People of Data

A hybrid approach for identifying drug repurposing candidates and their mechanisms: An interview with Vanessa Lage-Rupprecht and two co-authors

Vanessa Lage-Rupprecht, Bruce Schultz, and Marcin Namysl

Senior researcher Vanessa Lage-Rupprecht and two collaborators talk about what data science means to them and illustrate how they managed to create a data and lab coexistence in their drug-repurposing project, which was recently published in Patterns. In this article, they have developed a drug-target-mechanism-oriented data model, Human Brain PHARMACOME, and have presented it as a resource to the community.
Human Brain PHARMACOME (HBP). Mutual analysis led to the identification of interesting repurposing candidates in the Alzheimer disease context.

Vanessa and Bruce, what's next for the project?
VLR: Our interplay of in-silico and in-vitro analysis has yielded far-reaching findings that are being further explored experimentally on the one hand and taken up in new studies on the other. The generalizability of our approach allows us to investigate a wide range of biological questions and to continue the successful symbiosis of computational and wet lab work.

BS: Though our approach published here will provide a lot of utility in generating new insights and hypotheses, it is only the beginning. We wish to improve our integration of quantitative data within our knowledge graph and develop new methods capable not only of pinpointing new modes of action for drugs, but also predicting dosages, likely side effects, and even unknown drug interactions.

Vanessa, how important was the collaboration to the success of the paper? How important do you think collaboration is to research in general?
VLR: The collaboration was enormously important for the success of the paper. The mutual exchange between data and laboratory scientists helped a lot to identify the needs of both sides and to achieve optimal results. We believe that cross-domain collaboration is a groundbreaking concept to overcome limitations, e.g., in therapeutic research.

Vanessa, what advice would you give other scientists/data scientists looking to establish collaborations?
VLR: It is important to have a common goal in mind and to continuously consult with each other. Communication and mutual feedback are the most important and limiting factors. As a data scientist, take the time to explain what modern data science is already capable of so the collaborators do not limit their ideas from the start. Vice versa, try to get the best possible understanding what types of data are relevant and what the key areas of possible new insights are that could drive experimental progress.

Marcin, how do you stay updated with both advances in data-science techniques and the advances in the field/domain that you work in?
MN: I keep updated with advances in research areas of interest by reading conference proceedings and journal issues as well as the latest preprints published online. Moreover, I enjoy going through the data and the code that were made publicly available by the authors. Last but not least, discussion of the latest research advances with peers gives me further insights into the up-to-date solutions to the most pressing problems in my fields of interest.

Marcin, which of the current trends in data science seem the most interesting to you? In your opinion, what are the most pressing questions for the data science community?
MN: From the perspective of my research interests, the development in the adversarial robustness field excites me the most. Moreover, there are many other interesting trends in data science such as fostering interpretability and reproducibility of machine learning research or promoting sustainable and efficient solutions that reduce the carbon footprint and financial cost of running computationally expensive experiments. In my opinion, advocating responsible and ethical AI research is one of the most pressing questions for the data science community.

Bruce, what is the fun part of being a data scientist?
BS: The rush a biological scientist gets when their experiment that they have been troubleshooting for months finally produces meaningful results is the same elation that a data scientist feels when their code runs to completion and returns what was expected. Being able to take the same collection of data and manipulate it in a creative, logical way to identify new, useful patterns is invigorating.

Bruce and Marcin, have you ever used your data science skills in your personal life? If yes, how?
MN: Actually, I use them quite often, especially to manage my finances, i.e., to analyze incomes and expenses, or to track my fitness activities.

BS: Data science skills have a wide variety of day-to-day applications such as being able to identify troublesome applications and bottlenecks in my home network as well as being able to automate repetitive tasks, especially those involving Excel spreadsheets. Currently, I am working on developing a machine learning model to find recipes I would enjoy online.

Bruce, what attributes make a data scientist successful?
BS: Data science requires two major skills: a strong foundation in an applicable programming language (Python, R, Julia, Java, etc.) as well as a decent understanding of both mathematics and statistics. Depending on the work environment, there are several other factors that can help one to succeed, but being able to find the faults and issues within a dataset, fix them, and create meaningful algorithms to apply to it are highly dependent on the skills mentioned above.

REFERENCES
1. Cummings, J. (2018). Lessons Learned from Alzheimer Disease: Clinical Trials with Negative Outcomes. Clin. Transl. Sci., 11, 147–152.
2. Yiannopoulou, K.G., Anastasiou, A.I., Zachariou, V., and Pelidou, S.H. (2019). Reasons for failed trials of disease-modifying treatments for Alzheimer disease and their contribution in recent research. Biomedicines 7, 97.

Figure 1. Left: Vanessa Lage-Rupprecht, middle: Bruce Schultz, right: Marcin Namysl
3. Lage-Rupprecht, V., Schultz, B., Dick, J., Namysl, M., Zaliani, A., Gebel, S., Pless, O., Reinshagen, J., Ellinger, B., Ebeling, C., et al. (2022). A hybrid approach unveils drug repurposing candidates targeting an Alzheimer pathophysiology mechanism. Patterns 3. https://doi.org/10.1016/j.patter.2022.100466.

About the authors
Vanessa Lage-Rupprecht is a research scientist at Fraunhofer SCAI Bioinformatics. After receiving her diploma in biology at the Rheinische Friedrich-Wilhelm-University in Bonn (Germany), she received her PhD in experimental neurophysiology at the University Hospital Bonn and continued her research in cellular neuroscience at McGovern Medical School (Houston, TX, USA) and the University of Regensburg (Germany). At Fraunhofer SCAI Bioinformatics she works in the field of applied semantics and operates as a bridge builder between biomedical domains and informatics implementation strategies with a focus on the interoperability of data and knowledge.

Bruce Schultz is a research scientist at Fraunhofer SCAI Bioinformatics. In 2012, he obtained a bachelor of science in biochemistry from the University of Michigan and thereafter worked with Dr. Hendrik Streeck in HIV vaccine research at the Henry Jackson Foundation. His work took him to Essen, Germany, where he began focusing on the data-science aspect of his research, thus leading him to get his master of science in bioinformatics from the University of Bonn in 2019. Bruce now works at Fraunhofer SCAI, developing tools and applications to help bridge the gap between biological researchers and data scientists.

Marcin Namysl is a research scientist for Fraunhofer IAIS and a PhD candidate in computer science at the University of Bonn. He completed his master’s degree in control engineering and robotics at Wroclaw University of Science and Technology in 2008. After his graduation, he gained nine years of industry experience by working as a software engineer and computer vision expert. Now he is doing research that bridges pattern recognition and natural language processing and is focusing on improving the robustness of information extraction systems for entity and table recognition.