the Bioconductor certified Carpentry instructor and the developer of the official Bioconductor training courses. In DKFZ, I also have experience of supervising students.

Curriculum Vitae

Experience

National Center for Tumor Diseases, German Cancer Research Center (DKFZ-NCT)

  Staff Scientist (permanent)   2022-present

Heidelberg Center for Personalized Oncology, German Cancer Research Center (DKFZ-HIPO)

  Postdoctoral researcher   2013-2021

ProtaGene GmbH

  Bioinformatics consultant (Minijob, Permanent)   2022-present

GeneWerk GmbH

  Bioinformatics consultant (Minijob)   2017-2021

Education

Nanjing University, China, Ph.D. Biochemistry and Molecular Biology   2009-2012

Nanjing University, China, M.D. Biochemistry and Molecular Biology   2005-2008

Nanjing University, China, B.D. Biochemistry and Molecular Biology   2001-2005

List of publications

First/corresponding-author publications

[#]corresponding author, *co-first author.

1. **Zuguang Gu[#]**, simona：a comprehensive R package for semantic similarity analysis on bio-ontologies. bioRxiv 2023.
2. **Zuguang Gu[#]**, Daniel Hübschmann, rGREAT: an R/bioconductor package for functional enrichment on genomic regions. Bioinformatics 2023.
3. **Zuguang Gu[#]**, On the dependency heaviness of CRAN/Bioconductor ecosystem. Journal of Systems and Software 2023.
4. **Zuguang Gu[#]**, Daniel Hübschmann, Pkgndep: a tool for analyzing dependency heaviness of R packages. Bioinformatics 2022.
5. **Zuguang Gu[#]**, Daniel Hübschmann, simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results. Genomics, Proteomics & Bioinformatics 2022.

6. **Zuguang Gu**[#], Daniel Hübschmann, Improve consensus partitioning via a hierarchical procedure. Briefings in Bioinformatics 2022.
7. **Zuguang Gu**[#], Daniel Hübschmann, Make Interactive Complex Heatmaps in R. Bioinformatics 2022.
8. **Zuguang Gu**[#], Daniel Hübschmann, spiralize: an R package for visualizing data on spirals. Bioinformatics 2022.
9. **Zuguang Gu**[#], Complex heatmap visualization. iMeta 2022.
10. Nina Hensel*, **Zuguang Gu***, Sagar*, Dominik Wieland, Katharina Jechow, Janine Kemming, Sian Llewellyn-Lacey, Emma Gostick, Oezlem Sogukpinar, Florian Emmerich, David A Price, Bertram Bengsch, Tobias Boettler, Christoph Neumann-Haefelin, Roland Eils, Christian Conrad, Ralf Bartenschlager, Dominic Grün, Naveed Ishaque, Robert Thimme, Maike Hofmann, Memory-like HCV-specific CD8+ T cells retain a molecular scar after cure of chronic HCV infection. Nature Immunology 2021.
11. **Zuguang Gu**, Matthias Schlesner, Daniel Hübschmann, cola: an R/Bioconductor package for consensus partitioning through a general framework. Nucleic Acids Research 2021.
12. Yonghe Wu*, Michael Fletcher*, **Zuguang Gu***, Qi Wang, Barbara Costa, Anna Bertoni, Ka-Hou Man, Magdalena Schlotter, Jörg Felsberg, Jasmin Mangei, Martje Barbus, Ann-Christin Gaupel, Wei Wang, Tobias Weiss, Roland Eils, Michael Weller, Haikun Liu, Guido Reifenberger, Andrey Korshunov, Peter Angel, Peter Lichter, Carl Herrmann, Bernhard Radlwimmer, Glioblastoma epigenome profiling identifies SOX10 as a master regulator of molecular tumour subtype. Nature Communications 2020.
13. **Zuguang Gu**, Roland Eils, Matthias Schlesner, Naveed Ishaque, EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations. BMC Genomics 2018.
14. **Zuguang Gu**, Roland Eils, Matthias Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016.
15. **Zuguang Gu**, Roland Eils, Matthias Schlesner, HilbertCurve: an R/Bioconductor package for high-resolution visualization of genomic data. Bioinformatics 2016.
16. **Zuguang Gu**, Roland Eils, Matthias Schlesner, gtrellis: an R/Bioconductor package for making genome-level Trellis graphics. BMC Bioinformatics 2016.
17. **Zuguang Gu**, Lei Gu, Roland Eils, Matthias Schlesner, Benedikt Brors, circlize Implements and enhances circular visualization in R. Bioinformatics 2014.
18. **Zuguang Gu**, Jin Wang, CePa: an R package for finding significant pathways weighted by multiple network centralities. Bioinformatics 2013.
19. **Zuguang Gu**, Jialin Liu, Kunming Cao, Junfeng Zhang, Jin Wang, Centrality-based pathway enrichment: a systematic approach for finding significant pathways dominated by key genes. BMC Systems Biology 2012.
20. **Zuguang Gu**, Chenyu Zhang, Jin Wang, Gene regulation is governed by a core network in hepatocellular carcinoma. BMC Systems Biology 2012.
21. **Zuguang Gu**, Jie Li, Song Gao, Ming Gong, Junling Wang, Hua Xu, Chenyu Zhang, Jin Wang, InterMitoBase: an annotated database and analysis platform of protein-protein interactions for human mitochondria. BMC Genomics 2011.

Co-author publications

22. Josefine Radke, Naveed Ishaque, Randi Koll, **Zuguang Gu**, Elisa Schumann, Lina Sieverling, Sebastian Uhrig, Daniel Hübschmann, Umut H Toprak, Cristina López, Xavier Pastor Hostench, Simone Borgoni, Dilafruz Juraeva, Fabienne Pritsch, Nagarajan Paramasivam, Gnana Prakash Balasubramanian, Matthias Schlesner, Shashwat Sahay, Marc Weniger, Debora Pehl, Helena Radbruch, Anja Osterloh, Agnieszka Korfel, Martin Misch, Julia Onken, Katharina Faust, Peter Vajkoczy, Dag Moskopp, Yawen Wang, Andreas Jödicke, Lorenz Trümper, Ioannis

Anagnostopoulos, Dido Lenze, Ralf Küppers, Michael Hummel, Clemens A Schmitt, Otmar D Wiestler, Stephan Wolf, Andreas Unterberg, Roland Eils, Christel Herold-Mende, Benedikt Brors, Reiner Siebert, Stefan Wiemann, Frank L Heppner, The genomic and transcriptional landscape of primary central nervous system lymphoma. Nature Communications 2022.

23. Marina Laplana, Matthias Bieg, Christian Faltus, Svitlana Melnik, Olga Bogatyrova, **Zuguang Gu**, Thomas Muley, Michael Meister, Hendrik Dienemann, Esther Herpel, Christopher I Amos, Matthias Schlesner, Roland Eils, Christoph Plass, Angela Risch, Differentially methylated regions within lung cancer risk loci are enriched in deregulated enhancers. Epigenetics 2022.

24. Mengyu Tu, Lukas Klein, Elisa Espinet, Theodoros Georgomanolis, Florian Wegwitz, Xiaojuan Li, Laura Urbach, Adi Danieli-Mackay, Stefan Küffer, Kamil Bojarczuk, Athanasia Mizi, Ufuk Günesdogan, Björn Chapuy, **Zuguang Gu**, Albrecht Neesse, Uday Kishore, Philipp Ströbel, Elisabeth Hessmann, Stephan A Hahn, Andreas Trumpp, Argyris Papantonis, Volker Ellenrieder, Shiv K Singh, TNF-α-producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. Nature Cancer 2021.

25. Ana Maia, **Zuguang Gu**, André Koch, Mireia Berdiel-Acer, Rainer Will, Matthias Schlesner, Stefan Wiemann, IFNβ1 secreted by breast cancer cells undergoing chemotherapy reprograms stromal fibroblasts to support tumour growth after treatment. Molecular Oncology 2021.

26. Daniel Hübschmann, Lea Jopp-Saile, Carolin Andresen, Stephen Krämer, **Zuguang Gu**, Christoph E Heilig, Simon Kreutzfeldt, Veronica Teleanu, Stefan Fröhling, Roland Eils, Matthias Schlesner, Analysis of mutational signatures with yet another package for signature analysis. Genes, Chromosomes & Cancer 2021.

27. Elisa Espinet, **Zuguang Gu***, Charles D Imbusch*, Nathalia A Giese, Magdalena Büscher, Mariam Safavi, Silke Weisenburger, Corinna Klein, Vanessa Vogel, Mattia Falcone, Jacob Insua-Rodríguez, Manuel Reitberger, Vera Thiel, Steffi O Kossi, Alexander Muckenhuber, Karnjit Sarai, Alex Y L Lee, Elyne Backx, Soheila Zarei, Matthias M Gaida, Manuel Rodríguez-Paredes, Elisa Donato, Hsi-Yu Yen, Roland Eils, Matthias Schlesner, Nicole Pfarr, Thilo Hackert, Christoph Plass, Benedikt Brors, Katja Steiger, Dieter Weichenhan, H Efsun Arda, Ilse Rooman, Janel L Kopp, Oliver Strobel, Wilko Weichert, Martin R Sprick, Andreas Trumpp, Aggressive PDACs Show Hypomethylation of Repetitive Elements and the Execution of an Intrinsic IFN Program Linked to a Ductal Cell of Origin. Cancer Discovery 2021.

28. Simon Raffel, Daniel Klimmeck, Mattia Falcone, Aykut Demir, Alireza Pouya, Petra Zeisberger, Christoph Lutz, Marco Tinelli, Oliver Bischel, Lars Bullinger, Christian Thiede, Anne Flörcken, Jörg Westermann, Gerhard Ehninger, Anthony D Ho, Carsten Müller-Tidow, **Zuguang Gu**, Carl Herrmann, Jeroen Krijgsveld, Andreas Trumpp, Jenny Hansson, Quantitative proteomics reveals specific metabolic features of acute myeloid leukemia stem cells. Blood 2020.

29. Enoch B Antwi, Ada Olins, Vladimir B Teif, Matthias Bieg, Tobias Bauer, **Zuguang Gu**, Benedikt Brors, Roland Eils, Donald Olins, Naveed Ishaque, Whole-genome fingerprint of the DNA methylome during chemically induced differentiation of the human AML cell line HL-60/S4. Biology Open 2020.

30. Maike Smits, Katharina Zoldan, Naveed Ishaque, **Zuguang Gu**, Katharina Jechow, Dominik Wieland, Christian Conrad, Roland Eils, Catherine Fauvelle, Thomas F Baumert, Florian Emmerich, Bertram Bengsch, Christoph Neumann-Haefelin, Maike Hofmann, Robert Thimme, Tobias Boettler, Follicular T helper cells shape the HCV-specific CD4+ T cell repertoire after virus elimination. The Journal of Clinical Investigation 2020.

31. Stephan M Tirier, Jeongbin Park, Friedrich Preußer, Lisa Amrhein, **Zuguang Gu**, Simon Steiger, Jan-Philipp Mallm, Teresa Krieger, Marcel Waschow, Björn Eismann, Marta Gut, Ivo G Gut, Karsten Rippe, Matthias Schlesner, Fabian Theis, Christiane Fuchs, Claudia R Ball, Hanno Glimm, Roland Eils, Christian Conrad, Pheno-seq - linking visual features and gene expression in 3D cell culture systems. Scientific Reports 2019.

32. Calvin Wing Yiu Chan, **Zuguang Gu**, Matthias Bieg, Roland Eils, Carl Herrmann, Impact of cancer mutational signatures on transcription factor motifs in the human genome. BMC Medical

Genomics 2019.

33. Nagarajan Paramasivam, Daniel Hübschmann, Umut H Toprak, Naveed Ishaque, Marian Neidert, Daniel Schrimpf, Damian Stichel, David Reuss, Philipp Sievers, Annekathrin Reinhardt, Annika K Wefers, David T W Jones, **Zuguang Gu**, Johannes Werner, Sebastian Uhrig, Hans-Georg Wirsching, Matthias Schick, Melanie Bewerunge-Hudler, Katja Beck, Stephanie Brehmer, Steffi Urbschat, Marcel Seiz-Rosenhagen, Daniel Hänggi, Christel Herold-Mende, Ralf Ketter, Roland Eils, Zvi Ram, Stefan M Pfister, Wolfgang Wick, Michael Weller, Rachel Grossmann, Andreas von Deimling, Matthias Schlesner, Felix Sahm, Mutational patterns and regulatory networks in epigenetic subgroups of meningioma. Acta Neuropathologica 2019.

34. Verena Körber, Jing Yang, Pankaj Barah, Yonghe Wu, Damian Stichel, **Zuguang Gu**, Michael Nai Chung Fletcher, David Jones, Bettina Hentschel, Katrin Lamszus, Jörg Christian Tonn, Gabriele Schackert, Michael Sabel, Jörg Felsberg, Angela Zacher, Kerstin Kaulich, Daniel Hübschmann, Christel Herold-Mende, Andreas von Deimling, Michael Weller, Bernhard Radlwimmer, Matthias Schlesner, Guido Reifenberger, Thomas Höfer, Peter Lichter, Evolutionary Trajectories of IDHWT Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. Cancer Cell 2019.

35. Naveed Ishaque, Mohammed L Abba, Christine Hauser, Nitin Patil, Nagarajan Paramasivam, Daniel Huebschmann, Jörg Hendrik Leupold, Gnana Prakash Balasubramanian, Kortine Kleinheinz, Umut H Toprak, Barbara Hutter, Axel Benner, Anna Shavinskaya, Chan Zhou, **Zuguang Gu**, Jules Kerssemakers, Alexander Marx, Marcin Moniuszko, Miroslaw Kozlowski, Joanna Reszec, Jacek Niklinski, Jürgen Eils, Matthias Schlesner, Roland Eils, Benedikt Brors, Heike Allgayer, Whole genome sequencing puts forward hypotheses on metastasis evolution and therapy in colorectal cancer. Nature Communications 2018.

36. Daniel B Lipka, Tania Witte, Reka Toth, Jing Yang, Manuel Wiesenfarth, Peter Nöllke, Alexandra Fischer, David Brocks, **Zuguang Gu**, Jeongbin Park, Brigitte Strahm, Marcin Wlodarski, Ayami Yoshimi, Rainer Claus, Michael Lübbert, Hauke Busch, Melanie Boerries, Mark Hartmann, Maximilian Schönung, Umut Kilik, Jens Langstein, Justyna A Wierzbinska, Caroline Pabst, Swati Garg, Albert Catalá, Barbara De Moerloose, Michael Dworzak, Henrik Hasle, Franco Locatelli, Riccardo Masetti, Markus Schmugge, Owen Smith, Jan Stary, Marek Ussowicz, Marry M van den Heuvel-Eibrink, Yassen Assenov, Matthias Schlesner, Charlotte Niemeyer, Christian Flotho, Christoph Plass, RAS-pathway mutation patterns define epigenetic subclasses in juvenile myelomonocytic leukemia. Nature Communications 2017.

37. Paul A Northcott, Ivo Buchhalter, A Sorana Morrissy, Volker Hovestadt, Joachim Weischenfeldt, Tobias Ehrenberger, Susanne Gröbner, Maia Segura-Wang, Thomas Zichner, Vasilisa A Rudneva, Hans-Jörg Warnatz, Nikos Sidiropoulos, Aaron H Phillips, Steven Schumacher, Kortine Kleinheinz, Sebastian M Waszak, Serap Erkek, David T W Jones, Barbara C Worst, Marcel Kool, Marc Zapatka, Natalie Jäger, Lukas Chavez, Barbara Hutter, Matthias Bieg, Nagarajan Paramasivam, Michael Heinold, **Zuguang Gu**, Naveed Ishaque, Christina Jäger-Schmidt, Charles D Imbusch, Alke Jugold, Daniel Hübschmann, Thomas Risch, Vyacheslav Amstislavskiy, Francisco German Rodriguez Gonzalez, Ursula D Weber, Stephan Wolf, Giles W Robinson, Xin Zhou, Gang Wu, David Finkelstein, Yanling Liu, Florence M G Cavalli, Betty Luu, Vijay Ramaswamy, Xiaochong Wu, Jan Koster, Marina Ryzhova, Yoon-Jae Cho, Scott L Pomeroy, Christel Herold-Mende, Martin Schuhmann, Martin Ebinger, Linda M Liau, Jaume Mora, Roger E McLendon, Nada Jabado, Toshihiro Kumabe, Eric Chuah, Yussanne Ma, Richard A Moore, Andrew J Mungall, Karen L Mungall, Nina Thiessen, Kane Tse, Tina Wong, Steven J M Jones, Olaf Witt, Till Milde, Andreas Von Deimling, David Capper, Andrey Korshunov, Marie-Laure Yaspo, Richard Kriwacki, Amar Gajjar, Jinghui Zhang, Rameen Beroukhim, Ernest Fraenkel, Jan O Korbel, Benedikt Brors, Matthias Schlesner, Roland Eils, Marco A Marra, Stefan M Pfister, Michael D Taylor, Peter Lichter, The whole-genome landscape of medulloblastoma subtypes. Nature 2017.

38. Marcus Winter, Loreen Thürmann, **Zuguang Gu**, Gerrit Schüürmann, Gunda Herberth, Denise Hinz, Martin von Bergen, Hauke Harms, Sven Olek, Stefan Röder, Michael Borte, Roland Eils,

Irina Lehmann, Saskia Trump, The benzene metabolite 1,4-benzoquinone reduces regulatory T-cell function: A potential mechanism for tobacco smoke-associated atopic dermatitis. The Journal of Allergy and Clinical Immunology 2017.

39. Kortine Kleinheinz, Isabell Bludau, Daniel Hübschmann, Michael Heinold, Philip Kensche, **Zuguang Gu**, Cristina López, Michael Hummel, Wolfram Klapper, Peter Möller, Inga Vater, Rabea Wagener, Benedikt Brors, Reiner Siebert, Roland Eils, Matthias Schlesner, ACEseq–allele specific copy number estimation from whole genome sequencing. BioRXiv 2017.

40. Federico Vita, Cosimo Taiti, Antonio Pompeiano, **Zuguang Gu**, Emilio Lo Presti, Larisa Whitney, Michele Monti, Giuseppe Di Miceli, Dario Giambalvo, Paolo Ruisi, Lorenzo Guglielminetti, Stefano Mancuso, Aromatic and proteomic analyses corroborate the distinction between Mediterranean landraces and modern varieties of durum wheat. Scientific Reports 2016.

41. Saskia Trump, Matthias Bieg, **Zuguang Gu**, Loreen Thürmann, Tobias Bauer, Mario Bauer, Naveed Ishaque, Stefan Röder, Lei Gu, Gunda Herberth, Christian Lawerenz, Michael Borte, Matthias Schlesner, Christoph Plass, Nicolle Diessl, Markus Eszlinger, Oliver Mücke, Horst-Dietrich Elvers, Dirk K Wissenbach, Martin von Bergen, Carl Herrmann, Dieter Weichenhan, Rosalind J Wright, Irina Lehmann, Roland Eils, Prenatal maternal stress and wheeze in children: novel insights into epigenetic regulation. Scientific Reports 2016.

42. Tobias Bauer, Saskia Trump, Naveed Ishaque, Loreen Thürmann, Lei Gu, Mario Bauer, Matthias Bieg, **Zuguang Gu**, Dieter Weichenhan, Jan-Philipp Mallm, Stefan Röder, Gunda Herberth, Eiko Takada, Oliver Mücke, Marcus Winter, Kristin M Junge, Konrad Grützmann, Ulrike Rolle-Kampczyk, Qi Wang, Christian Lawerenz, Michael Borte, Tobias Polte, Matthias Schlesner, Michaela Schanne, Stefan Wiemann, Christina Geörg, Hendrik G Stunnenberg, Christoph Plass, Karsten Rippe, Junichiro Mizuguchi, Carl Herrmann, Roland Eils, Irina Lehmann, Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. Molecular Systems Biology 2016.

43. Jenny Wegert, Naveed Ishaque, Romina Vardapour, Christina Geörg, **Zuguang Gu**, Matthias Bieg, Barbara Ziegler, Sabrina Bausenwein, Nasenien Nourkami, Nicole Ludwig, Andreas Keller, Clemens Grimm, Susanne Kneitz, Richard D Williams, Tas Chagtai, Kathy Pritchard-Jones, Peter van Sluis, Richard Volckmann, Jan Koster, Rogier Versteeg, Tomas Acha, Maureen J O'Sullivan, Peter K Bode, Felix Niggli, Godelieve A Tytgat, Harm van Tinteren, Marry M van den Heuvel-Eibrink, Eckart Meese, Christian Vokuhl, Ivo Leuschner, Norbert Graf, Roland Eils, Stefan M Pfister, Marcel Kool, Manfred Gessler, Mutations in the SIX1/2 pathway and the DROSHA/DGCR8 miRNA microprocessor complex underlie high-risk blastemal type Wilms tumors. Cancer Cell 2015.

44. Lei Gu, Sandra C Frommel, Christopher C Oakes, Ronald Simon, Katharina Grupp, Cristina Y Gerig, Dominik Bär, Mark D Robinson, Constance Baer, Melanie Weiss, **Zuguang Gu**, Matthieu Schapira, Ruprecht Kuner, Holger Sültmann, Maurizio Provenzano, Marie-Laure Yaspo, Benedikt Brors, Jan Korbel, Thorsten Schlomm, Guido Sauter, Roland Eils, Christoph Plass, Raffaella Santoro, BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence. Nature Genetics 2015.

45. Chenfeng He, Ying-Xin Li, Guangxin Zhang, **Zuguang Gu**, Rong Yang, Jie Li, Zhi John Lu, Zhi-Hua Zhou, Chenyu Zhang, Jin Wang, MiRmat: mature microRNA sequence prediction. PloS One 2012.

46. Lingyun Zhu, Qiang Wang, Lin Zhang, Zhixiang Fang, Fang Zhao, Zhiyuan Lv, **Zuguang Gu**, Junfeng Zhang, Jin Wang, Ke Zen, Yang Xiang, Dongjin Wang, Chen-Yu Zhang, Hypoxia induces PGC-1α expression and mitochondrial biogenesis in the myocardium of TOF patients. Cell Research 2010.

The 5 most important publications overall

1. **Zuguang Gu**, Roland Eils, Matthias Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016. [*5906 Google Scholar*

citations, by 06.02.2024, most cited paper on *Bioinformatics* 2016]
2. **Zuguang Gu**, Lei Gu, Roland Eils, Matthias Schlesner, Benedikt Brors, circlize Implements and enhances circular visualization in R. Bioinformatics 2014. [*2735 Google Scholar citations, by 06.02.2024*]
3. **Zuguang Gu#**, Matthias Schlesner, Daniel Hübschmann, cola: an R/Bioconductor package for consensus partitioning through a general framework. Nucleic Acids Research 2021.
4. **Zuguang Gu#**, Daniel Hübschmann, simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results. Genomics, proteomics & bioinformatics 2022.
5. **Zuguang Gu#**, On the dependency heaviness of CRAN/Bioconductor ecosystem. Journal of Systems and Software 2023. [*A software-engineering study*]

Books

The three books are freely online for everyone. They have not been published by any publisher, but they are in plan.

1. **Zuguang Gu**. Circular Visualization in R. https://jokergoo.github.io/circlize_book/book/
2. **Zuguang Gu**. ComplexHeatmap Complete Reference.
   https://jokergoo.github.io/ComplexHeatmap-reference/book/
3. **Zuguang Gu**. Gene Set Enrichment Analysis with R and Bioconductor (ongoing project).
   https://github.com/jokergoo/GSEA_book

Conference talks

useR! Is the official and the largest conference for R programming language. Bioconductor conference as well the European and Asian branches are the official conferences in the Bioconductor community.

1. circlize: circular visualization in R, poster presentation on **useR! 2015**, Aaborg, Denmark.
2. Make interactive complex heatmaps, poster presentation on **Bioconductor conference 2021**, online conference. Awarded as Best Innovative poster ideas.
3. spiralize: an R package for visualizing data on spirals, contributed talk on **BioC Asia 2021**, online conference.
4. On the heaviness of package dependencies, contributed talk on **useR! 2022**, online conference.
5. cola: a general framework for consensus partitioning, package demo on **Bioconductor conference 2022**, online conference.
6. simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results, package demo on **EuroBioc 2022**, Heidelberg Germany.
7. Complex Heatmap visualization, invited talk on **China Gut conference 2023**, online conference.
8. rGREAT: an R/Bioconductor Package for Functional Enrichment on Genomic Regions, package demo on **Bioconductor conference 2023**, online conference.
9. On the Dependency Heaviness of CRAN/Bioconductor Ecosystem, short talk on **Bioconductor conference 2023**, online conference.

<u>Teaching</u>

1. Teaching assistant of the Perl programming course (for undergraduate students) in 2014.
2. Prepare practice materials for the RNASeq course (for master students) in 2017.
   https://github.com/jokergoo/teaching_2017

3. Lectures on the topic of clustering (for master students) in 2018. https://github.com/jokergoo/teaching_2018
4. Supervised two intern students (2015, 2017) and two master students (2018, 2023).
5. Intensive course on the topic of gene set enrichment analysis in May 2022, February 2023, and January 2024. https://www.physalia-courses.org/courses-workshops/gse-in-r/

I am a certified Carpentries instructor for the teaching/training in the R/Bioconductor community. I am part of the Bioconductor teaching group and I am involved in the development of the official Bioconductor teaching material (https://carpentries-incubator.github.io/bioc-rnaseq/06-gene-set-analysis.html). I also plan to develop another free course on data visualization for the global bioinformatics community.

List of major scientific projects

My scientific projects can be categorized into two major categories, software-based and research-based. The software-based projects also include new method development for biological data analysis. The complete list of software that I developed can be found in the "**List of software**" section.

Software-based projects

1. **ComplexHeatmap**: The ComplexHeatmap package implements a new concept "complex heatmap visualization" which can efficiently visualize multi-dimensional omics data by a list of heatmaps and complex annotation graphics, making it very easy to associate information and find correspondence between different data types (**Gu Z**, 2022 iMeta, **Gu Z**, 2016 Bioinformatics). ComplexHeatmap has become a standard tool for visualizing high-dimensional omic data. The total downloads of ComplexHeatmap exceed 1M. The ComplexHeatmap paper has been cited over 5900 times. ComplexHeatmap has more than 1200 GitHub stars and around 110 other CRAN/Bioconductor packages have direct dependency on it.
2. **Circlize**: Circular visualization is a popular way for visualizing multi-dimensional genomic data, by arranging multiple tracks for associating various genomic signals and relations between elements. The circlize package provides a general-purpose solution for circular visualization, not only widely used in bioinformatics but in the general data science (Gu Z, 2014 Bioinformatics), with total downloads of 1.8M and total paper citations over 2700. It has around 900 GitHub stars and 119 other CRAN/Bioconductor packages directly depend on it.
3. **cola**. Consensus clustering is one of the standard methods for high-throughput data analysis, and it is widely used especially in finding new subtypes in cancer research. I developed the *cola* package which provides a standard framework for consensus clustering, in which various methods can be freely integrated (**Gu Z**, 2021 Nucleic Acids Research). *cola* provides rich visualization and web reporting interface, which allows to generate detailed consensus clustering reports. In addition, in an extended research based on *cola*, I proposed a new method called hierarchical consensus partitioning (HCP) (**Gu Z**, 2022 Briefings in Bioinformatics), to solve the problem that the stability of consensus clustering decreases when the number of clusters becomes larger. HCP will play an important role in dealing with highly heterogeneous data.
4. **simplifyEnrichment**. Gene Set Enrichment Analysis (GSEA) is one of the standard analysis methods in bioinformatics, which is used to evaluate the significant biological functions. However, a common difficulty is that in most cases, GSEA will generate a very long enrichment list (often hundreds), which makes it difficult to summarize. I developed the *simplifyEnrichment* package, which provides a new method called "binary cut" to cluster the enrichment results

based on the semantic similarities of functional terms, and to summarize the enrichment results through an effective word cloud visualization (**Gu Z**, 2022 Genomics, Proteomics & Bioinformatics).

5. **rGREAT**. This is a tool for gene set enrichment analysis (**Gu Z**, 2023 Bioinformatics) directly applied on genomic intervals. In the tool, I proposed a new background model which allows to perform enrichment analysis within a specific genomic context (such as high CG density genomic regions). *rGREAT* provides a very scalable way to easily integrate any gene set database and species.

6. **spiralize**. The R package *spiralize* provides a general solution for visualizing data on Archimedean spirals (**Gu Z**. 2022 Bioinformatics). Spiral visualization has two advantages: it is able to visualize data with long axes, which greatly improves the resolution of visualization; second, it is efficient for time series data to reveal periodic patterns.

7. **InteractiveComplexHeatmap**. The *ComplexHeatmap* package (**Gu Z**. 2016 Bioinformatics) also developed by me has become a stand tool for making static heatmaps in bioinformatics. Here the *InteractiveComplexHeatmap* package brings interactivity to *ComplexHeatmap* (**Gu Z**. 2022 Bioinformatics). *InteractiveComplexHeatmap* is designed with an easy-to-use interface where static complex heatmaps can be directly exported to an interactive web-based application only with one additional line of code. *InteractiveComplexHeatmap* also provides flexible functionalities for integrating interactive heatmap widgets to build more complex and customized web applications.

8. **simona**. Ontology defines semantic relations of biological entities in a hierarchical structure. With the *simona* package, I aim to define a robust data structure for ontologies, to implement an efficient way to compute semantic similarities, to provide an easy-to-use interface for general ontology data and to implement visualizations for large scale ontologies.

9. **epik**. This is also an on-going project. It is a collection of functions for integrative analysis on epigenomics datasets, e.g. methylation, ChIPseq and gene expression.

10. **Omics data browser**. I am involved in the precision oncology program in DKFZ-NCT, which has a complex set of data of thousands of samples in various data types. I developed a web-based data browser (https://github.com/CO-DKFZ/dataMASTER_browser) which can query and perform analysis on this dataset on various levels (i.e., gene-level, patient-level, or cohort-level). The tool will be deployed for other bioinformaticians and clinicians to use.

Software engineering project

11. Based on more than 10 years of experience of working with R, I conducted a software engineering study on the R language, aiming to find new patterns of the dependency relations in the R package ecosystem. Different from the traditional software engineering researches, I proposed a new metric "dependency heaviness", which measures which of the parent dependencies introduces more additional dependencies to a child package. Based on this, I developed a new tool *pkgndep*, to analyze any R packages to find the heavy dependencies and to provide directions for future optimization (**Gu Z**, 2022 Bioinformatics). I also conducted a systematic analysis on the entire R package ecosystem, revisiting the R language ecosystem from a new perspective (**Gu Z**, 2023 Journal of Systems and Software). I studied short-range dependencies and also long-range indirect dependencies, explaining the core graphs in the ecosystem. This study provides new directions for understanding how software dependencies are propagated throughout the ecosystem and revealing the most vulnerable parts of the ecosystem.

Research-based projects

As the "core bioinformatician", I have participated in many collaborative projects in the German Cancer Research Center, Heidelberg University Hospital, Freiburg University Hospital and other research institutions. The following 4 research projects are published in the past 5 years where I am one of the main contributors:

1. **Glioblastoma Project** ([Wu Y*, Fletcher M*, **Gu Z***, et al., 2020 Nature Communications](#)). This project has a huge amount of data, including whole genome sequencing, exome sequencing, whole genome methylation sequencing, RNAseq and ChIPseq. The project studies the four subtypes of Glioblastoma with the aim of identifying the epigenetic characteristics of the four subtypes and the mechanisms that lead to switching between the four subtypes. In this project, I conducted a large number of integrated analyses to reveal in detail how the changes in gene expression, DNA methylation, and chromatin activity are associated in the four subtypes of Glioblastoma.
2. **Pancreatic Cancer Project** ([Espinet E#, **Gu Z**, et al., 2021 Cancer Discovery](#)). This project also includes a large amount of data, including exome sequencing, RNAseq, and genome-wide methylation sequencing. In addition to the integrative analysis similar as project 1, we also discovered two new tumor subtypes using my *cola* package. We found that in both subtypes, changes in the methylation of repeat sequences in the genome affect mRNA transcription, which in turn affect the downstream INF pathway. There was a statistically significant difference in survival between the two subtypes.
3. **Primary central nervous system lymphoma** ([Radke J#, Ishaque N#, Koll R, **Gu Z**, et al., 2022 Nature Communications](#)). This study systematically investigated subtypes of primary central nervous system lymphoma from the perspective of genome and transcriptome. My work focused on the analysis of the transcriptome. I applied the *cola* package, performed subtype analysis, and successfully found the marker genes that distinguish different tumor subtypes.
4. **Research on HCV-specific T cells** ([Hensel N*, **Gu Z***, Sagar*, et al., 2021 Nature Immunology](#)). My main contribution is the mining and analysis of transcriptomics data, revealing the characteristics of gene expression of two subtypes of HCV-specific CD8+ T cells: memory and terminal exhaustion.

List of software

I am active in developing new methods and tools for biological data analysis. I have *independently* developed >20 tools. The total downloads are > 5.8 million. I am also an active and responsible maintainer of all my tools. I have answered > 1500 questions from users on GitHub and by email. My software can be put in the five categories. The full list of my tools can be found from https://jokergoo.github.io/software/ or https://github.com/jokergoo.

Data visualization

1. [**circlize**](#): Circular visualization in R. It has been a major software for making circular visualizations. It has been widely applied not only in bioinformatics, but also in varieties of other research topics such as climate changes, migration flows, information science, environment science and more. It has been downloaded for more than 1.8M times.
2. [**ComplexHeatmap**](#): Flexibly arrange multiple heatmaps and self-define annotation graphics. It is a standard tool for making heatmaps for bioinformatics studies. It has been introduced in textbooks also in YouTube videos. It has been downloaded for 1M times.
3. [**spiralize**](#): Visualize data on spirals.
4. [**InteractiveComplexHeatmap**](#): Make interactive complex heatmaps.

5. **EnrichedHeatmap**: Visualize the enrichment of genomic signals on specific target regions.
6. **HilbertCurve**: Making 2D Hilbert Curve.
7. **gtrellis**: Genome-level trellis layout.
8. **BioCartaImage**: BioCarta pathway images

Statistical analysis

9. **simplifyEnrichment**: Simplify functional enrichment results.
10. **cola**: A Framework for Consensus Partitioning.
11. **rGREAT**: Functional Enrichment on Genomic Regions.
12. **simona**: Semantic Similarity on Bio-ontologies.
13. **CePa**: Centrality-based pathway enrichment.

Biological datasets

14. **GeneSummary**: RefSeq Gene Summaries.
15. **UniProtKeywords**: Keywords from the UniProt database.
16. **BioMartGOGeneSets**: Gene Ontology Gene Sets from BioMart.

Software engineering

17. **pkgndep**: Analyzing Dependency Heaviness of R Packages.

Utility tools

18. **bsub**: Submitter and monitor of LSF cluster.
19. **GlobalOptions**: Setting global options in R.
20. **GetoptLong**: Wrapper of the Perl module Getopt::Long, also provides a simple variable interpolation in R.

List of awards

1. World top 2% most cited scientists in the year 2021.
2. World top 2% most cited scientists in the year 2022.
3. The best method 2022 in the journal *iMeta*.
4. Most cited author on *Bioinformatics* 2016.
5. Top 10 authors who published the most first-author papers on *Bioinformatics* of all time.

Professional Service

Manuscripts reviewed for Bioinformatics (12), PLoS Computational Biology (1), BMC Bioinformatics (5), BMC Genomics (2), Frontiers in Genetics (1), Journal of Computational and Graphical Statistics (1), F1000Research (1), iMeta (2), GigaScience (1).