Seminar:
Statistical Methods for High-Dimensional Biomedical Data

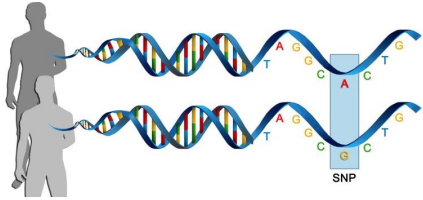
Christian Staerk

Department of Statistics
TU Dortmund University

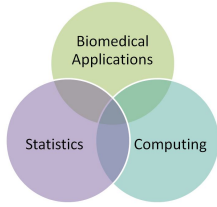
Winter Semester 2024 / 2025

Statistical Methods for High-Dimensional Biomedical Data

Krutmann et al. (2017)



https://atlasofscience.org/wp-content/uploads/2016/03/Fig1-Iacobucci.jpg



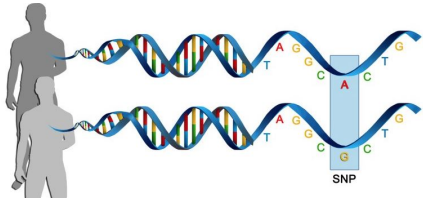
High-dimensional Biomedical Data

- Exposome: Environmental factors like pollution, radiation, nutrition, ...
- Genome-wide association study (GWAS) with $p \approx 10^7$ SNPs
- Gene expression data with $p \approx 20{,}000$ genes

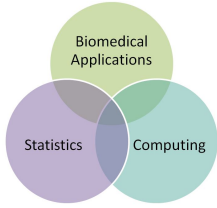
$\rightarrow$ **High-dimensional Statistics**: Many variables p, often $p > n$ for sample size n

Statistical Methods for High-Dimensional Biomedical Data

Krutmann et al. (2017)



https://atlasofscience.org/wp-content/uploads/2016/03/Fig1-Iacobucci.jpg



Seminar Topics

- Classical variable selection and prediction methods, e.g., Lasso and variants
- Ensemble methods like Random Forest, Boosting, ...
- Multiple testing
- Graphical models
- ....

Seminar Organization

Target Group

- Seminar for Statistics & Data Science (Bachelor & Master)
- No formal prerequisites
- Basic knowledge in statistical modelling and R is expected

Parts of the Seminar

- Oral presentation ($\sim$ 30 min B.Sc. / 45 min M.Sc.) and discussion ($\sim$ 10 min)
- Active participation in student discussions
- Written report (15-20 pages)

Preliminary Timeline

- First meeting (topic selection) in October 2024 (TBA)
- Block seminar in February 2025 (TBA)
- Deadline for reports in March 2025 (TBA)