

Bioinformatics @ JKU LINZ

Günter Klambauer

Institute of Bioinformatics
Johannes Kepler University, Linz, Austria

Institute of Bioinformatics

- Bioinformatics with strong background in machine learning
 - Deep Learning, neural networks, Support Vector Machines, probabilistic models
- Bachelor & Master & PhD studies
 - Bachelor in cooperation with University Budweis
- Cooperations with pharmaceutical companies
 - J&J, Merck, Sanofi, Bayer
- Deep Learning cooperation with AUDI

Hardware

- “Raptor”: 20 CPUs, 256 GB RAM, 8 GeForce GTX Titan X GPUs
- “Cobra”: 20 CPUs, 384 GB RAM, 2 Tesla K80 GPUs
- “K40”: 16 CPUs, 128 GB RAM, 4 Tesla K40 GPUs
- “Tiger”: 32 CPUs, 512 GB RAM

- Computing facilities at JKU
 - SGI Altix UltraViolet 1000, 2048 Intel Xeon cores, 16 TB RAM
 - SGI Altix UltraViolet 1000, 512 Intel Xeon cores, 1 TB RAM

Personnel & infrastructure

- 2 professors
- 2 postdoc researchers
- 5 researchers
- Technician



Sepp Hochreiter



Ulrich Bodenhofer



Gundula Povysil



Andreas Mayr



Thomas Unterthiner



Martin Heusel



Karin Schwarzbauer



Bernhard Nessler



Günter Klambauer

Scientific use cases

- Probabilistic models for biotechnologies
 - Hochreiter, S., Clevert, D. A., & Obermayer, K. (2006). A new summarization method for Affymetrix probe level data. *Bioinformatics*, 22(8), 943-949.
 - Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D. A., Mitterecker, A., Bodenhofer, U., & Hochreiter, S. (2012). cn. MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic acids research*, 40 (9), e69.
 - Clevert, D. A., Mitterecker, A., Mayr, A., Klambauer, G., Tuefferd, M., De Bondt, A., ... & Hochreiter, S. (2011). cn. FARMS: a latent variable model to detect copy number variations in microarray data with a low false discovery rate. *Nucleic acids research*, 39(12), e79-e79.
 - Klambauer, G., Unterthiner, T., & Hochreiter, S. (2013). DEXUS: identifying differential expression in RNA-Seq studies with unknown conditions. *Nucleic acids research*, 41(21), e198-e198.
- Deep Learning applications
 - Mayr, A., Klambauer, G., Unterthiner, T., & Hochreiter, S. (2015). DeepTox: Toxicity Prediction using Deep Learning. *Frontiers in Environmental Science*, 3, 80.
 - Unterthiner, T., Mayr, A., Klambauer, G., Steijaert, M., Wegner, J. K., Ceulemans, H., & Hochreiter, S. (2014). Deep learning as an opportunity in virtual screening. In *Proceedings of the Deep Learning Workshop at NIPS*.

Scientific use cases

- New machine learning methods
 - Clevert, D. A., Mayr, A., Unterthiner, T., & Hochreiter, S. (2015). Rectified factor networks. In *Advances in Neural Information Processing Systems* (pp. 1846-1854).
 - Clevert, Djork-Arné, Unterthiner, T., and Hochreiter, S. (2015). "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)." *To appear*.
- Data analysis and development of predictive models in large-scale projects
 - Federica Eduati, Lara M Mangravite, Tao Wang, Hao Tang, ... , Sepp Hochreiter, Günter Klambauer, Andreas Mayr, ... , Ivan Rusyn, Fred A Wright, Gustavo Stolovitzky, Yang Xie, and Julio Saez-Rodriguez (2015). Prediction of human population responses to toxic compounds by a collaborative competition. *Nature Biotechnology*, 33, 933–940.
 - Günter Klambauer, Bie Verbist, Liesbet Vervoort, Willem Talloen, QSTAR Consortium, Ziv Shkedy, Olivier Thas, Andreas Bender, Hinrich W.H. Göhlmann, Sepp Hochreiter (2015). Using transcriptomics to guide lead optimization in drug discovery projects. *Drug Discovery Today*, 20(5).
 - SEQC/MAQC-III Consortium. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology*, 32(9), 903-914.

Thoughts and comments

- Data remains at a fixed location
- Data contributors give access rights to Bioinf groups
- Algorithms are applied without duplicating data

