Sustainable Data Management for Single-Cell Sequencing: Tools, Platforms, and Challenges

Nils Rosenboom[a], Sabine A. Smolorz[a], Robert Kossen[a], Ulrich Sax[a, b], Sara Y. Nussbeck[a, c], Harald Kusch[a, b, d]

[a]Department of Medical Informatics, University Medical Center Göttingen (UMG), Germany; [b]Campus-Institute Data Science (CIDAS), Göttingen, Germany; [c]Central Biobank UMG, University Medical Center Göttingen, Germany; [d]Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Germany.

presented at GMDS 2024 by

LinkedIn — Nils Rosenboom — ORCiD — LinkedIn — Harald Kusch — ORCiD

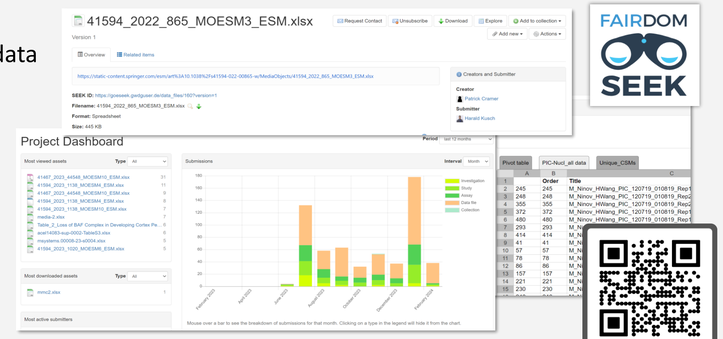
Motivation

Managing and integrating single-cell sequencing (SCS) data sustainably poses significant challenges for today's institutions and researchers. This exploration focuses on the research data management (RDM) tools and platforms available to handle and analyze such data in a "Findable, Accessible, Interoperable, and Reusable" (FAIR) manner. The challenges encompass the formidable size of the data, storage and computational complexity, and the need for comprehensive metadata, especially in respect the generally complex preprocessing of the SCS-data. Different concepts, such as the usage of different cloud concepts to match these requirements, are analyzed and discussed as solutions are required not only for individual institutions, and scalable systems are necessary to handle SCS data effectively.

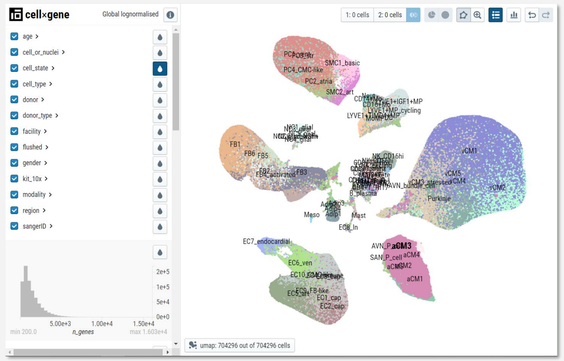
Data Management and Visualization Tools

FAIRDOM SEEK[(1)]

- Structured approach to manage metadata
- Organize data and processing steps
- Focus on FAIRification
- Not specialized
- Searchable files
- Open source
- Does not support large files
- Does not integrate analytic functions
- Requires addional tools



seek4science.org



CELLxGENE

- Visualize single cell data through dimensional reduction
- Export pseudo bulks for downstream analysis
- Needs external tools for differential analysis
- Funded by Chan-Zuckerberg-Initiative

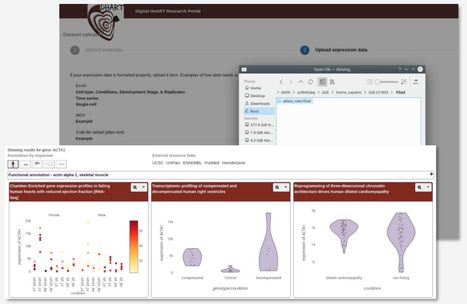
Heart Cell Atlas[(2)]

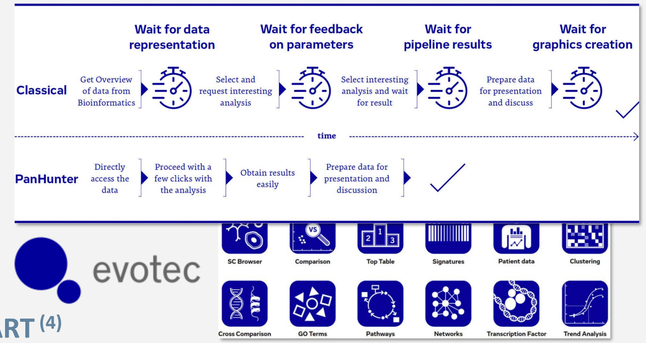
- Public dataset in CELLxGENE instance

Integrated Platforms

PanHunter by Evotec[(3)]

- Software as a Service (SaaS)
- All-in-One industry solution
- Integrates CELLxGENE
- Streamlined workflow through pre-built apps
- FAIRness will be evaluated





evotec

DHART [(4)]

- Focus on cardiac single cell data analysis
- Based on gEAR framework
- Still limited functionality
- Depends of project specific funding

DZHK
Deutsches Zentrum für Herz-Kreislauf-Forschung

Conclusion

Focus and Functionalities

- Not all tools are specialized
- All-in-One and linked system approaches
- Software as a Service (SaaS) or local hosting
- App based approaches

Results

- No simple, „catch-all" solution available
- Promising tools still in development

Challenges

- Applicability on certain aspects of transcriptomics
  - Spatial transciptomics, single cell sequencing
- Usability (too complicated, limited functionality)
- Communication with reseachers
- Integration of FAIRification
- Hosting
- License binding

Acknowledgement We thank Luca Freckmann, Linus Weber, Fabian Rakebrandt and Sven Bingert for excellent technical and consulting support.

Conflict of Interest N. Rosenboom will be employed at Evotec as a working student for the duration of his following master thesis project

References **(1)** Wolstencroft K, Owen S, Krebs O, Nguyen Q, Stanford NJ, Golebiewski M, et al. SEEK: A Systems Biology Data and Model Management Platform. BMC Systems Biol 2015;9(1):33. https://doi.org/10.1186/s12918-015-0174-y **(2)** Heart Cell Atlas. Human heart single cell atlas. [Internet]. Available from: https://www.heartcellatlas.org/ **(3)** PanHunter by Evotec. [Internet]. Available from: https://www.evotec.com/panomics/panhunter **(4)** Orvis J, Gottfried B, Kancherla J, Adkins RS, Song Y, Dror AA, et al. gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. Nat Methods. 2021;18:843–844. https://doi.org/10.1038/s41592-021-01200-9