Jörn Walter and Hannah Schickl (Editors)

Single-Cell Analysis in Research and Medicine : Report oft he Interdisciplinary Research Group *Gene Technology Report*

Berlin-Brandenburg Academy of Sciences and Humanities, 2019
ISBN: 978-3-939818-85-4

Persistent Identifier: urn:nbn:de:kobv:b4-opus4-32879

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Namensnennung 4.0 International Lizenz zur Verfügung gestellt.

Jörn Walter and
Hannah Schickl
(Editors)

SINGLE-CELL ANALYSIS IN RESEARCH AND MEDICINE

Report of the Interdisciplinary Research Group
Gene Technology Report

BERLIN-BRANDENBURG ACADEMY OF SCIENCES AND HUMANITIES

berlin-brandenburgische
AKADEMIE DER WISSENSCHAFTEN

SINGLE-CELL ANALYSIS IN RESEARCH AND MEDICINE

Report of the Interdisciplinary Research Group *Gene Technology Report*

Jörn Walter and Hannah Schickl
(Editors)

Editors: Jörn Walter and Hannah Schickl for the Interdisciplinary Research Group
Gene Technology Report, Berlin-Brandenburg Academy of Sciences and Humanities

Translation: Übersetzungsagentur24
Layout: angenehme gestaltung/Thorsten Probst
Coverphoto: Agnieszka Rybak-Wolf
Print: PIEREG Druckcenter Berlin GmbH

© Berlin-Brandenburg Academy of Sciences and Humanities, 2019
Jägerstraße 22/23, 10117 Berlin, www.bbaw.de, info@gentechnologiebericht.de

This publication was funded by the Friede Springer Stiftung and supported by the network *Single Cell Omics Germany* (SCOG) which is funded by the Federal Ministry of Education and Research (BMBF).

friede springer stiftung

No part of this booklet may be reproduced without express permission of the publisher.

ISBN: 978-3-939818-85-4

TABLE OF CONTENTS

Foreword ... 6

1. Introduction ... 7
(Jörn Walter, Nina Gasparoni)
1.1 From complex tissues to single-cell signatures – new horizons for modern cell biology ... 7
1.2 Technological developments and the birth of modern single-cell biology ... 8
1.3 Efforts in the field of single-cell omics ... 9
1.4 Upcoming developments in single-cell omics ... 10
1.5 Single-cell analysis and developmental biology ... 11
1.6 Single-cell analysis in biomedical research ... 12
1.7 Challenges and limitations of single-cell technologies ... 12
1.8 Final remark ... 13
1.9 References ... 14

2. Single-Cell Genomics is Transforming Developmental Biology
(J. Philipp Junker, Christian Popp, Nikolaus Rajewsky) ... 16
2.1 Introduction ... 16
2.2 Perturbation analysis in developmental model systems ... 16
2.3 Spatial information ... 18
2.4 Temporal information – pseudo-temporal ordering ... 19
2.5 Temporal information – high-throughput lineage analysis ... 19
2.6 Measuring other parameters beyond RNA ... 20
2.7 LifeTime, a new initiative built on single-cell analysis ... 22
2.8 References ... 23

3. The Past, Present and Future of Single-Cell Omics in Biomedicine
(Anna C. Aschenbrenner, Elvira Mass, Joachim L. Schultze) ... 26
3.1 Single-cell omics forges new paths ahead by changing our perception and treatments of major diseases ... 26
3.2 The past of single-cell omics ... 26
3.3 The present of single-cell omics ... 28
3.4 The future of single-cell omics ... 29
3.5 References ... 31

4. Data Analytics in Single-Cell Genomics Using Machine Learning
(Hananeh Aliee, Anna Sacher, Fabian J. Theis) ... 33
4.1 Introduction ... 33
4.2 Machine learning in single-cell transcriptomics ... 34
4.3 Outlook ... 36
4.4 References ... 37

5. Single-Cell Transcriptome Analysis in Plants
(Bernd Müller-Röber) 38
5.1 Plant single-cell transcriptomes 38
5.2 Transcriptome analyses using isolated cell nuclei 39
5.3 Future research 40
5.4 Summary 40
5.5 References 41

6. Single-Cell Analyses and Ethical Considerations
(Heiner Fangerau, Lilian Marx-Stölting, Angela Osterheider) 42
6.1 Introduction 42
6.2 Ethical topics 43
6.3 Conclusion 45
6.4 References 45

7. Problem Areas and Indicators on the Topic of Single-Cell Analysis
(Angela Osterheider, Yaroslav Koshelev, Marlen Reinschke, Lilian Marx-Stölting) 46
7.1 Introduction: motivation and purpose 46
7.2 Problem areas 46
7.3 Indicators 49
7.4 Conclusion 53
7.5 References 54

8. Core Statements and Recommendations for Action on Single-Cell Analysis
(Interdisciplinary Research Group *Gene Technology Report*) 55
8.1 Core statements on single-cell analysis 55
8.2 Recommendations for action for the handling of single-cell technology and single-cell data 59

9. Appendix
9.1 Authors 60
9.2 Members of the Interdisciplinary Research Group *Gene Technology Report* 63
9.3 Publications of the Interdisciplinary Research Group *Gene Technology Report* 64

FOREWORD

The Interdisciplinary Research Group (IAG) *Gene Technology Report* at the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) has been observing and monitoring new developments in gene technologies and their relevance for science and society since 2001. Its tasks at the interface between science, politics, business and the public include addressing current topics as promptly as possible and comprehensively examining them in order to initiate and promote an objective, fact-based public discussion.

Single-cell analyses comprise a multitude of analytical methods that share a common feature, namely the focus on individual cells. This is in contrast to previous methods that provided summarized data for cell clusters, groups of cells, tissues and organs. The new field offers huge potential not only for basic research, but also for medical and biotechnological applications, as it opens up new levels in the context-related and personal interpretation of biological interconnections. This brochure on single-cell analysis provides an overview on the new possibilities from the viewpoint of developmental biology, biomedicine and bioinformatics, but also addresses possible social implications and consequences.

Author-attributed articles do not necessarily reflect the editors' or the group's opinion. However, the group shares responsibility for the chapter "Core Statements and Recommendations for Action on Single-Cell Analysis". The recommended actions presented have been agreed on by the members of the IAG, but might not represent views of all members of the academy; however, the BBAW unreservedly stands behind the quality of the work carried out.

Heartfelt thanks to the Friede Springer Foundation for promoting the work of the IAG at the BBAW. Thanks also to the authors of the articles as well as the editorial team and the office of the IAG.

This brochure has been compiled on the initiative of the IAG *Gene Technology Report* at the BBAW. We are delighted that our work has been supported by the network *Single Cell Omics Germany* (SCOG).

Boris Fehse
Spokesperson of the Interdisciplinary Research Group *Gene Technology Report*
at the Berlin-Brandenburg Academy of Sciences and Humanities.

Hamburg, August 2019

Jörn Walter, Nina Gasparoni

1. INTRODUCTION

1.1 FROM COMPLEX TISSUES TO SINGLE-CELL SIGNATURES – NEW HORIZONS FOR MODERN CELL BIOLOGY

Since the birth of modern cell biology in the early 20th century, scientists have searched for technologies allowing them to capture the molecular mechanisms that regulate the biological programs of individual cells in a complex organism. The recent development of single-cell analysis provides science with the means to generate comprehensive and highly precise data about the molecular character and functioning of individual cells. The interpretation of these data opens up entirely new possibilities of understanding complex biological processes within cells, from complex developmental processes and aging to adaptation to environmental conditions, and from complex processes of organ development to the cause and consequence of diseases. With the help of new technologies, these processes can be captured precisely for thousands to millions of individual cells at once. New techniques enable researchers to use the obtained data for modelling the spatial allocation of an individual cell in the tissue as well as its developmental dynamic. Thus single-cell analysis brings biologists closer to their goal of precisely understanding and influencing the properties and functioning of individual cells in the organism.

Until recently, functional concepts of cell programs were based on a combination of genetic, biochemical, and molecular data generated from cell *populations*. All comprehensive analyses prior to single-cell omics[1] had to be performed on cells that were isolated in large quantities from tissues or body fluids as "homogeneous" cell populations. The molecular signatures (such as the gene expression patterns) gathered for such cell populations always reflect the sum of individual cells and hence face serious restrictions: they do not allow to capture individual functional variation, changes during development, differences in cell cycle states, in the individual age of a cell, or its response to its spatial localization. These individual properties cannot be determined adequately by the analysis of cell populations; with single-cell analysis it became possible to reveal the molecular differences between single cells. Moreover, the analysis of cell populations often depends on sorting/selection procedures, so that not all cells from the same tissues can be collected and analyzed simultaneously.

For many of these problems, far-reaching solutions are now emerging with the development of comprehensive single-cell omics technologies. Combined with new unbiased sorting techniques,

1 "Omics" is a neologism that describes several research areas in the field of life sciences that contain the suffix "omics", such as genomics, transcriptomics, metabolomics and proteomics. The suffix indicates, that the focus of the study lies on the whole cellular content of the molecules being studied (e.g., an entirety of the genes, transcripts of genes, metabolites or proteins in the cells).

extended microscopic techniques such as multi-RNA-FISH,[2] and novel bioinformatics approaches, single-cell analyses will provide answers to hitherto unsolvable questions and open up new systemic insights into the function of individual cells in a complex biological environment.

1.2 TECHNOLOGICAL DEVELOPMENTS AND THE BIRTH OF MODERN SINGLE-CELL BIOLOGY

The basis of modern single-cell omics was the functional annotation of the human genome and the genomes of all major model organisms. The localization of genes and other functional/regulatory parts of the genome boosted numerous studies in "functional genomics" in order to assign individual molecular programs to cell types. This functional genome revolution was made possible by a fast development of novel technological advances in massive parallel sequencing methods, known as next-generation sequencing (NGS) technologies. NGS methods, originally developed for genome sequencing, were rapidly adapted for functional analyses of cells, such as comprehensive gene expression profiling[3] using NGS-based RNA-sequencing (RNA-seq)[4] methods. About 10 years ago, the first manual attempts were made to obtain comprehensive mRNA[5]-seq signatures from a few sorted single mammalian cells (Tang et al., 2009). The first successful applications boosted this new field of research and very rapidly novel high throughput methods for single-cell isolation and NGS processing were developed to obtain comprehensive RNA-seq-based gene expression profiles for *many* single cells. All these novel technologies combine microprocessing with sophisticated molecular protocols for the preparation of complex sequencing "libraries".[6] The new technologies are expanding the possibilities in two important directions: i) an increasingly comprehensive capturing of molecular signatures in individual cells and ii) the possibility to analyze high numbers of individual cells in cheaper and massive parallelized sequencing systems.[7]

2 FISH (fluorescence in situ hybridization) is a method where specific molecules in a sample are labeled with fluorescent markers that can then be detected. The term "in situ" expresses that the molecules are being detected at the position where they naturally occur. Multi-mRNA-FISH can detect many mRNAs at the same time.

3 Gene expression profiling studies which genes are being expressed in cells.

4 RNA-seq is a method that uses NGS in order to detect the quantity and presence of RNA transcripts in a sample at the timepoint of the investigation (also called "whole transcriptome shotgun sequencing"). Since the method is used to analyze the cellular transcriptome it is a method of transcriptomics.

5 mRNA is the abbreviation for messenger RNA, the molecule that is the product of gene expression. In a process called transcription DNA is used as a template for the generation of RNA, which in turn gets processed to mRNA that leaves the cell nucleus and is being translated into an amino acid sequence, thus building up a protein. mRNA is therefore the transcript of the corresponding DNA and the study of RNA content in a cell is called transcriptomics. The transcriptome of a cell consists in all of its RNA, which allows conclusions on which genes are expressed in this particular cell (at a given time).

6 DNA copies are generated of the transcriptome (mRNA is re-transcribed into DNA, called cDNA for complementary DNA), i.e. a "library" of the individual mRNA molecules present in a cell is created. Such a library can then be read out using high-throughput sequencing, determining the presence and number of mRNA copies of a gene.

7 Current RNA-seq methods are either based on microfluidic systems such as the most popular methods of "Drop-seq", or they use microwell "cell-container-like" solutions. In both systems, the RNA-libraries of individual cells are labeled by sophisticated adaptor barcoding techniques to distinguish the expression profiles of individual cells. Some of the new technologies such as "MARS-seq" (Jaitin et al., 2014; Keren-Shaul et al., 2019), "Drop-seq" (Macosko et al., 2015), "Seq-Well" (Gierahn et al., 2017), "SPLiT-seq" (Rosenberg et al., 2018) – to name just a few of the current "leaders" – have reached a high throughput level that allows RNA-seq signatures to be obtained for millions of single cells at reasonable sequencing costs.

The field started by producing a series of deep single-cell maps for blood cells, providing new insights into different immune cell types in both healthy development and disease (Kowalczyk et al., 2015; Wilson et al. 2015). Soon after, the first comprehensive maps of model organisms were produced, followed by recent publications on tissues and organs in human and mouse (Han et al., 2018), such as liver (Halpern et al., 2017; MacParland et al., 2018; Aizarani et al., 2019), brain (Darmanis et al., 2015; Lake et al., 2018; Rosenberg et al., 2018), kidney (Magella et al., 2018), lung (Treutlein et al., 2014; Xu et al., 2016), or even whole animals (Drosophila, mouse embryonic stages: Karaiskos et al., 2017; Mohammed et al., 2017). The first comprehensive analyses of tissues and developing organisms show that, besides identifying novel (previously unspecified) cell types or cell states, it is possible to identify similarities and changes of cell functions across tissues, to capture the transition of cell populations, i. e., the dynamics of appearance and disappearance during development, to determine the heterogeneity of cell types in diseased tissues (e.g., cancer), and to follow the variation of cell composition in aging tissues, to name only a few of the most intriguing conclusions.

1.3 EFFORTS IN THE FIELD OF SINGLE-CELL OMICS

With the emergence of single-cell applications, it soon became clear that comparative analysis would require some kind of standardization at both the experimental and data interpretation levels. International research consortia such as the Human Cell Atlas (HCA)[8] or LifeTime[9] were established in order to take the lead in these tasks and to rapidly develop this fast growing field of research by providing high-quality single-cell data with defined standards and controls in reference databases. The first databases for mouse, human and Drosophila have been established, from which cell-specific single-cell data can already be retrieved.[10] The Human Cell Atlas was the first consortium formed in 2017 with the goal of generating a comprehensive single-cell atlas of all human cells and of developing new cloud-based informatics solutions for data storage and analysis. The European LifeTime initiative launched in 2018 complements these efforts by focusing on medical applications in several disease-related areas. A major goal of LifeTime is to develop and analyze disease-related models and to produce novel approaches that can be transferred into clinical use of single-cell data. In Germany, the Single Cell Omics Germany (SCOG) network[11] supported by the Federal Ministry for Education and Research (BMBF) was founded in 2018 with the aim of establishing the first network of laboratories performing single-cell analyses and offering further education in single-cell technologies, especially in emerging fields such as single-cell multi-omics, comprehensive data analysis and interpretation.

The intention of all of these joint efforts is to establish a scientific community working together on a complete atlas of all cell types of the human body at single-cell resolution, in the context

8 See: https://www.humancellatlas.org/ [13.08.2019].

9 See: https://lifetime-fetflagship.eu/ [13.08.2019]. See also Junker, Popp, Rajewsky, Chapter 2.

10 Comprehensive list of databases: https://www.singlecell.de/index.php/resources/databases/ [13.08.2019].

11 See: https://www.singlecell.de/ [16.08.2019].

of the body/tissue, organ aging and disease, to make these data freely available to the research community, and to develop new informatics approaches for deep biological interpretation. The complexity of single-cell data and the many new questions that can be addressed with this type of high-resolution data require the development of new bioinformatics approaches which go far beyond the applications developed for bulk NGS data. Aliee, Sacher and Theis outline this in more detail in Chapter 4.

1.4 UPCOMING DEVELOPMENTS IN SINGLE-CELL OMICS

Currently, the majority of single-cell omics assays focus on RNA-seq, mainly capturing expression signatures of the last exon[12] of a gene (e.g., Chromium (Zheng et al., 2017), Drop-seq). While such approaches are fairly robust and are well-suited for generating cell signatures that allow the distinction of major cell types, they do not investigate more sophisticated changes in gene regulatory programs such as alternative transcriptional or spliced isoforms of genes which often play a different functional role. Therefore, deeper and more comprehensive RNA-seq methods are emerging to capture the entire spectrum of gene transcription, such as Smart-seq2 (Picelli et al., 2013) and others (Chen et al., 2019). While such approaches are still rather costly, decreasing sequencing costs will make them become more and more a routine – simply because the richness of such data allows much better and deeper interpretations.

The generation of a high resolution signature from single cells is at the expense of the loss of spatial orientation of the individual cells. This means that the researchers have no information about the evironment and position the cell had within its tissue. However, this knowledge is important for the interpretation of the single-cell data and their integration with existing knowledge about the tissue or the organism. To overcome this restriction problem, methods are being developed that generate a frame for the spatial reconstruction of single-cell omics data into "virtual tissues". The main current methods use high-resolution imaging data to localize expression signatures in tissues for a sufficient number of genes and cells, for example by multicolor RNA-FISH. Such methods allow the quantification of the relative expression of genes "in situ" in single cells of tissue slices. The expression map of these "marker" genes can then be used as anchors for a spatial reconstruction of the single-cell RNA-seq signatures in order to subsequently generate a kind of "virtually" reconstructed tissue from the single-cell data. This means it is possible to know exactly when and where a gene is being expressed in which cell of an organism. This aspect is further discussed by Junker, Popp and Rajewsky in Chapter 2. An alternative upcoming approach to generate anchor points is to collect few or single cells from defined regions within tissue slices by laser capture microscopy followed by deep (single-)cell sequencing (Nichterwitz et al., 2016; Chen et al., 2017).

A comprehensive and mechanistic interpretation of single-cell omics data comes with a comparison to other omics data, including genomic and functional epigenomics data (changes in

12 Exons are the part of the original RNA transcript that make up the mRNA after a processing step known as "splicing". To detect the last exon serves as proof that the gene has been transcribed.

chromatin and DNA modifications), ideally generated in single cells. So far, genomics and functional genomics reference data have been produced for "bulk"[13] cell types by consortia like the International Cancer Genome Consortium (ICGC),[14] 4DNucleome,[15] the Encyclopedia of DNA Elements (ENCODE),[16] the International Human Epigenome Consortium (IHEC)[17] and others. For some, such as e.g., ATAC-seq, single-cell-based applications have been developed (Buenrostro et al., 2015) and already have been commercialized. However, most of the genomics and epigenomics NGS-based methods are technically challenging and difficult to apply at the single-cell level. Until recently it seemed impossible to link the expression of single cells directly to the epigenetic profiles of single cells. Pilot experiments by Clark et al. (2018) have now shown that also single-cell gene expression, DNA-methylation and chromatin data[18] can be obtained simultaneously from the same cell. In addition, the first highly technical approaches were developed to determine the three-dimensional configuration of chromosomes in single cells, providing insights into the spatial organization of genes in the cell nucleus and the importance for the regulation of gene activity (Nagano et al., 2017).

The integrated interpretation of such multi-omics single-cell data constitutes an important emerging field in single-cell biology, as it builds bridges between (descriptive) transcriptional signatures of individual cells and the mechanisms by which these gene programs are established and executed. Functional multi-omics data will allow researchers to address biomedical questions at a resolution never reached before and pave the road for a precise understanding of mechanisms regulating gene expression. However, the complexity of the generated data poses great challenges for single-cell bioinformatics, as different data types (with different dynamic ranges) have to be integrated and analyzed.

1.5 SINGLE-CELL ANALYSIS AND DEVELOPMENTAL BIOLOGY

Studies in developmental biology will benefit greatly from the use of single-cell (multi-)omics. Single-cell resolution will provide a novel comprehensive view on cell program changes and their dynamic adaptation during development. It is likely that single-cell data will change our current view on (stochastic and directed) mechanisms that drive differentiation processes, and possibly also our (rather static and pre-knowledge-based) view on cell type definition. It will certainly also enhance our understanding of how cells adapt to changing environmental conditions.

13 "Bulk" refers to more or less homogeneous cell type mixtures, where the average gene expression or epigenetic modifications are measured over thousands to millions of cells. In contrast, single-cell omics data focuses on individual cells, not bulks of cells.

14 See: https://icgc.org/ [13.08.2019].

15 See: https://www.4dnucleome.org/ [13.08.2019].

16 See: https://www.encodeproject.org/ [13.08.2019].

17 See: http://ihec-epigenomes.org/ [13.08.2019].

18 "Chromatin" is a complex of the DNA strand and associated proteins. The DNA is wrapped around the so-called histone proteins and twisted in itself. The degree of twisting (condensation) influences the accessability of the chromatin for further binding proteins which can for example activate or inactivate genes. DNA-methylation is a biochemical modification of the DNA that affects the binding behavior of regulatory proteins and the chromatin conformation and thus influences gene expression.

First impressive examples of "developmental fate maps" have been generated for planarians (Cao et al., 2017), early mouse embryos (Peng et al., 2016; Mohammed et al., 2017) and Drosophila larvae (Karaiskos et al., 2017). These data provide insights into the dynamics of changes in cellular programs that occur during rapid phases of self re-organizing processes (e.g., during gastrulation). They allow to track the formation and organization of cells spatially and temporally during brain (and kidney, heart etc.) development and to identify the determined and stochastic mechanisms driving cellular diversification and differentiation (e.g., during early mammalian development). Complex computational models have been established to infer the dynamics of developmental trajectories of cell lineages and to follow the cellular transitions across lineage commitment. In combination with genetic labeling or overlaying with microscopic reference data, such high resolution omics data will provide a deep understanding of mechanisms regulating spatial and temporal organization of developmental transitions in various organisms. These aspects will be outlined in more detail in Chapter 2 by Junker, Popp, Rajewsky, and Chapter 4 by Aliee, Sacher, Theis.

1.6 SINGLE-CELL ANALYSIS IN BIOMEDICAL RESEARCH

The composition and relative localization of cells in a tissue and an organ is an important parameter to understand the physiological functions related to natural organ function, homeostasis, aging, regeneration, but also to diseases. Single-cell omics offer an unbiased approach to investigate the precise relationship between cellular composition and organ biology.[19] Moreover, consequences of local dysfunction of cells in the organ can be traced, for example, in processes leading to wounding, scar formation, fibrosis, steatosis etc. Single-cell analysis will allow to directly address changes in cell composition that occur in pathological situations like in abnormal organ development, in autoimmune diseases, in chronic diseases or in cancer. The determination of the cellular heterogeneity in solid tumors or in leukemic cells will open up a new diagnostic level to determine origin, progression and heterogeneity of the tumor and offers a tumor specific diagnosis and prediction for therapy responses (see Aschenbrenner, Mass, Schultze, Chapter 3 for further details). The extensive possibilities for medical application of single-cell analysis raise the question as to whether this field of work may evoke ethical questions. This aspect will be discussed by Fangerau, Marx-Stölting and Osterheider in Chapter 6.

1.7 CHALLENGES AND LIMITATIONS OF SINGLE-CELL TECHNOLOGIES

As with all new technological developments, single-cell analyses come with technical and conceptual challenges. One challenge is the preparation of high quality single-cell suspensions from complex tissues as well as limitations in obtaining sufficient high-quality RNA or DNA (but also lipids and proteins) from single cells (see Müller-Röber, Chapter 5 for further discussion of this

19 "Unbiased" in this context means that researchers can obtain results without narrowing down possible results according to their own prior hypothesis before running the experiment.

problem regarding plant cells). Physical constraints like the difficulty to singularize cells (e.g., in brain tissues) can compromise the experiment. New approaches of using isolated nuclei and analyzing nascent RNA may overcome some of these problems (Krishnaswami et al., 2016) even allowing to study diseases such as Alzheimer's disease in preserved post mortem tissues (see also Aschenbrenner, Mass, Schultze, Chapter 3).

The analysis of non-nucleic acid-based parameters, such as the presence of proteins, lipids and metabolites, at the single-cell level will be essential for the interpretation and modeling of single-cell multi-omics data. First successful implementations have been published, but the technical possibilities for a comprehensive representation of proteome and metabolome as well as lipidome data from single cells remain very limited (Marx, 2019; Duncan et al., 2019; Pasarelli et al., 2019). Single-cell analyses depend on methods in which cells are grouped according to their similarities in expression profiles, it is important to control whether the obtained distinct groups/patterns represent the expected and complete spectrum of cells in the starting single-cell suspension. For many applications, particularly the detection of rare cells, a sufficient efficiency in single-cell library construction and in sequencing depth needs to be considered. This is usually associated with high preparation and sequencing costs. Furthermore, experimental and bioinformatical standards need to be established to avoid over-interpretation of single-cell NGS data due to "single gene" dropout effects[20] (Van den Berge et al., 2018). Finally, the identification and biological interpretation of grouped cells require some a priori knowledge, for compositional estimates an approximate knowledge of the number of cell types, and for spatial reconstruction an orientation by cell specific „marker genes" (Aliee, Sacher, Theis, Chapter 4). New and old approaches will have to be developed (further) to meet this need.

1.8 FINAL REMARK

Single-cell omics is a fast growing and extremely important area in functional genomics. Its broad spectrum of applications and data usage will revolutionize and enrich modern biology and medicine in many aspects and drive them into a new deep molecular dimension. It will shed new light on concepts of cell and systems biology which will be explored in greater depth. NGS-based single-cell data will influence almost every biological field ranging from basic cell biology to developmental biology, from physiology to pathology, from taxonomy to ecology. Single-cell diagnostics is furthermore one of the hottest emerging fields in personalized medicine with high potential to raise precision diagnostics to a new level.The success of single-cell analysis very much depends on the development of novel experimental and bioinformatical solutions. The core structures for such a development are given but progress in this field is extremely fast and requires a constant investment and adjustment.

20 A dropout effect has occurred, if expressed transcripts are not being detected due to technical reasons, e.g., inefficient re-transcription of RNA into cDNA. Such effects can lead to an excess of zero read counts when single-cell data is compared to bulk RNA-seq data.

1.9 REFERENCES

Aizarani, N. et al. (2019): A human liver cell atlas reveals heterogeneity and epithelial progenitors. In: Nature 572: 199–204.

Buenrostro, J. D. et al. (2015): Single-cell chromatin accessibility reveals principles of regulatory variation. In: Nature 523(7561): 486.

Cao, J. et al. (2017): Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357(6352): 661–667.

Chen, J. et al. (2017): Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. In: Nature protocols 12(3): 566.

Chen, G./Shi, T. (2019): Single-cell RNA-seq technologies and related computational data analysis. In: Frontiers in genetics 10: 317.

Clark, S. J. et al. (2018): scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. In: Nature communications 9(1): 781.

Darmanis, S. et al. (2015): A survey of human brain transcriptome diversity at the single cell level. In: Proceedings of the National Academy of Sciences 112(23): 7285–7290.

Duncan, K. D. et al. (2019): Advances in mass spectrometry based single-cell metabolomics. In: Analyst 144(23): 7285-7290.

Gierahn, T. M. et al. (2017): Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. In: Nature methods 14(4): 395.

Halpern, K. B. et al. (2017): Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. In: Nature 542(7641): 352.

Han, X. et al. (2018): Mapping the mouse cell atlas by microwell-seq. In: Cell 172(5): 1091–1107.

Jaitin, D. A. et al. (2014): Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. In: Science 343(6172): 776–779.

Karaiskos, N. et al. (2017): The Drosophila embryo at single-cell transcriptome resolution. In: Science 358(6360): 194–199.

Keren-Shaul, H. et al. (2019): MARS-seq2. 0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. In: Nature protocols 14(6): 1841.

Kowalczyk, M. S. et al. (2015): Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. In: Genome research 25(12): 1860–1872.

Krishnaswami, S. R. et al. (2016): Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. In: Nature protocols 11(3): 499.

Lake, B. B. et al. (2018): Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. In: Nature biotechnology 36(1): 70.

Macosko, E. Z. et al. (2015): Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. In: Cell 161(5): 1202–1214.

MacParland, S. A. et al. (2018): Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. In: Nature communications 9(1): 4383.

Magella, B. et al. (2018): Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. In: Developmental biology 434(1): 36–47.

Marx, V. (2019): A dream of single-cell proteomics. In: Nature methods, online publication 12.08.2019. DOI: 10.1038/s41592-019-0540-6.

Mohammed, H. et al. (2017): Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. In: Cell reports 20(5): 1215–1228.

Nagano, T. et al. (2017): Cell-cycle dynamics of chromosomal organization at single-cell resolution. In: Nature 547(7661): 61.

Nichterwitz, S. et al. (2016): Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. In: Nature communications 7: 12139.

Passarelli, et al. (2013): Single-cell lipidomics: characterizing and imaging lipids on the surface of individual Aplysia californica neurons with cluster secondary ion mass spectrometry. In: Analytical chemistry 85(4): 2231–2238.

Peng, G. et al. (2016): Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. In: Developmental cell 36(6): 681–697.

Picelli, S. et al. (2013): Smart-seq2 for sensitive full-length transcriptome profiling in single cells. In: Nature methods 10(11): 1096.

Rosenberg, A. B. et al. (2018): Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. In: Science 360(6385): 176–182.

Tang, F. et al. (2009): mRNA-Seq whole-transcriptome analysis of a single cell. In: Nature methods 6(5): 377.

Treutlein, B. et al. (2014): Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. In: Nature 509(7500): 371.

Van den Berge, K. et al. (2018): Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. In: Genome biology 19(1): 24.

Wilson, N. K. et al. (2015): Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. In: Cell stem cell 16(6): 712–724.

Xu, Y. et al. (2016): Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. In: JCI insight 1(20).

Zheng, G. X. Y. et al. (2017): Massively parallel digital transcriptional profiling of single cells. In: Nat. Commun. 8(14049), online publication 16.01.2017. DOI: 10.1038/ncomms14049.

J. Philipp Junker, Christian Popp, Nikolaus Rajewsky

2. SINGLE-CELL GENOMICS IS TRANSFORMING DEVELOPMENTAL BIOLOGY

2.1 INTRODUCTION

New developments in single-cell genomics have transformed developmental biology over the past few years. Researchers were quick to recognize the potential of single-cell transcriptomics[1] for unbiased and systematic identification of cell types which constitutes a major improvement on previously used approaches based, for example, on cell morphology or a small number of marker genes. New comprehensive cell-type atlases are an extremely valuable resource for the scientific community: for instance, they enable a more systematic analysis of the effects of mutations by revealing the cell type in which the mutated gene is expressed (e.g., Human Cell Atlas, tabula muris, fly cell atlas, which aim to identify all cell types in the respective organism based on single-cell transcriptomics).[2] However, as discussed in more detail below, current efforts in single-cell genomics in developmental biology are moving beyond cell-type identification toward functional information about effects of perturbations, the origin of cell types, differentiation trajectories, spatial architecture of tissues, and mechanisms of gene regulation (Griffiths et al., 2018). Due to their genetic accessibility, their high degree of experimental reproducibility, and the detailed understanding of major developmental mechanisms accumulated over decades of research, developmental biology is currently serving as a testbed for new experimental and computational methods, which often go on to be applied to disease models or human patient samples. Single-cell transcriptomics is by far the most advanced of the single-cell omics technologies and will hence take up the largest part of this review. However, single-cell measurement of other parameters, in particular protein abundance, DNA-methylation, and open chromatin profiling,[3] are progressing rapidly and will also be discussed.

2.2 PERTURBATION ANALYSIS IN DEVELOPMENTAL MODEL SYSTEMS

Developmental models that are frequently used for single-cell analysis include the classical animal model organisms like fruit fly, zebrafish and mouse, which are very well suited for genetic

1 Transcriptomics is the study of the entirety of transcripts within a cell, encompassing its RNA.

2 All of these projects are about establishing reference maps of all cells of the organisms studied (e.g., human, mouse, fly). By way of example, it is the aim of the Human Cell Atlas "to create comprehensive reference maps of all human cells – the fundamental units of life – as a basis for both understanding human health and diagnosing, monitoring, and treating disease". See: https://www.humancellatlas.org/ [24.06.2019].

3 Protein abundance refers to the amount of protein available in a single cell; DNA-methylation is a common biochemical modification of DNA that influences the state of the chromatin (the complex of DNA and accompanying proteins), which in turn influences and regulates gene expression. Open chromatin refers to the parts of the genome that are accessible for binding of regulatory proteins. These regions are typically involved in controlling gene expression.

perturbation studies.[4] An important recent addition are organoids[5] derived from human patient material, which, for the first time, make human tissue accessible for genetic manipulation (Camp/Treutlein, 2017). This is particularly important for human brain organoids, due to the unique properties of the human brain, which are often recapitulated poorly in animal models. The single-cell analysis enables studies of the effect of genetic perturbations on cell fate decisions to be performed. This is already carried out regularly by comparison of wildtype and mutant animals, or, as in the case of a recent publication focusing on early mesoderm specification, by analyzing mosaic animals (Pijuan-Sala et al., 2019). In this study, the authors created a chimeric mouse embryo consisting of wildtype and $Tal1^{-/-}$ cells, which allowed them to directly compare the differentiation potential of the wildtype and mutant cells in the same animal.[6]

Beyond the classical genetic perturbations, single-cell analysis is also ideally suited to dissect the molecular and cellular impact of other perturbations. Regeneration of the axolotl limb after amputation is a particularly powerful example of this type of application (Gerber et al., 2018). In this study the authors focused on identifying the cell types that transiently appear at the site of injury to drive skeletal regeneration. By combining single-cell analysis with Brainbow-based lineage tracing (a method that uses fluorescent proteins to stain individual cells), they found that there are no pre-existing stem cells. Instead, they observed that a heterogeneous population of fibroblasts dedifferentiated to form a multipotent skeletal progenitor expressing the embryonic limb program.

In the past few years, we have witnessed an increased interest in non-standard model organisms. This is largely due to the fact that single-cell transcriptomics makes identification of cell types and differentiation pathways much easier. Furthermore, the emergence of CRISPR/Cas9[7] gene editing provides a simple tool for making transgenic animals in many species. Besides the work on axolotl mentioned above, other notable examples include evolutionary studies on annelids and the cnidarian *nematostella* (Achim et al., 2018; Sebé-Pedrós et al., 2018). These projects, together with novel computational methods for comparing single-cell datasets across different species, are beginning to yield interesting insights into the evolution of cell types.

The combination of single-cell genomics and CRISPR/Cas9 genome editing has not only led to a renewed interest in non-standard model systems, but is also an inspiring method development in other fields such as perturbation screens and lineage tracing (with the latter being discussed in more detail below). In CRISPR screens, cultured cells are transfected with Cas9 and a library

4 Perturbation studies interrupt specific genes in order to examine the effect that this perturbation has on the development of the cell studied. Thus, the function of the gene can be inferred.

5 Organoids are three-dimensional stem cell cultures that resemble organs. They are multicellular entities, have the ability to form three-dimensional structures, and display functions that are typical for the resembled organ (Bartfeld/Clevers, 2018).

6 Mosaic animals are animals consisting of at least two different cell populations with a different chromosome content. The described chimeric mouse embryo is one such mosaic, containing normal cells ("wildtype") as well as so-called $Tal1^{-/-}$ cells that lack the transcription factor Tal1 on both chromosomes. Tal1 (the abbreviation represents the name "T-cell acute lymphocytic leukemia protein 1") plays a role in regulating genes connected to leukemia.

7 CRISPR/Cas is a method used to edit genes in a specific manner by cutting at desired loci in the genome that are determined by a guide RNA.

of sgRNAs[8] that target a large number of different genes. Readout by single-cell transcriptomics then allows for association of the activity of a specific sgRNA with changes in gene expression (Dixit et al., 2016; Jaitin et al., 2016; Datlinger et al., 2017). While these methods are currently limited to cultured cells and are not yet applicable to developmental model systems, they hold great promise for systematic identification of gene regulatory networks.

2.3 SPATIAL INFORMATION

In tissues and organs, cells are organized in intricate spatial structures that are necessary for their proper function. Furthermore, cells are heavily influenced by their surroundings (e.g., stem cell niches) and the signals they send to each other. However, single-cell genomics typically requires dissociation of samples into a single-cell suspension, so in most approaches all information about spatial organization is lost. It is currently a major focus of both academic and industrial research alike to retain spatial information in single-cell analysis. These approaches can roughly be grouped into three categories:

1) Methods that make use of additional spatial information recorded by microscopy. If the spatial expression patterns of cell-type specific marker genes are known, cells can be positioned according to these landmark genes (Satija et al., 2015; Karaiskos et al., 2017). Importantly, the Rajewsky and Friedman labs have recently shown that spatial expression patterns can largely be derived from first principles even without additional information, since the majority of genes is expressed in simple patterns and smooth transitions (novoSpaRc[9]).

2) Another class of methods uses sequential rounds of single-molecule fluorescence in situ hybridization (FISH)[10] to detect transcripts with spatial resolution. While until very recently these methods were limited to profiling hundreds of genes, transcriptome-wide transcription imaging was recently reported (Eng et al., 2017). One important advantage of these imaging-based techniques is that they have much higher transcript recovery rates than sequencing-based approaches, since they rely on hybridization rather than on the often inefficient reverse transcription reaction. However, still today these methods remain laborious to set up and time-consuming to operate.

3) Finally, there are novel methods that add molecular barcodes (in the form of short DNA sequences) encoding spatial information directly in tissue slices. These approaches are either based on arrays of barcoded primers for reverse transcription spotted on a surface (Ståhl et al., 2016) or on barcoded beads that are positioned on a surface (Rodriques et al., 2019).

8 A library in this context is a collection of similar molecules, in this case sgRNA. SgRNA is single guide RNA and directs Cas9 to the place where it is supposed to cut.

9 novoSpaRc is a computational method that predicts locations of single cells in space by solely using single-cell RNA sequencing data. It transposes distances of single cells in expression space to their physical distances across tissues.

10 FISH is a technique that uses the specific binding of fluorescent probes to nucleic acid sequences.

2.4 TEMPORAL INFORMATION – PSEUDO-TEMPORAL ORDERING

Besides spatial information, another important challenge in single-cell genomics is inclusion of temporal information. Since the cells are destroyed during sequencing, it is impossible to follow their expression changes and fate decisions in real time. However, if the number of cells that are sampled is big enough, even extremely transient (and hence rare) states can be detected in the dataset. This allows for an ordering of cells along an inferred pseudo-temporal trajectory[11] (Moignard et al., 2015; Setty et al., 2016; Haghverdi et al., 2016; Kester/van Oudenaarden, 2018). For short-term processes that occur continuously (e.g., hematopoiesis), the entire process of transcriptional changes can therefore be sampled and reconstructed computationally in a single experiment (see also: Aliee, Sacher, Theis, Chapter 4).

While methods for pseudo-temporal ordering of single-cell transcriptomics efficiently orient cells along continuous trajectories, the directionality of the differentiation process is not obvious from the data alone. However, La Manno et al. (2018) recently introduced RNA velocity, a computational method that infers the direction in gene expression space in which cells are moving based on unspliced vs. spliced[12] (i.e. "old" vs. "new") transcript molecules. Another emerging method for looking into the immediate future of cells is RNA metabolic labeling (Hendriks et al., 2018; Erhard et al., 2019), which allows for separation of old from new molecules based on labels that are introduced experimentally into RNA molecules during a defined time window.

Despite the relative novelty of pseudo-temporal ordering, there are already numerous biological applications. These include a complete differentiation trajectory of planaria (Plass et al., 2018) and a study that revealed transitions between veins and arteries during coronary development in mice (Su et al., 2018). Expansions of the approach measure single-cell transcriptomes at different developmental stages and then computationally stitch the individual time points together to form continuous trajectories (Farrell et al., 2018; Wagner et al., 2018).

2.5 TEMPORAL INFORMATION – HIGH-THROUGHPUT LINEAGE ANALYSIS

While pseudo-temporal ordering and RNA metabolic labeling yield short-term temporal information, it is often desirable to record relationships of cells over longer periods of time, ranging from days to months and years. The field of lineage tracing has a long history of using visual markers (e.g., fluorophores) to label and track cells. More recently, with the emergence of single-cell genomics, it has become possible to use the enormous information storage capacity of the genome to determine the lineage relationships of cells. Sequence-based methods for lineage

11 For pseudo-temporal analysis, the sequenced cells are ordered by the similarity of their transcriptome. The resulting sequence of single-cell transcriptomes is called a "trajectory" and is interpreted as a temporal succession of cell states, e.g., a gradual transition from the stem cell state to a differentiated state.

12 After transcription, RNA is subject to modifications leading to a maturation of the RNA. "Splicing" is the process through which certain parts of the original RNA (introns) are cut and discarded, while the remaining parts (exons) are connected to establish a mature RNA.

analysis generally fall into two categories: those that use naturally occurring mutations; and those that seek to actively modify the genome.

In theory, naturally occurring somatic mutations (such as single nucleotide variants or copy number variations) are powerful lineage markers that can be read by sequencing. Since lineage tracing by somatic mutations is non-invasive and does not require continuous observation, it is ideally suited for studying human samples. In the last few years, pioneering studies have started to apply this strategy to early embryonic lineage decisions. In organoids derived from single mouse cells (Behjati et al., 2014) and in human blood samples analyzed in bulk (Lodato et al., 2015), analysis of somatic mutations allowed reconstruction of early embryonic lineage trees. In a recent landmark paper published by the Walsh lab, the authors placed neurons from postmortem human brains in a developmental lineage tree after whole genome amplification and sequencing of single cells (Ju et al., 2017). However, general applicability of this approach is currently hampered by the high cost of sequencing the whole genome of large numbers of single cells. Lineage tracing based on mutations in mitochondria (which have a much higher mutation rate) offers a promising alternative for high-throughput lineage tracing in humans (Ludwig et al., 2019).

While these approaches are ideally suited for human samples, for model organisms, lineage tracing techniques that are based on experimental manipulation are typically the better choice due to the higher degree of control. Experimentally controlled genome modifications for lineage tracing can be achieved via recombination of synthetic Cre-lox cassettes[13] (Pei et al., 2017) or by using CRISPR/Cas9 technology. High-throughput lineage tracing based on CRISPR/Cas9, combined with cell-type identification by single-cell RNA sequencing, has recently been established in zebrafish (Alemany et al., 2018; Raj et al., 2018; Spanjaard et al., 2018) and in mice (Kalhor et al., 2018; Chan et al., 2019). While many experimental and computational challenges remain, CRISPR/Cas9 lineage tracing holds great promise as a general approach to identify the developmental origin of cell types and to understand the mechanisms of cell-type dependent diseases.

2.6 MEASURING OTHER PARAMETERS BEYOND RNA

As mentioned above, RNA sequencing is by far the most advanced technology in single-cell genomics. However, measurement of other parameters is rapidly catching up, in particular with regard to identification of open chromatin profiling, DNA-methylation, and single-cell protein detection. Single-cell ATAC-seq (scATAC-seq), a transposase-based method for open chromatin profiling, can now routinely be performed in thousands of cells due to new protocols for combinatorial barcoding of single cells. Applications include atlases of chromatin accessibility in mice (Cusanovich/Hill et al., 2018) and in drosophila development (Cusanovich/Reddington et al., 2018). In a remarkable recent publication, Yoshida et al. (2019) generated matched epigenome and transcriptome measurements in 86 primary cell types that span the mouse immune system and its differentiation cascades. They found that genes fall into two distinct classes, controlled by either

13 Cre-lox cassettes are a system of enabling gene deletions in specific cell lineages in living animals. Using this technology, specific cell types or tissues may be genetically modified, while others are not.

enhancer- or promoter-driven logic. Relating transcription factor expression to the genome-wide accessibility of their binding motifs classifies them as predominantly openers or closers of local chromatin accessibility.

While scATAC-seq is rapidly being adopted by the scientific community, the use of single-cell DNA-methylation analysis has so far remained restricted to relatively few laboratories, which is probably largely due to the high cost of DNA-methylation analysis. However, single-cell DNA-methylation has already yielded important insights, in particular in early development. For instance, Rulands et al. (2018) identified unexpected genome-scale oscillations in DNA-methylation during exit from pluripotency. Importantly, detection of DNA-methylation has already been combined successfully with measurement of RNA from the same single cells (Clark et al., 2018).

While protein detection in single cells has not yet been successfully established on the level of the full proteome, there are already highly promising approaches for detection of panels of proteins: Single-cell mass cytometry (Bendall et al., 2011) allows for parallel detection of a large number of proteins in single cells by using specific antibodies labeled with heavy metals. Antibodies coupled to distinct transition element isotopes are used to bind to their epitopes. Individual cells are then vaporized and ionized in a plasma, and elemental ions are detected by time-of-flight mass spectroscopy.[14] Another sequencing-based approach, CITE- seq (Stoeckius et al., 2017) enables simultaneous detection of oligonucleotide-labeled antibodies and transcriptome measurements in an efficient single-cell readout.

With more and more datasets using different measurement techniques which became available, there is a growing realization in the field that novel computational methods for data integration are needed. This includes, for instance, matching of cell types identified by scRNA-seq and scATAC-seq, but also removal of differences between scRNA-seq datasets that are due to technical artifacts (e.g., batch effects caused by dissociation techniques). Several promising computational approaches have recently been proposed (Barkas et al., 2018; Butler et al., 2018; Haghverdi et al., 2018).

Single-cell analysis is transforming our understanding of development. With the help of new methods and approaches as described in the review above, developmental biologists have gained tools that allow them to unlock long-kept secrets in spatial and temporal tissue organization. But this fundamental knowledge does not just give us a better view of biological processes in a healthy state. It also gives us the opportunity to hone in on deviations from the norm that lead to disease. The last paragraph of this review will focus on a new initiative, LifeTime, that aims to harness the power of single-cell analysis to advance understanding, early diagnosis, interception and treatment of a wide range of diseases toward innovative and personalized medicine.

14 In time-of-flight mass spectrometry, an ion's mass-to-charge ratio is determined via a time-of-flight measurement. Ions are accelerated by an electric field of known strength.

2.7 LIFETIME, A NEW INITIATIVE BUILT ON SINGLE-CELL ANALYSIS

The power of single-cell analysis technologies has not only been recognized by the journal *Science* recently, who named it "Breakthrough of the year 2018" (Pennisi, 2018) but has also inspired the creation of the European Research Initiative, LifeTime. The consortium comprises hundreds of researchers in 18 European countries and is supported by more than 70 companies, by all major European science academies and many national governments. Its mission is to map, understand and target human cells for treatment during disease in patients. By harnessing the full potential of single-cell technologies, artificial intelligence and individualized experimental disease models (such as organoids), LifeTime researchers want to be able to better predict the onset of diseases and/or cure them by analyzing a patient's own tissue. To achieve this, it will be necessary to be able to understand how genomes function within cells – something only the cell itself is currently capable of – to decipher how cells form tissues and to identify the dynamics that lead from a healthy cell or tissue to a pathological state.

Single-cell technologies offer a great opportunity in overcoming some of the fundamental shortcomings in our current scientific approaches, such as resolving spatial cellular heterogeneity or capturing cellular changes in time. Importantly, the advent and utilization of new computational tools and artificial intelligence will be essential to achieve LifeTime's mission. It provides the power required to integrate the data generated and will allow for not only an understanding of the healthy state but also of the cause and biology of disease. The improvement of experimental disease models by employing new technologies in genome manipulation and cell reprogramming will enable LifeTime researchers to manipulate the genomes and cells from patient tissues.

To facilitate the profiling of multiple layers of genome regulation – an important step in achieving LifeTime's goals – single-cell multi-omics and imaging will need to be further developed and integrated (e.g., transcriptome, epigenome, metabolome, proteome, etc.). Also, major efforts in experimental scaling will be necessary to arrive at the required sample throughput with the appropriate analytic resolution. Judging by the experience in the evolution of other technologies in the past (e.g., DNA sequencing), adequate progress can be expected within several years, accompanied by significant reductions in cost. This is in fact already happening with some single-cell technologies currently operating in the order of millions of cells per sample.

Further developments will also be required in the other technological pillars that LifeTime is being built on, not least at the intersection of technological fields. Some are of course intrinsically tied to each other: for example, the adaptation of computational and statistical techniques to the scaling and integration of single-cell multi-omics. Similarly, as detailed molecular and spatial cell reference maps are becoming available, new machine learning tools will be required to facilitate the integration of patient trajectories and to predict a patient's disease trajectory from electronic health records. It will also be necessary to develop new computational methods to understand mechanistically what drives a cell's transcriptional state and ultimately its precise function in the context of a specific tissue (see also Aliee, Sacher, Theis, Chapter 4).

This improved knowledge of cause and effect in cellular regulation profiles will enable a true move towards personalized medicine. By applying it to patient-matched organoids and organ-on-chip models, which will also be developed further and improved, LifeTime researchers want to facilitate the translation into clinics. This will spur the transformation of personalized experimental disease models into powerful predictive systems.

The LifeTime tool kit of methods and technologies will be amenable to a wide range of diseases. These include neurological disorders, infectious diseases, cancers and many other disease areas. The diseases that will be studied using the LifeTime Technology Platform will be selected through an interactive, transparent and peer-reviewed mechanism, termed the *LifeTime Launchpad*. It will take into account a range of parameters (such as societal impact, heterogeneity on a cellular level, availability of cell models, clinical feasibility etc.) and remain in place during LifeTime's implementation to ensure that new ideas and opportunities can be explored as they arise.

2.8 REFERENCES

Achim, K. et al. (2018): Whole-Body Single-cell sequencing reveals transcriptional domains in the annelid larval body. In: Molecular biology and evolution 35(5): 1047–1062.

Alemany, A. et al. (2018): Whole-organism clone tracing using single-cell sequencing. In: Nature, online publication 28.03.2018. DOI: 10.1038/nature25969.

Barkas, B. et al. (2018): Wiring together large single-cell RNA-seq sample collections. In: BioRxiv, online publication 02.11.2018. DOI: 10.1101/460246.

Bartfeld, S./Clevers, H. (2018): Aus Stammzellen abgeleitete Organoide und ihre Bedeutung für die biomedizinische Forschung und Therapie. In: Zenke, M. et al.: Stammzellforschung. Nomos, Baden-Baden.

Behjati, S. et al. (2014): Genome sequencing of normal cells reveals developmental lineages and mutational processes. In: Nature 513: 422–425.

Bendall, S. C. et al. (2011): Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. In: Science 332(6030): 687–696.

Butler, A. et al. (2018): Integrating single-cell transcriptomic data across different conditions, technologies, and species. In: Nature Biotechnology 36: 411–420.

Camp, J. G./Treutlein, B. (2017): Human organomics: a fresh approach to understanding human development using single-cell transcriptomics. In: Development 144: 1584–1587.

Chan, M. M. et al. (2019): Molecular recording of mammalian embryogenesis. In: Nature 570: 77–82.

Clark, S. J. et al. (2018): ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in s. In: Nature Communication, online publication 22.02.2018. DOI: 10.1038/s41467-018-03149-4.

Cusanovich, D. A. et al. (2018): A single-cell atlas of in vivo mammalian chromatin accessibility. In: Cell 174(5): 1309–1324.

Cusanovich, D. A. et al. (2018): The cis-regulatory dynamics of embryonic development at single-cell resolution. In: Nature, online publication 14.03.2018. DOI: 10.1038/nature25981.

Datlinger, P. et al. (2017): Pooled CRISPR screening with single-cell transcriptome readout. In: Nature Methods 14: 297–301.

Dixit, A.. et al. (2016): Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. In: Cell 167(7): 1853–1866.e17.

Eng, C. H. L. J. et al. (2017): Profiling the transcriptome with RNA SPOTs. In: Nature Methods 14: 1153–1155.

Erhard, F. et al. (2019): scSLAM-seq reveals core features of transcription dynamics in single cells. In: BioRxiv, online publication 21.01.2019. DOI: 10.1101/486852.

Farrell, J. A. et al. (2018): Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. In: Science 360(6392): eaar3131.

Gerber, T. et al. (2018): Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. In: Science 362(6413): eaaq0681.

Griffiths, J. A. (2018): Using single cell genomics to understand developmental processes and cell fate decisions. In: Molecular Systems Biology, online publication 16.04.2018. DOI: 10.15252/msb.20178046.

Haghverdi, L. et al. (2016): Diffusion pseudotime robustly reconstructs lineage branching. In: Nature Methods 13: 845–848.

Haghverdi, L. et al. (2018): Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. In: Nature Biotechnology 36: 421–427.

Hendriks, G. et al. (2018): NASC-seq monitors RNA synthesis in single cells. In: BioRxiv, online publication 17.12.2018. DOI: 10.1101/498667.

Jaitin, D. A. et al. (2016): Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. In: Cell 167(7): 1883–1896.e15.

Ju, Y. S. et al. (2017): Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. In: Nature 543: 714–718.

Kalhor, R. et al. (2018): Developmental barcoding of whole mouse via homing CRISPR. In: Science 361(6405): eaat9804.

Karaiskos, N. et al. (2017): The drosophila embryo at single-cell transcriptome resolution. In: Science 358(6360): 194–199.

Kester, L./van Oudenaarden, A. (2018): Single-cell transcriptomics meets lineage tracing. In: Cell Stem Cell 23(2): 166–179.

La Manno, G. et al. (2018): RNA velocity of single cells. In: Nature 560: 494–498. Lodato, M. A. et al. (2015): Somatic mutation in single human neurons tracks developmental and transcriptional history. In: Science 360(6256): 94–98.

Ludwig, L. S. et al. (2019): Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. In: Cell 176(6): 1325–1339.

Moignard, V. et al. (2015): Decoding the regulatory network of early blood development from single-cell gene expression measurements. In: Nature Biotechnology 33: 269–276.

Pei, W. et al. (2017): Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. In: Nature 548: 456–460.

Pennisi, E. (2018): Development cell by cell. In: Science 362(6421): 1344–1345. Pijuan-Sala, B. et al. (2019): A single-cell molecular map of mouse gastrulation and early organogenesis. In: Nature 566: 490–495.

Plass, M. et al. (2018): Cell type atlas and lineage tree of a whole complex animal by single cell transcriptomics. In: Science 360(6391): eaaq1723.

Raj, B. et al. (2018): Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. In: Nature Biotechnology 36: 442–450.

Rodriques, S. G. et al. (2019): Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. In: Science 363(6434): 1463–1467.

Rulands, S. et al. (2018): Genome-scale oscillations in DNA methylation during exit from pluripotency. In: Cell Systems 7(1): 63–76.e12.

Satija, R. et al. (2015): Spatial reconstruction of single-cell gene expression data. Nature Biotechnology 33: 495–502.

Sebé-Pedrós, A. et al. (2018): Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-seq. In: Cell 173(6): 1520–1534.e20.

Setty, M. et al. (2016): Wishbone identifies bifurcating developmental trajectories from single cell data. In: Nature Biotechnology 34: 637–645.

Spanjaard, B. et al. (2018): Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. In: Nature Biotechnology 36: 469–473.

Ståhl, P. L. et al. (2016): Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. In: Science 353(6294): 78–82.

Stoeckius, M. et al. (2017): Simultaneous epitope and transcriptome measurement in single cells. In: Nature Methods 14: 865–868.

Su, T. et al. (2018): Single-cell analysis of early progenitor cells that build coronary arteries. In: Nature 559: 356–362.

Wagner, D. E. et al. (2018): Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. In: Science 360(6392): 981–987.

Anna C. Aschenbrenner, Elvira Mass, Joachim L. Schultze

3. THE PAST, PRESENT AND FUTURE OF SINGLE-CELL OMICS IN BIOMEDICINE

3.1 SINGLE-CELL OMICS[1] FORGES NEW PATHS AHEAD BY CHANGING OUR PERCEPTION AND TREATMENTS OF MAJOR DISEASES

Cellular heterogeneity within tissues has been a major obstacle in understanding and treating diseases such as cancer, chronic inflammatory diseases, autoimmune diseases, infections, or neurodegeneration. Previous approaches in genomics were restricted to bulk analyses providing only results averaged across all sampled cells. Probing cellular heterogeneity at single-cell resolution became possible only in the past few years and is now applied world-wide to understand the underlying mechanisms and thereby the pathogenesis of these diseases. Here, we provide an overview of how single-cell omics starts to revolutionize our view on cancer and neurodegenerative diseases, how basic research is currently implemented in the clinical setting and with which innovative and experimental ideas we will be able to forge new paths in order to help patients and allow individualized treatment.

3.2 THE PAST OF SINGLE-CELL OMICS

Cancer as the most advanced example for applying single-cell omics to diseases

The tumor microenvironment is characterized not only by different compositions of cancer cell clones but also by infiltrating immune cells and stromal cells. It is unsurprising that comprehensively analyzing cancer on the single-cell level has been a long-sought goal and therefore tumor research has been a driving force in the single-cell omics field. To understand the tumor's ecosystem, single-cell atlases of breast (Wagner et al., 2019), head and neck (Puram et al., 2017), lung (Lavin et al., 2017; Zilionis et al., 2019) and kidney (Chevrier et al., 2017) cancers – just to name a few examples – have already been generated and provide us with valuable insights into the biology of tumors, describe novel biomarkers that allow for conclusions to be drawn about pathogenic processes and define new attractive targets for therapeutic interventions.

Mass cytometry, a method for assessing up to 50 proteins on a single cell, already allows for large cell throughputs. Indeed, profiling of 26 million cells from 144 breast cancer tumors by mass cytometry revealed that 18 % of tumors exhibited patterns of strong T-cell exhaustion (Wagner

1 Single-cell omics is an area of research that focuses on the collective characterization and quantification of single cells that translate into the function and dynamics of tissues. It encompasses several different technologies (e.g., transcriptomics, proteomics or metabolomics) and opens up many different layers of information about the cells (e.g., about the transcripts, the proteins or the metabolites).

et al., 2019), which was accompanied by the expression of the co-suppressive molecule PD-1. T-cells typically recognize and attack foreign or "non-self" cells (such as cancer cells) in our body. However, some tumors escape this protection mechanism of our immune system. They inactivate the cytotoxic activity of T-cells via expression of PD-L1, the ligand for PD-1, thereby contributing to the T-cells' exhaustion. In that case, patients typically do not respond well to standard therapies or develop therapy resistance and metastases. However, armed with this particular knowledge of the tumor heterogeneity and stratification of patients, physicians could adjust the treatment such that some patients can benefit from anti-PD-1 and anti-PD-L1 therapy, also called checkpoint blockade, an immunotherapy approach that has been introduced successfully for many different tumor types.

In contrast to the large numbers of cells that can be assessed with mass cytometry, single-cell sequencing approaches are more limited due to significant costs, thereby restricting the analysis to a smaller number of cells derived from tumor biopsies. The assessment of a limited number of cells may therefore not represent the full picture needed to correctly diagnose and classify the tumor which is a prerequisite for finding the correct treatment. However, these aspects are addressed in multiple studies around the world and efforts to combine these studies are already underway.

Single-cell omics reaches the brain and its diseases

Despite the difficulty in obtaining brain biopsies, the first study applying single-cell RNA sequencing to postmortem biopsies of brains from Alzheimer's patients has been reported (Mathys et al., 2019). Alzheimer's disease (AD) is a progressive neurodegenerative disorder that accounts for the vast majority of age-related dementia in the world. Despite enormous research efforts, mainly in animal models, we still do not have a comprehensive understanding of AD, which is also reflected in the failure of clinical trials targeting molecules mainly identified and characterized in animal models. There is an urgent need to move into much more detailed analyses – preferably in humans – to be able to fight the disease. A recent study analyzing approximately 80,000 single-cell transcriptomes from 48 individuals with varying degrees of AD pathology (Mathys et al., 2019) identified myelination – a process that allows nerve impulses to travel faster – as a key factor in AD pathophysiology. Moreover, there is a sex-dependent molecular response in several cell types including oligodendrocytes, cells that produce myelin. This study is an excellent example of how single-cell omics is enhancing our understanding of major diseases which are relevant in our societies. Only single-cell resolution has allowed to unravel these novel pathophysiological mechanisms that can now be targeted by completely new therapeutic strategies. It is only the tip of the iceberg and we anticipate many more findings of this type relating to diseases of the brain and other major organs.

3.3 THE PRESENT OF SINGLE-CELL OMICS

Single-cell omics is fast-evolving into clinical research

Basic research is currently profiting greatly from the advances made by novel single-cell technologies. This new layer of resolution allows for an unprecedented view of heterogeneous cell populations, tissue composition and altered immune cell infiltration in disease. The recent emergence of techniques omitting sample disaggregation for integration of spatial information is pushing boundaries even further. The most prominent example of the current endeavors in the field is the Human Cell Atlas (HCA) – an enormous consortium effort committed to systematically mapping all cells of the human body at high resolution as a basis for understanding fundamental human biological processes and consequently use it as a reference resource to be able to gain insights into different pathologies (HCA, 2017; Regev et al., 2017). While the HCA is building the framework, large consortia are forming throughout the world, addressing the application of single-cell omics for disease-related questions. For example, within the European FET Flagship program, the LifeTime initiative (see also Junker, Popp, Rajewsky, Chapter 2) has been received as one of the most promising networks of experts eager to tackle future challenges of precision medicine by applying single-cell omics technologies.

Clinical trials utilizing single-cell omics are now within reach

Progress in the field has made it possible to move from proof-of-concept experiments to apply single-cell omics in broader settings. Clearly, in basic research, major goals are to understand differentiation processes during development, to gain insights into immune cell heterogeneity in classically defined cell populations, or studying pathogenesis in model systems. However, recent technical and computational advances have made it possible to move toward larger clinical studies, opening up new possibilities to study diseases, but also for the development of diagnostics, therapies and therapy management (see Junker, Popp, Rajewsky, Chapter 2).

As blood is the most accessible human tissue biopsy sample, isolated peripheral blood cells have long been the focus for studying disease or to serve as a surrogate for disease in other organs. Single-cell approaches have already helped us to learn about the cellular heterogeneity of expanded circulating immune cell populations in leukemia and have produced data that can be linked to clinical outcomes, for example to develop signatures for survival prediction (Gawad, 2014; Levine, 2015). Human solid tissue samples are harder to come by but are tremendously useful for gaining insights into cell composition and functional priming of cells present in diseased tissue.

3.4 THE FUTURE OF SINGLE-CELL OMICS

Scaling single-cell technologies to larger patient cohorts

The field has reached a point where single-cell omics can be applied to larger patient cohorts of clinically relevant diseases – an avenue that will elevate our understanding of disease to new heights. Unbiased comprehensive molecular profiling using single-cell omics technologies will enable us to define the molecular pathways and molecules involved in pathophysiological processes within every individual cell. This will be critical for the foundation of precision medicine, which requires a linking of molecular mechanisms with single-cell resolution to clinical phenotypes. We anticipate that this will trigger a change in traditional disease classifications. We will be able to much better stratify patients based on single-cell omics information. This will lead to much more precise identification of corresponding biomarkers including those for predicting and monitoring disease. Furthermore, it will be possible to comprehensively characterize novel and/or less defined rare diseases or clinical cases of uncertain diagnosis. Clearly, the knowledge gained will benefit approaches in precision medicine.

As it will become possible to make predictions relating to the reaction toward available medication, guided decisions on suitable therapeutic avenues will become available by screening individual patients. Another area which will be impacted is cell-based therapies, where single-cell analysis will aid in better characterizing and refining the utilized cell populations. For example, the purity of CAR (chimeric antigen receptor) T-cell therapy products could be improved using targeted single-cell omics analysis prior to administering a product to the patient.

Clinical application to immune-mediated diseases is most promising

Human immunology has been at the forefront of applications of single-cell omics technologies, as immune cells are easily obtained from peripheral blood (Bassler et al., 2019; Schultze/Aschenbrenner, 2019). Moreover, blood does not need any tissue disintegration as the cells are already in solution for further downstream analytical processing. Furthermore, immunologists can build upon a profound knowledge obtained by single-cell molecular profiling by flow cytometry. Thus, we expect diseases involving an immunological component, for example autoinflammatory conditions, chronic infections, metabolic syndrome, neurodegenerative diseases, or cancer, to be the frontrunners for being profiled.

Approaches from these types of systems applied to increasing numbers of individuals will also shed light on the functional variability between individuals with respect to complete organ systems, including the immune system. It has become clear that the combination of genetic susceptibility plus environmental factors influences an individual's well-being. For the immune system, it has recently been shown that immunosenescence, the aging of the immune system, is greatly affected by genetics and environmental influences and correlates better with disease

outcome and development than the actual chronological age (Alpert et al., 2019). The great potential of gaining insights into the variation of immune responses in healthy individuals has also been demonstrated by the Human Functional Genome Project (HFGP; Ter Horst et al., 2016; Li et al., 2016). We foresee that such large cohorts will build the backbone for future single-cell omics approaches to be applied to clinical questions. This is of particular relevance since these cohort studies already provide important information about the influence of environment and genetic susceptibility on immune functions disturbed during major inflammatory diseases (cancer, autoimmunity, chronic infections). Therefore, in addition to characterizing disease pathophysiology on the single-cell level, one of the tasks in the near future is to further delineate the factors influencing variation in the human immune system, as this will support the development of precision medicine approaches and disease risk prediction. For example, the frequency of a particular monocyte subset was highly predictive for therapy outcomes with anti-PD-1 immunotherapy, indicating how high-dimensional single-cell analysis predicts a response to checkpoint blockade (Krieg et al., 2018).

Multi-disciplinary teams will tackle clinical questions

The collaborative work of clinicians, biologists, bioinformaticians, computational biologists, Artificial Intelligence (AI) specialists, but also medical device developers and engineers will be particularly important to deduce medically relevant implications from the wealth of information produced by single-cell omics approaches. We envision single-cell approaches to derive disease-specific signatures that will be used for diagnostics instead of, for example, commonly used analytical parameters that are often not specific enough, such as white blood cell counts. As techniques and experience in the field are improving, costs will decrease over time, which will make these much more in-depth approaches clinically applicable. The gained knowledge may favor a switching to defined-marker tests, or even back to bulk sequencing at some point, making single-cell omics as a discovery tool even more attractive for the clinics.

Outlook – Future developments and requirements

Single-cell omics is an extremely fast-moving field and the following major aspects will drive this field over the next decade. In addition to "in-solution" single-cell omics, technologies preserving *spatial information* of the origin of individual cells will become major players within the clinical setting. This is not only due to the need to understand pathophysiological mechanisms in a spatial context, but also since the diagnostic framework of pathology is already spatial. Numerous different technologies are currently developed for spatial single-cell omics. This is an area of great potential, particularly in the context of clinical applications. More problematic will be the assessment of the temporal component of disease. Except in the case of blood, it is more difficult to envision repetitive biopsies derived from solid organs for single-cell omics analysis. Here, a strategy might be to develop blood-based single-cell omics as a surrogate. The second requirement will be the application of *AI methods* to the complete analytical pipelines of single-cell omics

data. These data are both big and sparse, which comes with particular challenges. Based on first successful applications of AI methods (Eraslan et al., 2019), we envision this field to be critical for clinical applications (see Aliee, Sacher, Theis, Chapter 4). To increase the predictive value of model systems, *human organoids* together with single-cell level analyses will further drive our understanding of human biology and major diseases (Roerink et al., 2018; Bolhaqueiro et al., 2019; Gehart et al., 2019; Klaus et al., 2019; Velasco et al., 2019; Xiang et al., 2019; see also Junker, Popp, Rajewsky, Chapter 2). The combination of single-cell level analysis, human organoids and AI will also drive the development for *better animal models*, which will still be necessary to determine causal relationships of molecular mechanisms responsible for major diseases. Probably the most important requirement is the development of sufficiently large structures – preferably international networks – that bring together all of this diverse expertise. It would be detrimental for any clinical development of single-cell omics if these sectors could not be efficiently linked to work together seamlessly.

3.5 REFERENCES

Alpert, A. et al. (2019): A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. In: Nat. Med. 25(3): 487-495.

Bassler, K. et al. (2019): The myeloid cell compartment-cell by cell. In: Annu. Rev. Immunol. 37: 269–239.

Bolhaqueiro, A. C. F. et al. (2019): Ongoing chromosomal instability and karyotype evolution in human colorectal cancer organoids. In: Nat. Genet. 51: 824–834.

Chevrier, S. et al. (2017): An immune atlas of clear cell renal cell carcinoma. In: Cell 169(4): 736–749.e18.

Eraslan, G. et al. (2019): Single-cell RNA-seq denoising using a deep count autoencoder. In: Nat. Commun. 10(1): 390.

Gawad, C. et al. (2014): Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. In: Proc. Natl. Acad. Sci. 111(50): 17947–17952.

Gehart, H. et al. (2019): Identification of enteroendocrine regulators by real-time single-cell differentiation mapping. In: Cell 176(5): 1158–1173.e16.

HCA (The Human Cell Atlas Consortium) (2017): The human cell atlas white paper, online publication 18.10.2017. Available at: https://www.humancellatlas.org/files/HCA_WhitePaper_18Oct2017.pdf [25.06.2019].

Klaus, J. et al. (2019): Altered neuronal migratory trajectories in human cerebral organoids derived from individuals with neuronal heterotopia. In: Nat. Med. 25(4): 561–568.

Krieg, C. et al. (2018): High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. In: Nat. Med. 24(4): 144–153.

Lavin, Y. et al. (2017): Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. In: Cell 169(4): 750–765.e17.

Levine, J. H. et al. (2015): Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. In: Cell 162(1): 184–197.

Li, Y. et al. (2016): A functional genomics approach to understand variation in cytokine production in humans. In: Cell 167(4): 1099–1110.e14.

Mathys, H. et al. (2019): Single-cell transcriptomic analysis of Alzheimer's disease. In: Nature, online publication 01.06.2019. DOI: 10.1038/s41586-019-1195-2.

Puram, S. V. et al. (2017): Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. In: Cell 171(7): 1611–1624.e24.

Regev, A. et al. (2017): The human cell atlas. In: eLife 6, online publication 05.12.2017. DOI: 10.7554/eLife.27041.

Roerink, S. F. et al. (2018): Intra-tumour diversification in colorectal cancer at the single-cell level. In: Nature 556(7702): 457–462.

Schultze, J. L./Aschenbrenner, A. C. (2019): Systems immunology allows a new view on human dendritic cells. In: Semin. Cell Dev. Biol. 86: 15–23.

Ter Horst, R. et al. (2016): Host and environmental factors influencing individual human cytokine responses. In: Cell 167(4): 1111–1124.e13.

Velasco, S. et al. (2019): Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. In: Nature, online publication 05.06.2019. DOI: 10.1038/s41586-019-1289-x.

Wagner, J. et al. (2019): A single-cell atlas of the tumor and immune ecosystem of human breast cancer. In: Cell 177(5): 1330–1345.e18.

Xiang, Y. et al. (2019): hESC-derived thalamic organoids form reciprocal projections when fused with cortical organoids. In: Cell Stem Cell 24(3): 487–497.e7.

Zilionis, R. et al. (2019): Single-cell transcriptomics of human and mouse lung cancers reveals conserved myeloid populations across individuals and species. In: Immunity 50(5): 1317–1334.e10.

Hananeh Aliee, Anna Sacher, Fabian J. Theis

4. DATA ANALYTICS IN SINGLE-CELL GENOMICS USING MACHINE LEARNING

4.1 INTRODUCTION

In both science and industry, datasets are growing at a faster rate than ever before because they are increasingly gathered by a great number of cheap devices and technologies resulting in so-called "Big Data". Big data is a synonym for large, complex, and often unstructured data, which therefore needs to be processed with statistical tools to reveal meaningful information. The analysis of such data is known as "data science", and it opens up new avenues in terms of combining data from various sources that helps to achieve deeper insights into a problem and make better decisions.

To extract value from data, methods of *Machine Learning* (ML), one of the main drivers of the Big Data revolution, are often used. ML is a subset of *Artificial Intelligence* (AI), which more generally aims to imitate human intelligence in particular tasks. More specifically, ML can be defined as computational algorithms applied to autonomously learn from both labeled and unlabeled data[1] to provide data-driven insights to guide decision-making and predictions. However, as the volume of data increases, conventional ML techniques may not be scalable so as to describe the complexity contained in the data. Hence, *Deep Learning* (DL) has emerged as a new area of ML. Deep learning is an ML technique based on artificial neural networks which concatenate simple nonlinear processing units[2] into multiple layers. DL architectures can capture complicated, hierarchical statistical patterns within data in *supervised* (e.g., for classification) and/or *unsupervised* (e.g., for clustering) modes.[3] The main advantage of DL algorithms is that they learn high-level features from data in an incremental manner. This eliminates the need of domain expertise for feature extraction, but commonly necessitates larger-scale, annotated datasets.

DL has revolutionized many fields such as computer vision and natural language processing in recent years, and has found applications ranging from astronomy to robotics, finance, healthcare, etc. In this chapter we focus on health research, in particular genomics, which itself has seen true exponential acceleration due to new advances in biomedical techniques from *next-generation sequencing* (NGS), which nowadays routinely creates a vast amount of genomic data. NGS-based

1 Unlabeled data consists of data without any information about the data, whereas labeled data contains additional information about the data (a label).

2 These units, called *artificial neurons*, loosely model the neurons in a biological brain. A connection can transmit a signal from one artificial neuron to another. The receiver neuron processes the input signal and signals to other artificial neurons connected to it.

3 "Supervised" indicates that, based on known datasets, functions are inferred that allow the classification of unknown data. "Unsupervised" means that unknown data are being examined and structures within the data are identified, which allows clustering.

technologies, like genomics, transcriptomics, proteomics, and epigenomics,[4] are now increasingly focused on profiling individual cells. Unlike traditional profiling methods that assess bulk populations, single-cell technologies isolate single cells, generate cell-specific sequencing libraries, and mark each cell individually with a cell-specific molecular barcode. Single-cell technologies then make it possible to profile the information of thousands to millions of single cells in a single experiment. This uncovers the heterogeneity among even similar cell types (see Aschenbrenner, Mass, Schultze, Chapter 3) and reveals potentially complex and rare cell populations, cellular dynamics, regulatory relationships between genes as well as developmental trajectories of distinct cell lineages (Hwang et al., 2018 ; see also Junker, Popp, Rajewsky, Chapter 2). However, the complexity of single-cell data coupled with the massive volume makes it a paradigm of Big Data. This makes it necessary to develop analytics capable of handling big datasets containing a large number of cells. As one of the most popular single-cell technologies with the largest scalability, this review will focus on single-cell transcriptomics and highlight its challenges and opportunities with a particular focus on modern analytics based on ML and DL.

4.2 MACHINE LEARNING IN SINGLE-CELL TRANSCRIPTOMICS

Single-cell RNA sequencing (scRNA-seq) entails the profiling of all messenger RNAs (mRNAs) presented in a single cell and provides the gene expression profile of hundreds of thousands and even millions of individual cells. Therefore, scRNA-seq represents truly Big Data with a superior statistical power that opens new horizons for applying Machine and Deep Learning for single-cell data analysis.

However, due to technical limitations and biological factors, the single-cell data generated is inherently sparse and noisy.[5] This gives rise to several computational and statistical challenges relating to recognition of patterns, like cell types, in gene expression. Commonly, additional quality control is performed to discard unreliable cells (e.g., outliers or possible doublets)[6] followed by *normalization*[7] which accounts for differences in read coverage and other technical confounders. Subsequently, *feature selection*[8] and *dimensionality reduction*[9] are performed, which filter the most informative genes and strongest signals from the background noise (Luecken/Theis, 2019).

4 Genomics entails the study of the entire genome, transcriptomics the study of all transcripts of genes (gene expression products, RNAs), proteomics the entirety of all proteins, and epigenomics all of the epigenetic data contained within cells. See Walter/Gasparoni, Chapter 1.

5 "Noise" in this context means that there are signals that are considered to be irrelevant or incidental to the question examined and that are obtained and need to be filtered in order to identify significant signals.

6 Outliers are cells that differ from the average expression level of their cell type, thereby making it harder to identify commonly expressed genes. Doublets are expression profiles that are accidentally generated from two cells instead of just one, often due to errors in cell sorting or capture. They can compromise the correct interpretation of results. For example, they can point towards the existence of intermediate populations or transitory states that do not actually exist.

7 Normalization typically scales count data to obtain correct relative gene expression abundances between cells.

8 Feature selection filters the dataset to keep only features/variables (for this concept, genes) that are informative of the variability in the data.

9 Dimensionality reduction is the process of reducing the number of random variables by obtaining a set of principal variables.

This is necessary for downstream analysis like clustering the cells to discover the sub-populations or inferring cell trajectories. In the following, we discuss some applications of supervised and unsupervised learning techniques in downstream analysis of transcriptome data.

Supervised learning in single-cell transcriptomics

Supervised learning is a domain of ML which requires training with labeled data to infer a function that can be used for mapping unlabeled data to output variables. A supervised learning model is first trained with a *training set* consisting of input-target pairs to learn the model parameters. In order to measure how well a function fits the training set, a loss function is defined for penalizing errors in prediction. The goal is then to optimize the model parameters by minimizing the prediction errors. The model is also validated with a distinct *validation set* followed by evaluating the performance of the inferred function using a *test set* that is separate from the training set. The accuracy of predictions is measured by different evaluation metrics such as the Pearson correlation coefficient. The main applications of supervised learning are classification and regression.[10]

In the field of single-cell transcriptomes, supervised learning is mainly employed for *cell annotation.* Cell annotation assigns cell types to unknown cells given a set of reference datasets with labeled cell types. For genomic data, this is the equivalent of using flow cytometry that is routinely employed for diagnosis of health disorders such as blood cancer. Conventionally, cells were annotated based on a set of markers which is labor intensive and requires extensive literature review of cluster-specific genes. Moreover, these genes often vary among different laboratories leading to difficulties in comparing their results (Pliner et al., 2019). Classical supervised learning techniques are therefore better as they automatically capture important features (or genes) from the labeled data enabling a more accurate cell annotation and reducing cross-laboratory classification discrepancies. In this regard, numerous classification models such as logistical regression, support vector machines, and random forests are used. However, with increasing data volumes, DL models might be preferred to the classical ML models in cell-type annotation tasks.

Unsupervised learning in single-cell transcriptomics

Unsupervised learning involves inferring useful structures or patterns from unlabeled datasets. Classically, unsupervised learning algorithms have been used for clustering data, dimensionality reduction, and visualization and embedding. Neural networks are able to generalize some of these approaches. For example, *Autoencoders* compress the data into a low-dimensional code and then decompress the code to reconstruct the original input data. An autoencoder allows to only approximately copy the input data into the output. This forces the model to engage in dimensionality reduction by learning how to ignore the noise. In the field of single-cell transcriptomes, autoencoders are employed for imputation/denoising as well as dimensionality reduction.

10 Regression is a set of statistical processes to estimate the relationship between the variables.

Embedding techniques like tSNE can then be performed on the latent space for mapping the compressed data onto a 2D plane. Specific noise characteristics of scRNA-seq data, such as *Zero-Inflated Negative Binomial* (ZINB), can also be addressed with customized loss functions within the autoencoder framework (Eraslan et al., 2019).

Another powerful application of unsupervised learning is *cluster analysis* to define cell types within scRNA-seq data. Broadly speaking, the goal of the cluster analysis is to cluster cells into groups based on the similarity of their gene expression profiles. Cluster analysis is the basis of several atlas projects, most notably the Human Cell Atlas (Human Cell Atlas[11]). These projects integrate several single-cell datasets into an atlas and build comprehensive reference maps of all human cells. For a cell atlas to be of practical use, reliable methods for unsupervised clustering of the cells will be one of the key computational challenges (Kiselev et al., 2019).

Cellular diversity may not sufficiently be described by a discrete classification system such as clustering. In fact, the biological processes that drive the development of the observed heterogeneity are continuous processes. Thus, in order to capture transitions between cell types, branching differentiation processes, or gradual, unsynchronized changes in biological function, we require dynamic models of gene expression. This class of methods is known as *trajectory inference*. In trajectory analysis (see Junker, Popp, Rajewksy, Chapter 2), the data is regarded as a snapshot of a dynamic process that lies on a connected manifold. The cells are then ordered along such a manifold and are described by a continuous variable called *pseudotime*. Pseudotime analysis – often based on transcriptional distance of cells from a root cell – describes development as a transition in transcriptomic state (i.e. trajectory) rather than a transition in real time. Pseudotemporal ordering of cells helps to understands how cell-type frequencies change in response to developmental and/or environmental signals that underlie physiological mechanisms of health and disease. For example, it determines how the frequency of a given cell type may decrease during a process because its death rate increases or because it differentiates to other cell types. It is important to understand the nature of this shift especially when the process is associated with a disease.[12] Another interesting question that pseudotime analysis can answer is how stem cells or progenitors are differentiated to develop an organ consisting of various cell types. In this regard, manifold learning approaches, categorized as nonlinear dimensionality reduction methods, are commonly used to learn the overall topology of the data and thereby infer the connectivity between the trajectories (Wolf et al., 2019).

4.3 OUTLOOK

Single-cell RNA sequencing is a powerful method for discovering intercellular heterogeneity. It focuses on the characterization of individual cells and can reveal complex and rare cell populations, uncover regulatory relationships between genes, and track the trajectories of distinct cell lineages in development. Several elegant studies have demonstrated the usefulness of scRNA-seq

11 See: https://www.humancellatlas.org/ [21.06.2019].

12 For example, decrease in pancreatic beta cell frequency being associated with diabetes.

namely to study the development of early embryos as well as to unravel the complexity of cancer and other diseases in ways that other techniques are unable to. However, the complexity of single-cell data coupled with its massive volume raise computational challenges in data analysis. Additionally, this is an emerging field for which standardized analysis methods are yet to be developed.

4.4 REFERENCES

Eraslan, G. et al. (2019): Single-cell RNA-seq denoising using a deep count autoencoder. In: Nat. Commun 10(1): 390.

Hwang, B. et al. (2018): Single-cell RNA sequencing technologies and bioinformatics pipelines. In: Exp. Mol. Med. 50(8): 96.

Kiselev, V. Y. et al. (2019): Challenges in unsupervised clustering of single-cell RNA-seq data. In: Nat. Rev. Genet 20(5): 273–282.

Luecken, M. D./Theis, F. J. (2019): Current best practices in single-cell RNA-seq analysis: a tutorial. In: Mol Syst Biol 15: e8746.

Pliner, H. A. et al. (2019): Supervised classification enables rapid annotation of cell atlases. In: bioRxiv, online publication 04.02.2019. DOI: 10.1101/538652.

Wolf, F. A. et al. (2019): PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. In: Genome Biol 20(1): 59.

Bernd Müller-Röber

5. SINGLE-CELL TRANSCRIPTOME ANALYSIS IN PLANTS

5.1 PLANT SINGLE-CELL TRANSCRIPTOMES

In plant research, single-cell transcriptome analysis (single-cell RNA-seq, scRNA-seq, see Walter and Gasparoni, Chapter 1) is only now being established, although it is quickly gaining in importance. Unlike animal cells, plant cells have a rigid cell wall, which consists of different carbohydrate polymers with a variable composition, depending on the cell type and state of differentiation. These include cellulose, hemicellulose and other components, among them proteins integrated into the cell wall. The cells form a stable tissue and must be first separated from each other prior to a typical single-cell transcriptome analysis. This is achieved by treating the plant tissue with different enzymes, which dismantle the cell wall; during this process, so-called "protoplasts" are created (i.e. plant cells without a cell wall), which are then subjected to a single-cell transcriptome analysis. Since the production of the protoplasts itself can already lead to a change in the transcription pattern, corresponding checks must be made, such as comparisons with already known gene expression patterns of untreated plant cells, in order to find out whether these patterns can also be found among the protoplasts.

The works published on single-cell transcriptome analysis in plants to date have focussed on roots of the plant *arabidopsis thaliana* (thale cress; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Turco et al., 2019). They are a long-studied and now well understood model system of developmental processes in plants. Numerous genes that control the development of plant roots, and their reaction to environmental influences, are known. This research has also led to the identification of marker genes that are only active in certain cell types of the root, for example, in stem cells or hair-forming cells of the root epidermis. The expression data obtained via single-cell transcriptome analysis can therefore be compared with gene activity maps of the roots obtained earlier, and thus validated for genes, the expression of which was already known in different cells. In the studies published to date, the transcriptomes of around 400 to 12,000 individual cells were analysed in each case.

To date, single-cell transcriptome analysis in plants has provided the following new information:

- scRNA-seq captures spatio-temporal information for high-precision gene expression,
- scRNA-seq allows the identification of new regulators for processes in individual cell types,
- by means of scRNA-seq, regulatory paths of cellular development can be studied to a higher degree of precision than has been possible to date using other methods,
- and it has been possible to identify subtypes of cells that have been unknown to date on the basis of their specific gene expression pattern.

In the scRNA-seq studies conducted to date, mutants and transgenic plants with defects in root development were also studied alongside wild-type plants (Ryo et al., 2019; Turco et al., 2019), and the effects of external factors such as the availability of sucrose in the growth medium (Shulse et al., 2019) or heat stress (Jean-Baptiste et al., 2019) were analysed. As sessile organisms, plants cannot escape through mobility from the environmental conditions to which they are exposed. However, they do have a pronounced development plasticity, i.e. in a number of different ways, they can adapt to different environmental situations through changes in their growth and morphology, without changing their genetic constitution (sequence of genetic information) (Bradshaw, 2006; Salazar-Henao et al., 2016). These different adaptation scenarios and the genes that form their basis have become established during the course of evolution and have led to the adaptation of plants to certain ecological niches. In the future, single-cell transcriptome analysis will make it possible to analyse the molecular and cellular mechanisms on which these complex – and variable – developmental processes are based in considerably greater detail than has been the case to date.

5.2 TRANSCRIPTOME ANALYSES USING ISOLATED CELL NUCLEI

As explained above, to date, single-cell transcriptome analysis in plants has still required the protoplasting of plant tissues. Since the composition of the cell wall varies between different cells in a plant, and cells of different plants, and in addition is also modulated by environmental influences, suitable protoplasting protocols must first be developed in each case. This alone can already take up a great deal of time, which makes access to single-cell transcriptomes more difficult. As a possible alternative one can analyse transcripts present in the nuclei of plant cells and their tissues, rather than taking protoplasts as the subject of investigation. Employing cell nuclei has the advantage that no protoplasting protocols need to be established for every plant and tissue type. In addition, well established protocols for a fast and uncomplicated enrichment of cell nuclei from complex plant tissues or organs are already available. With the aid of such methods, it could become possible in the near future to analyse the single-cell transcriptomes of plants that have not to date been included in the typical model systems, but which are of particular relevance from an ecological perspective or in respect of their specific physiological or biotechnological properties. For example, it is possible in principle to isolate cell nuclei "en bloc", as it were, from an organ (with its different cell types), in order to then subject them to further analyses. With this method, it is possible to forego the use of transgenic plants entirely. However, the option is also available to use isolated cell nuclei in a targeted manner from specific cell types. To do so, nuclei of the corresponding cell types must be labelled. This can be achieved, for example, by equipping them with certain proteins, such as the green fluorescent protein (GFP); to this end, the plants are genetically modified with suitable gene constructs. Cell nuclei that are labelled in this way can then be isolated through suitable biochemical methods, such as immune precipitation using antibodies that detect GFP, or through FACS (fluorescence-activated cell sorting), and then subjected to a transcriptome analysis. However, here, the gene expression patterns in cells that have not been specifically marked are not captured, as a result of which the experimenter may fail to obtain important information for a more comprehensive interpretation of cellular processes in tissues.

5.3 FUTURE RESEARCH

Future scientific questions that can be assessed with the aid of single-cell transcriptome analysis in plant research are extremely varied in nature, and can currently only be foreseen to a limited degree. Some examples of these, among many others, are:

1. the analysis of gene expression patterns in plants changing growth and development due to environmental influences,

2. the decoding of gene regulatory networks with a high degree of spatio-temporal precision from which bioinformatics will benefit in particular,

3. through the comparison of single-cell gene expression patterns in different plants, it will become possible to better understand the evolution of plants and their diversification and adaptation to different ecological niches at molecular level,

4. synthetic-biological approaches are also increasingly gaining in importance in plants. There is no doubt that here, single-cell transcriptome analyses will make a significant contribution towards understanding the variability between cells. As a result, it will become possible to create a more solid basis for the robustness of synthetic-biological modifications in plants.

5.4 SUMMARY

In the field of plant research, single-cell transcriptome analyses have to date only been reported in scientific publications in connection with the model plant *arabidopsis thaliana*, and here, with a sole focus on roots. Thus, there are currently still no comprehensive data records available for ongoing analysis by the scientific community. However, despite the still very small number of publications in this field of research, it can be assumed that single-cell transcriptome analysis will have a considerable impact on plant research in the future. This also relates particularly to research in cultivated plants that are important for feeding humans and animals. Here, it is of particular interest that not only "traditional" cultivated plants can be included in the analyses, but also those that have tended to be underrepresented to date, and which will require further genetic optimization in the coming years. Sufficient research funds should be made available for this purpose.

5.5 REFERENCES

Bradshaw, A. D. (2006): Unravelling phenotypic plasticity – why should we bother? In: New Phytol 170(4): 644–648.

Denyer, T. et al. (2019): Spatiotemporal developmental trajectories in the arabidopsis root revealed using high-throughput single-cell RNA sequencing. In: Dev Cell 48(6): 840–852.e5.

Jean-Baptiste, K. et al. (2019): Dynamics of gene expression in single root cells of arabidopsis thaliana. In: Plant Cell 31(5): 993–1011.

Ryu, K. H. et al. (2019): Single-cell RNA sequencing resolves molecular relationships among individual plant cells. In: Plant Physiol 179(4): 1444–1456.

Salazar-Henao, J. E. et al. (2016): The regulation and plasticity of root hair patterning and morphogenesis. In: Development 143(11): 1848–1858.

Shulse, C. N. et al. (2019): High-throughput single-cell transcriptome profiling of plant cell types. In: Cell Rep 27(7): 2241–2247.e4.

Turco, G. M. et al. (2019): Molecular mechanisms driving switch behavior in xylem cell differentiation. In: Cell Report 28 (2): 342–351.e4.

Heiner Fangerau, Lilian Marx-Stölting, Angela Osterheider

6. SINGLE-CELL ANALYSES AND ETHICAL CONSIDERATIONS

6.1 INTRODUCTION

From an epistemological viewpoint, the biomedical model rests on the pillars of "universalism", "reductionism" and "modeling" (Strasser, 2014). As the smallest or "most basic functional unit" (Regev et al., 2017),[1] which regulates its own genetic expression, the cell has long played a prominent role as the anchor point in this triad for each of these approaches. For a long time, the dogma applied that: a) what was valid for one cell was valid universally; b) the analysis of living beings could be reduced to the cellular level; and c) one cell type could in turn be drawn on as a model for another in the research process.

The single-cell analysis field of research draws on exactly this point, on the one hand, while on the other hand differentiating the specified pillars further by departing from the principle that individual cells are far more different from each other than was previously assumed. The special features of individual cells with regard to their genome, epigenome, transcriptome or proteome[2] have increasingly come into focus in research efforts over recent years.

Although it is still valid to assert, as Michael Speicher stated in 2013, that single-cell analysis as a field of research is in its infancy, the potential which he ascribed to the approach at the time has lost none of its power. Thus, the corresponding methods serve, among other things, for a better understanding of how various cells differentiate, age or react to contaminants and how different cell types can be characterized.[3] In the medical application, in turn, this knowledge can facilitate a better understanding of tumor formation or metastasizing cells. Also in preimplantation diagnosis and other forms of disease prediction, a gain in momentum is expected (Speicher, 2013). Overall, the hope is that single-cell analyses should contribute to the realization of lab-centered, predictive and so-called "personalized" or "individualized" medicine. They could do this, for example, by helping to reliably predict when a disease will take hold and how it will progress, or by making it possible to test treatments on a cellular level (Shalek/Benson, 2017).

1 A good overview of the various single-cell genomics methods is also provided here (p. 4).

2 A cell's genome is the sum of all genes; the epigenome is all epigenetic modifications; the transcriptome refers to the sum of all genes transcribed (rewritten from DNA to RNA) in a cell; and the proteome consists of the total number of all proteins contained in the cell (see also Walter/Gasparoni, Chapter 1). The term "omics technologies" refers to the processes that enable the capture of data on these characteristics.

3 Various international initiatives are being carried out for this purpose, such as the "Human Cell Atlas" project for the characterization of the various human cell types, or the "LifeTime Initiative", in which large volumes of data are being gathered and evaluated. Regarding the Human Cell Atlas, see Regev et al., 2017. Regarding LifeTime see: https://www.mdc-berlin.de/de/lifetime [14.05.2019]. See also Junker, Popp, Rajewsky, Chapter 2.

The ongoing explicit reference to potential areas of application of single-cell analysis in medicine begs the question: to what extent might this field of work evoke its own or new medical ethics questions? In any case, this process, still in its infancy, is an area that has hopes of having a significant medical impact and thus an area in which patients, doctors and other healthcare stakeholders can come into direct contact both personally and professionally, and in their dealings with one another. Furthermore, the question arises as to whether the linking of single-cell analysis methods with complexes such as genome editing (for model formation) merely reproduces the ethical dilemmas debated here or whether this gives them a new face.

6.2 ETHICAL TOPICS

To come to the point: It does not seem as if an entirely new sub-field of moral theory needs to be established in order to ethically reflect single-cell analysis. On the contrary, the methods touch on a range of ethical questions which have already long been discussed in bioethics within the scope of other biotechnologies. At the heart of these are value conflicts which have been discussed repeatedly within the field of medicine in connection with new (mass) data gathering and processing methods. The interest in data protection and data sovereignty or the concerns around data misuse is opposed to the interest of the most highly-comprehensive data gathering possible aimed at the acquisition of knowledge. Indeed, the fact that single-cell analysis can be linked, used and connected across disciplines to the most diverse research and application areas nevertheless makes a compilation of the ethical topics on which it touches extensive and, in cases of doubt, incomplete. The following is an endeavor on our part to tease out the ethical challenges that are specific to single-cell analysis, with the possible medical value and freedom of research standing on one side of the debate, and (mostly social or individual) values which have the potential of coming into conflict with these on the other.[4]

Significance and validity of data

The implementation of single-cell analyses opens up the possibility of gathering and interpreting large volumes of data for the first time. Like in other fields of medicine, there are hopes that large volumes of data may also lead to better diagnoses. Similarly to other areas of bioinformatics, medical diagnosis and data-driven research, it must be ensured that statistical standards are upheld, correlation problems are noted and data is thus interpreted correctly and uniformly. Otherwise, there is a risk of over- or underestimation of correlations, misinterpretations and, in case of doubt, bad medicine working on the basis of inaccurate data.[5] Single-cell analysis must bear in mind, epistemologically speaking, that the investigated cells are being observed as removed from their cell cluster, their system, which can impact on their behavior. Similarly, in relation to possible data gathering, it is necessary to bear in mind, from an overall self-critical

4 For further information on the following overall topics, see: Lenk et al., 2014 and Düwell, 2011; an overview of “Big Data” in: German Ethics Council, 2018.

5 The error rate of large volumes of data, see for instance Bertram, 2019.

perspective, that an individual's income situation may be a better risk indicator for certain disease risks than omics data.

Data protection

As always, the gathering and processing of biological data for research raises questions of data protection and data ownership. It is necessary to clarify whether the potentially obtained human cell data belongs to the researchers or to the test subject or patient. Furthermore, data protection and appropriate data management are essential, as for example anonymization of the data may be difficult if a link with individuals is required for subsequent research activities. This is also associated with the practical problem of data exchange between researchers who might even be located across borders, as is applicable in the case of international research projects with partners whose data protection provisions do not necessarily correspond to those in the European Union. Since data protection is not a purpose in and of itself, but rather it should serve to protect people from stigmatization or the misuse of knowledge about them, it is above all necessary to define the domains in which single-cell analysis is able to generate corresponding data in the first place.

Informed consent

Data protection questions may be addressed through appropriate informed consent if those persons affected agree to participate in research and to the storage or saving and use of their biomaterial for data capture following a specific and understandable explanation. There is some dispute in relation to the scope of agreement, which extends from targeted, specific ("narrow consent") to all-encompassing, open-ended ("broad consent"). As there is also the problem of the understandable clarification on unknown usage possibilities, which may always occur in view of the openness of research, the question arises as to how far consent can go in this respect, and to what extent approvals can be kept dynamic if researchers want to avoid constantly having to obtain new or updated consent ("dynamic consent").

Additional and incidental findings

Closely linked to the informed consent issue is the question of the handling of findings and ancillary findings which are gathered outside of an initially clearly delineated diagnosis. Dilemmas particularly occur here, if relevant information on treatable or non-treatable future diseases is obtained and the previous informed consent excluded the question of notification on such findings or if the corresponding decision of the person affected contradicts the researcher's moral intuition (e.g., in cases of refusal to notify of a treatable, untreated or terminal illness).

Social implications/equitable distribution

As is the case for any new medical or biotechnology, above all questions of justice and prioritization arise, with regard to society. These questions range from research funds being allocated on a higher level, right down to the level of individual patient care. If the costs for access to the gathered data from single-cell analyses become very high, these may be provided for only a limited number of patients or self-paying patients for treatment planning. Stratification into patient subgroups could lead to relatively small groups requiring relatively expensive medications and/or specific groups being excluded from care due to excessively high costs. Patients with rare illnesses are especially at risk in this regard.

6.3 CONCLUSION

None of the groups of ethical topics outlined here is new or specific to single-cell analysis. Strictly speaking, moral decision-making situations brought about by this method alone seem to be relatively rare. Nevertheless, it presents extensive common ground for problems from other areas and suggestions for solutions. In conclusion, the current hype surrounding single-cell analysis as *Science* journal-nominated "breakthrough of the year" in 2018[6] must not be allowed to bring about a reduction or negation of ethical standards that are already established in other fields. As always, it is a clear requirement that the standards of good scientific, clinical and ethical practice be upheld.

6.4 REFERENCES

Bertram, I. (2019): Big data, big errors – Interview with Gerd Antes. In: GID 248, February 2019. Available at: https://www.gen-ethisches-netzwerk.de/biobanken-und-big-data/gentests-und-genomsequenzierung/big-data-big-errors [14.05.2019].

German Ethics Council (2018): Big data und Gesundheit. Datensouveränität als informationelle Freiheitsgestaltung: Stellungnahme [Big data and health. Data sovereignty as the shaping of informational freedom: opinion]. German Ethics Council, Berlin.

Düwell, M. et al. (ed.) (2011): Handbuch Ethik [Manual of ethics]. 3rd updated edition. Metzler, Stuttgart, Weimar.

Lenk, C. et al. (ed.) (2014): Handbuch Ethik und Recht der Forschung am Menschen. [Human experimentation: ethical and legal manual]. Springer, Berlin, Heidelberg.

Regev, A. et al. (2017): The Human Cell Atlas. In: eLife 6, online publication 05.12.2017. DOI: 10.7554/eLife.27041.

Shalek, A. K./Benson, M. (2017): Single-cell analysis to tailor treatments. In: Sci Transl Med, online publication 20.09.2017. DOI: 10.1126/scitranslmed.aan4730.

Speicher, M. R. (2013): Single-cell analysis. Toward the clinic. In: Genome Medicine 5(8), online publication 27.08.2013. DOI: 10.1186/gm478.

Strasser, B. J. (2014): Biomedicine. Meanings, assumptions, and possible futures. Report to the Swiss Science and Innovation Council (SSIC) 1/2014. Swiss Science and Innovation Council, Bern.

6 See: https://www.sciencemag.org/tags/2018-breakthrough-year [14.05.2019].

Angela Osterheider, Yaroslav Koshelev, Marlen Reinschke, Lilian Marx-Stölting

7. PROBLEM AREAS AND INDICATORS IN THE FIELD OF SINGLE-CELL ANALYSIS

7.1 INTRODUCTION: MOTIVATION AND PURPOSE

The Interdisciplinary Research Group (IAG) *Gene Technology Report* at the Berlin-Brandenburg Academy of Sciences and Humanities is tasked with observing the various developments in the field of gene technology in Germany over the long term, and with making them available to interested members of the public in form of publications and events. The results it publishes are intended to provide a source of generally accessible information and thus promote well-informed discussion in the public domain on subject areas which are dynamic in nature and in some cases contentious in society. Alongside the qualitative analysis of various aspects of gene technologies (e.g., the natural sciences, law and ethics), the IAG has undertaken to open up the complex field of gene technologies to interested members of the public and present it in a (publicly) accessible and measurable form (Diekämper/Hümpel, 2015: 16 ff., 2012: 51–60). The problem area and indicator analysis method, which originates from the social sciences, is used as a core instrument in the process. Taking qualitative data gathering (problem area analysis) as a starting point, quantitative data (indicators) are collated.[1]

7.2 PROBLEM AREAS

Many gene technology topics are intensely debated in the public domain and particularly in the media. The IAG *Gene Technology Report* applies the method of problem area capture in order to break these complex discussions down into subject areas and aspects (problem areas). Thus, the problem area analysis is aimed at presenting the public perception of gene technologies in a clear way (Diekämper/Hümpel, 2015: 16). Various print and online media are evaluated within the scope of the analysis. Following this evaluation, the identified problem areas are allocated within a selected coordinate system. This coordinate system is founded on the cornerstones of four guiding dimensions that primarily stand out in the context of gene technologies. These four dimensions are of an economic, scientific, ethical and social nature. In a final step, these problem areas are assigned to relevant indicators.

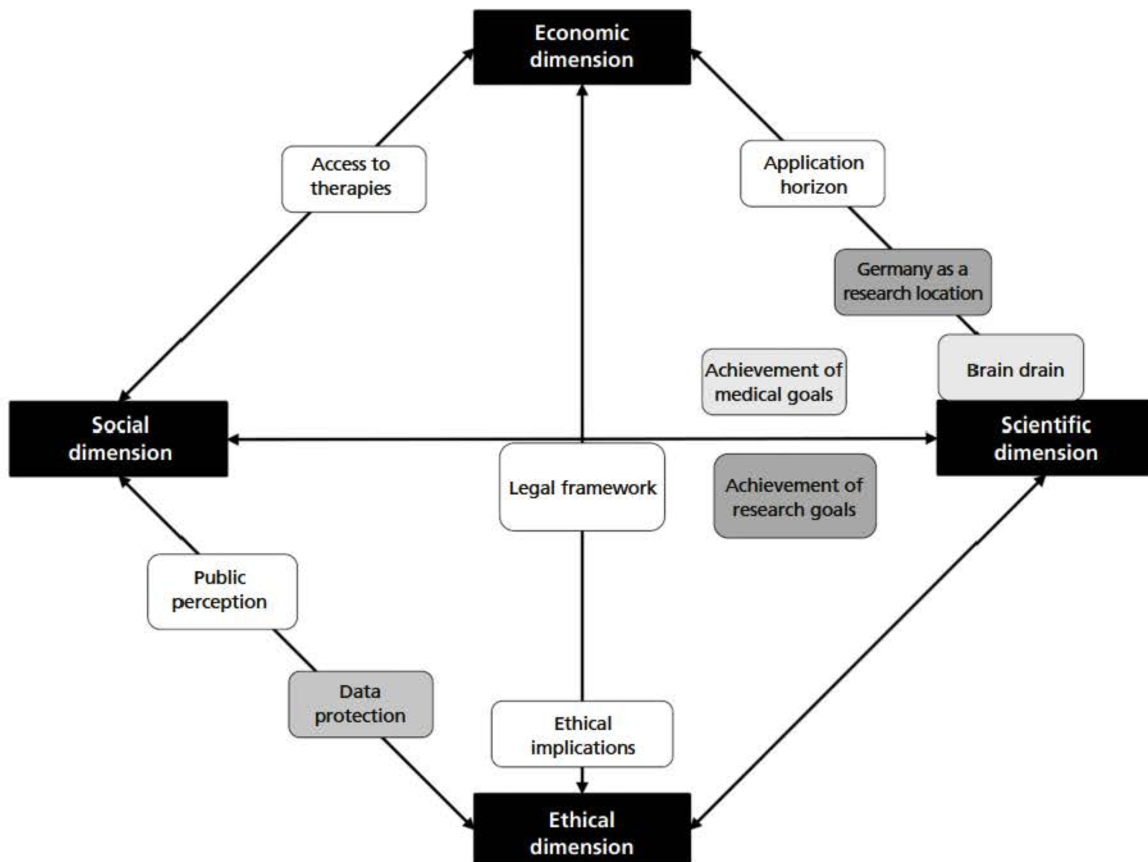
The problem area is captured on the basis of a qualitatively evaluated text corpus. This text corpus is gathered using keyword research in main print media, the daily newspapers *Süddeutsche*

1 Problem area and indicator analysis are one of the IAG's main methods. Thus, introductory and general considerations as well as statements regarding this approach have already been put forward in previous IAG publications (see e.g.: Marx-Stölting, 2017; Diekämper/Hümpel, 2012).

Zeitung and *Frankfurter Allgemeine Zeitung* and the weekly magazines *Die Zeit* and *Der Spiegel* on the one hand, and in online search engines (Google and Metager) on the other.[2]

Figure 1 shows the identified problem areas for the subject area of single-cell analysis and its quantitative weighting within the analyzed text corpus. The size and coloration show the quantitative weighting of the problem areas. The more frequently the problem area was discussed in the text corpus, the larger the size and the darker the color in which it is shown.

Figure 1: Problem areas of single-cell analysis



2 For the print media, a search was carried out using the following German keywords for the period from 5 to 12 March 2019: "Einzelzellsequenzierung", "Einzelzellbiologie", "Einzelzell-Transkriptomik", "Einzelzell-Genomik", "Einzelzelldiagnostik" and "Einzelzellanalyse" ["single-cell sequencing", "single-cell biology", "single-cell transcriptomics", "single-cell genomics", "single-cell diagnostics" and "single-cell analysis"]. Within the scope of the research for other subject areas of the IAG, as a rule only German keywords were used. Since only very few articles could be found, the following English terms were also used for this topic: "single-cell analysis", "single-cell biology", "single-cell sequencing", "single-cell genomics", "single-cell diagnostics", "single-cell transcriptomics". Four articles were found in total. The research using the Google and Metager search engines was carried out from 12 to 26 March 2019. The aforementioned German search terms were used again as well as the combination of these search terms (with truncation) and the word "Stellungnahme" [opinion]. The first ten hits from the search engines were merged and compared. The hits from the search engine and print media research jointly form the text corpus, which was then qualitatively evaluated with regard to the problem areas.

The following problem areas were determined using the qualitative evaluation of the text corpus:

Application horizons: Application horizons in the area of single-cell analysis are already subject to ongoing discussion. They comprise visionary goals with high innovation potential, whose feasibility is accordingly unknown.

Germany as a research location: A variety of factors contribute to the international attractiveness of a research location: the existing scientific infrastructure and the scale and type of funding measures, but also national legal provisions which influence scientific practice. The international reputation and networking within the globalized research landscape also play a role.

Access to therapies: If the costs for potential medical applications or for access to the data gathered from single-cell analysis are very high, the question as to the coverage of costs by the statutory health insurance bodies arises. At this point, questions around prioritization and distribution must be discussed.

Brain drain: In a dynamic, globalized research landscape with its demand for mobility, Germany is at risk of losing scientific talent without attracting scientists to the same extent in return. For the area of single-cell analysis, this can mean that highly-qualified scientists leave the country for professional, scientific or legal reasons. Thus, in the global research race and competition for location dominance, important know-how can be lost and economic potential can go untapped.

Achievement of medical goals: One of the aims of the research in the area of single-cell analysis is to acquire new findings in order to promote developments in the area of personalized medicine, among others. Problems occur if not all targets are achievable, or if targets turn out to be more difficult or time-consuming than was initially assumed.

Achievement of research goals: Scientific research strives to generate new findings and technologies. Limited planning ability and openness to unforeseen results is an inherent aspect of its nature. Nevertheless, the existing framework conditions, such as the scientific infrastructure, funding options or applicable law, influence the achievement of defined research goals – goals which are reflected in a quantifiable way in publications, research awards or academic statements, for instance.

Legal framework: The legal frameworks at national, European and international level determine the permissibility of research, particularly the handling of research data. The legal frameworks define the application in scientific practice or formulate the necessary framework conditions. They play a role in the assessment of other conflicting interests and protected assets. The data exchange beyond group boundaries within the scope of international projects is just one of the points under discussion within the area of single-cell analyses.

Ethical implications: Research – especially in the life sciences and more intensively in biomedical research – generates knowledge and applications which demand an analysis of potential

consequences for the individual, society as a whole and the environment. Social or legal aspects play a role here as much as ethical considerations. These must be discussed in the public domain and may ultimately require that political action be taken. In the case of single-cell analysis for example, the handling of volumes of diverse data and the associated value conflicts pose significant potential for discussion. The question regarding the handling of incidental and additional findings additionally plays a role in this area.

Data protection: The gathering and saving of research data in the area of single-cell analysis in principle enables a more extensive use which may affect individual rights. The right to informational self-determination as well as a "right not to know" are under discussion in this context.

Public perception: How new technological processes are publicly perceived is of essential significance in their use and establishment. Discussion of single-cell analysis in the print media and online as well as the number of public events and publications that are accessible to the public illustrate the interest in the topic within the public domain.

7.3 INDICATORS

Based on the qualitative problem area analysis, indicators[3] (quantitative data) are collated in order to illustrate current developments. A selection of the indicators is presented and evaluated below. Using this data, initial indications can be provided on the current status as well as developments in the area of single-call analysis.

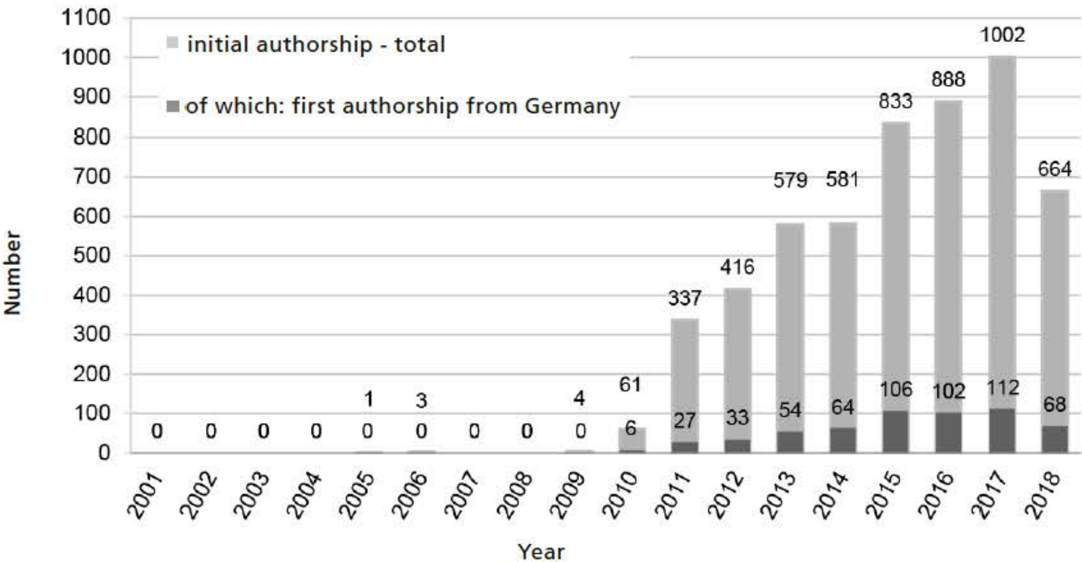
The indicators "number of international publications", "online search queries" and "new publications" currently appear suitable for illuminating selected problem areas in the field of single-cell analysis – especially in light of the fact that this is in an area in which there are many new developments at present.

The indicator "number of international publications" on single-cell analysis was allocated to the problem areas "achievement of research goals" and "Germany as a research location". PubMed, the free and publicly accessible online citation database of the American National Center for Biotechnology Information (NCBI), was used during the search (accessed: March 2019, status: 2018). The database claims to currently hold approx. 24 million citations for biomedical literature from MEDLINE (= Medical Literature Analysis and Retrieval System Online), relevant specialist journals and e-books. In general, specialist articles starting from 1946 are taken into account, and in some cases also older ones. The focus is on English-language literature. Research can be conducted using freely selected keywords on the one hand, or, on the other, the Medical Subject Headings (MeSH) catalog, which is used for indexing the PubMed citations and is continually maintained and expanded by the American National Library of Medicine (NLM) (see: www.nlm.

3 The indicators were provided in previous publications of the IAG using standardized indicator sheets. They were recently published in the fourth genetic technology report (Marx-Stölting et al., 2018: 299–340).

nih.gov/mesh [03.04.2019]).[4] A relevant MeSH from the current MeSH catalog was used for the "single-cell analysis" research. In addition, first authorships from Germany were also identified. The data presented here starts from 2001 (the year in which the IAG *Gene Technology Report* started its work) to 2018. The indicator reflects the global research activities in the area of single-cell analysis. Based on the scope of publications published to date, it is possible to observe how intensively a subject area is being researched over the years and which countries occupy a prominent position in the "international research race" within that area. However, it is important to note that, despite the large scope of the database, a comprehensive collection of citations cannot be expected: relevant publications may not be in the database in the first place or may not be indexed by keyword under the MeSH categories used. It must also be taken into account that even though a publication may represent an equal collaboration of authors from several countries, the MEDLINE database only collects the nationality of first authors as standard practice. The representation for 2018 may be incomplete, since it may be the case that not all publications are as yet included in the database.

Figure 2: Number of international publications on single-cell analysis (total and with German first authorship [2001–2018])



Year	2001	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Total	0	0	0	0	1	3	0	0	4	61	337	416	579	581	833	888	1002	664
German first authorship	0	0	0	0	0	0	0	0	0	6	27	33	54	64	106	102	112	68

4 A piece of research was also carried out in the "Web of Science" data tape (accessed: 05.04.2019). The development of the publication figures is similar to that from "PubMed". There is a steady increase from the start of the gathering period. However, the data is only similar to a limited extent, as no MeSH terms are used in the "Web of Science".

The indicator "online search queries" was gathered using the free Google Trends tool (see: https://www.google.com/trends [26.03.2019], status: March 2019). This indicator was assigned to the problem area "public perception". This online tool analyses a percentage of the search inputs into the Google search. However, the analysis algorithm used and the absolute figures on the search queries are not visible to the public. The data reflects the demand for a certain search term in relation to the overall search volume in Google within a selected period of time. A large proportion of the population in Germany uses the internet for personal purposes on a regular basis (87 % in 2018; see: https://www.destatis.de [03.04.2019]). The research was mainly carried out using internet search engines: Top of the list is Google (see: http://de.statista.com [03.04.2019]). Thus, online search queries can be regarded as an indicator for the interest of the public in various topics.

German search terms have usually been used in the previous publications of the IAG *Gene Technology Report*. As a first step, the following German keywords were researched: "Einzelzell-sequenzierung", "Einzelzellbiologie", "Einzelzell-Transkriptomik", "Einzelzell-Genomik", "Einzelzellanalyse", "Einzelzelldiagnostik" [translation: "single-cell sequencing", "single-cell biology", "single-cell transcriptomics", "single-cell genomics", "single-cell analysis", "single-cell diagnosis"] (truncations such as "single-cell*" are not possible in Google Trends). However, there was insufficient data for these search terms ("Search volume is too low" = 0). This shows that the subject area is still very young and not being widely discussed in the public domain. In a subsequent step, the following English terms were queried: "single-cell analysis", "single-cell biology", "single-cell sequencing", "single-cell genomics", "single-cell diagnostics" and "single-cell transcriptomics". The only English keywords that received hits were: "single-cell analysis" and "single-cell sequencing". Search results may be filtered under regions (countries, cities) and defined search categories.[5] In addition, it is possible to search for several keywords at the same time. The data is visible to the public effective from 2004. Thus, a search was carried out for Germany in the period from January 2004 to March 2019 (gathering date: 26.03.2019, status: March 2019).

The relative demand for the keyword "single-cell analysis" peaks in the years 2004, 2005 and 2006. From then until now, the demand has been at a lower level. The keyword "single-cell sequencing" is only searched for more often starting from 2010, while demand dropped to a medium level in the subsequent years, peaking in 2013 and 2014. From 2017 to 2019, demand increases again to approximately the level of 2010.

5 To narrow down the significance: A high search volume cannot be equated with an increase in search queries, the calculations are based on random samples, multiple meanings of the search terms may play a role and the reason for the search for information cannot be traced.

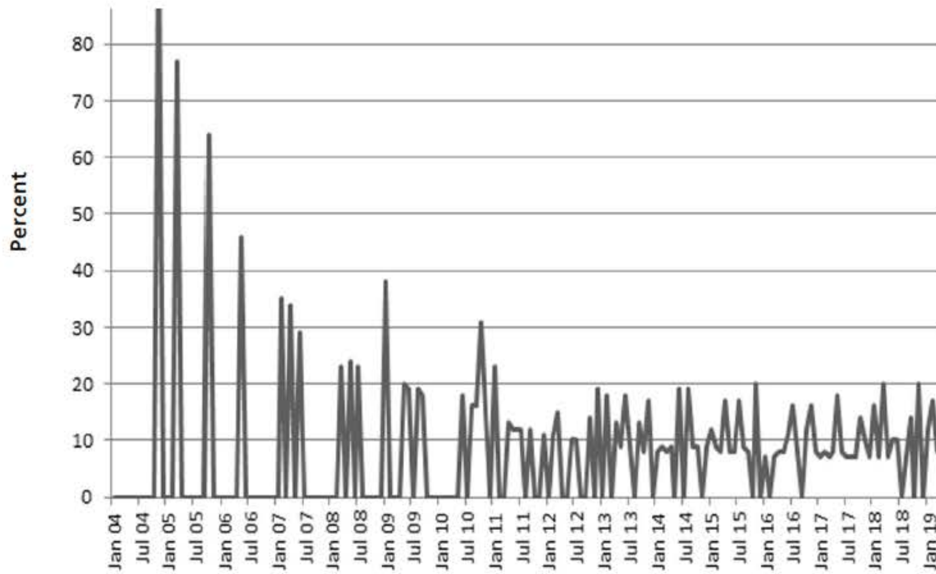


Figure 3: Relative demand for the keyword "single-cell analysis" in Google Trends for Germany (2004–2019)

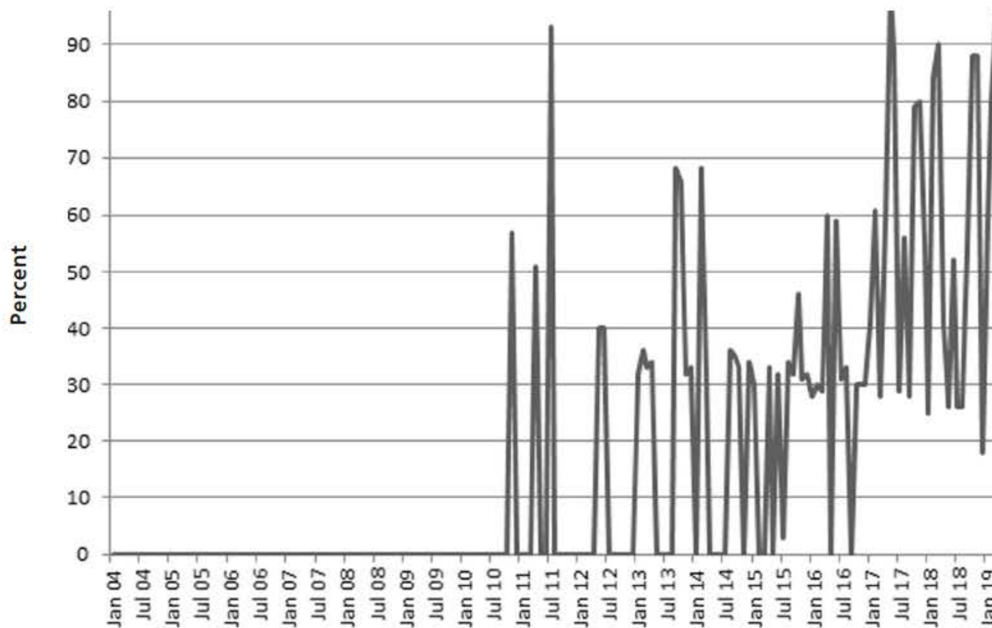


Figure 4: Relative demand for the keyword "single-cell sequencing" in Google Trends for Germany (2004–2019)

The indicator "new publications" shows the publication density of books in Germany. This indicator was assigned to the problem areas of "public perception" and "Germany as a research location". A keyword research was carried out in the database of DNB (Deutsche Nationalbibliothek [German National Library]) to gather this indicator (date of gathering: 26.03.2019, status: March 2019). DNB is a public law institution directly accountable to the federal government. Its task is the archiving and bibliographic collection of publications (monographs, newspapers, journals, loose-leaf binders, cards, sheet music, sound recordings, electronical publications) published in

Germany. This additionally involves the collection of German-language publications published outside Germany, translations of German-language publications published outside Germany, foreign-language publications about Germany and exile publications of German-speaking emigrants between 1933 and 1950. Online publications have also been included in a systematic way since 2006. DNB has permitted research in its comprehensive library stocks free of charge since 1913. Based on information from the provider, publications received are added to the catalog and the DNB following a processing period of approx. one month. Relevant titles were queried using the following German search terms: "Einzelzellsequenz*", "Einzelzellbiolog*", "Einzelzell-Transkriptom*", "Einzelzell-Genom*", "Einzelzellanaly*", "Einzelzelldiagnost*" [translation: "single-cell sequence", "single-cell biology*", "single-cell transcriptome*", "single-cell genome*", "single-cell analysis*", "single-cell diagnosis*"]. Since it was a search for special terms, search functions beyond the title fields (Index = woe) were used. During previous instances of indicator gathering, the academic papers mentioned in the stocks were excluded because they are difficult to access for the interested layperson.[6] However, a look at the researched publications gives the following picture: Only 17 German-language academic papers were found during the gathering timeframe of the German publications.

Publications that are listed in DNB and are visible to the public represent an indicator of a possible yardstick for the public perception of a subject area.

7.4 CONCLUSION

In conclusion, it is possible to make the following points:

- Single-cell analysis is a research field of increasing relevance around the world. Thus, the publication figures increase steadily starting in 2009. The number of articles with German first authorship also reflects this development.
- Even if the indicator "online search queries" only shows the relative search frequency, it is interesting that the search with German keywords in Google Trends showed an excessively low search volume and the only English terms that achieved relatively high search frequency figures were single-cell analysis and single-cell sequencing.
- The new publications collected in DNB comprised only a limited number of German-language publications. In addition, only higher education publications (doctoral theses and habilitation treatises) were recorded. This shows that although the knowledge in the area of single-cell analysis is scientifically prepared and published in the specialist community, it has not yet taken hold or become a topic of discussion in the public domain.

The consistent and strong increase in international specialist publications in the area of single-cell analysis, the partially low relative frequency of online search queries and the low number of new publications reflect the fact that this method is very new. In addition, increasing publication

6 The following were excluded: Periodicals, standard data for individual persons, organizations, events, geographics, specialist terms and work titles, double entries (physical and online publication). English-language publications were removed by hand. No further qualitative filtering of the search results was carried out.

figures may denote the increase in research activities at international and national level. The fact that the establishment of the methods is still relatively new may be a reason why the subject area of single-cell analysis is only discussed in the public domain to a limited extent and is not highly visible in the media. Going forward and with an eye to the increasing data quantity which is associated with the development and establishment of these methods, ethical implications and legal aspects such as questions around data protection, informed consent or various social implications could play a role in the public discussion (see Fangerau, Marx-Stölting, Osterheider, Chapter 6). To this end, the qualitative evaluation of the text corpus and the gathering of problem areas provide initial indications.

7.5 REFERENCES

Diekämper, J./Hümpel, A. (2012): Synthetische Biologie in Deutschland. Eine methodische Einführung [Synthetic Biology in Germany. A Methodical Introduction]. In: Köchy, K./Hümpel, A. et al. (ed.): Synthetische Biologie. Entwicklung einer neuen Ingenieurbiologie [Synthetic Biology. Developing a New Bioengineering]. Forum W, Dornburg: 51–60.

Diekämper, J./Hümpel, A. (2015): Introduction: Gentechnologien in Deutschland im Langzeit-Monitoring [Gene Technologies in Germany in Long-term Monitoring]. In: Müller-Röber, B. et al. (ed.): Dritter Gentechnologiebericht. Analyse einer Hochtechnologie [Third *Gene Technology Report*: Analysis of an Advanced Technology]. Nomos, Baden-Baden: 13–23.

Marx-Stölting, L. et al. (2018): Ausgewählte Indikatoren zu den unterschiedlichen Gentechnologien [Selected Indicators of the Various Gene Technologies]. In: Hucho, F. et al. (ed.): Vierter Gentechnologiebericht [Fourth Technology Report]. Nomos, Baden-Baden: 299–340.

Interdisciplinary Research Group *Gene Technology Report*
Jörn Walter, Heiner Fangerau, Boris Fehse, Jürgen Hampel, Ferdinand Hucho, Martin Korte,
Bernd Müller-Röber, Jens Reich, Jochen Taupitz, Martin Zenke

8. CORE STATEMENTS AND RECOMMENDATIONS FOR ACTION ON SINGLE-CELL ANALYSIS

8.1 CORE STATEMENTS ON SINGLE-CELL ANALYSIS

The significance of single-cell analyses for biomedicine

Higher animals consist of a multitude of different cells – the adult human, for example, has approx. 38 trillion cells. The composition and functioning of cells change over the course of a person's life, development, regeneration, aging, and in the event of illness. With modern single-cell analysis, an area of research is developing that involves the gathering of fundamentally new biological data that opens deep insights into cellular processes at the molecular level. Single-cell analysis generates new approaches for the interpretation of biological interconnections in a context-related and individualized way, which is highly significant for the life sciences, biotechnology, medicine and pharmaceutical research. Up to now, interpretations were mostly based on the analysis of groups of cells or entire tissues and organs, thus they reflected "average values". The functioning and variation range of individual cells could only be captured under certain conditions, or not at all. The new methods and applications of single-cell analysis offer deep insights that have been unachieved to date and will influence biological research and medicine in a sustainable way. For example, single-cell analyses of cells, which were previously classified (e.g., by surface proteins) as a uniform "cell type", show that often presumably identical cells are endowed with similar but not identical programs. This opens up a new and deeper understanding of natural biological variance or a cleaner classification of cell types, enhancing our understanding of fundamental principles of biology, the mechanisms of pathogenesis and the origin of individual diseases. Individual cells isolated from patients or, for example, from organoids established from patient cells, can be classified as "normal" or "deviant". This not only allows to reach profound conclusions on the backgrounds of individual diseases, but also to test how cells in the body respond to specific treatments. Thus, single-cell analysis represents an important step towards personalized medicine.

Single-cell analysis through next-generation sequencing and other omics technologies

After the human genome was decoded around the turn of the millennium, it became clear that the sequence of genomes alone does not deliver conclusive information but requires additional interpretations to comprehend the molecular functioning of cells. Genome sequences have

to be translated into RNA and proteins, both contributing to cell function. In addition, other (downstream) processes such as metabolic status strongly influence individual cell programs. A comprehensive capturing of cell-specific molecular programs takes place at multiple levels: that of the genome (genomics), RNA transcripts (transcriptomics), proteins (proteomics), metabolic products (metabolomics), lipids (lipidomics) and epigenetic programs (epigenomics), to name the main areas. Thus, research is no longer exclusively focused on the analysis of individual genomic programs but now also addresses their complex realization in individual cells.

Novel single-cell technologies have been developed on the basis of existing omics platforms, mostly on the basis of next-generation sequencing (NGS) technologies. The rapid development of new NGS technologies over the last decade allows for fast and efficient sequencing of billions of individual DNA molecules in a short time. One key step for making these deep NGS techniques applicable for single-cell analysis was their combination with microfluidic technologies, allowing massive parallel sequencing of RNA and DNA molecules from single cells. The massive parallelization of sequencing facilitates capture of molecular signatures such as transcriptomes (RNA-seq) or epigenomes (DNA-methylation, open chromatin) from thousands to millions of cells in one sequencing run. Single-cell NGS approaches are complemented by new sensitive single-cell mass spectrometry applications, allowing high (single-cell) resolution profiling of proteins and metabolic products. Finally, new single-cell multi-omics assays are emerging, which demonstrate that simultaneous gathering of transcriptome, chromatin and DNA-methylation can be obtained, opening up a new level of understanding of the link between gene-activity and the consequences of gene regulation within a single cell.

Range of application of single-cell analysis in biology, biotechnology and medicine

Through modern NGS-based single-cell omics technologies, the molecular signatures of up to several million individual cells can be captured. This opens up completely new perspectives for biology. Complex processes such as structural formation in fly larvae or the development of organs can be captured in a precise manner at the level of individual cells. The addition of high-resolution and dynamic imaging techniques allows for the modeling of the spatial allocation and developmental biology dynamic of single cells in the organ or tissue. The future potentials for new insights into developmental processes and diseases are immense. For humans, single-cell omics immediately indicates a wide range of new direct medical applications. These range from the exact determination of the composition and distribution of cell populations (e.g. stem cells, immune cells) through the capture of cellular changes in chronic diseases and definition of the effects of genetic diseases on individual cell types to the high-resolution analysis of individual tumors for individualized treatment (personalized medicine). Single-cell analysis will also play an important role in the rapidly developing research field of organoids.

However, profound single-cell analysis is applicable not only to humans and animals, but also to microorganisms and plants. In bacteria, for example, investigations are underway to determine how individual cells of a bacterial colony differ and whether these differences impact

pathogenicity. Plants are more difficult to investigate due to the solidity of plant cell walls. The problems that are being investigated in the context of plant cultivation include questions around cellular reactions to pathogen attacks and resistance mechanisms to pathogens, the influence of variations in environmental conditions on cellular and developmental processes and the role of genetic networks. New findings could lead to more targeted cultivation and to improvements in the properties of food crops.

Data analysis and infrastructure

Single-cell analysis is already being carried out at many specialized centers in Germany. Beyond an experimental infrastructure, in most cases, these centers have developed methods for data capture, storage and interpretation. The German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) recently equipped four new DNA sequencing centers with the latest infrastructure, which can also generate data for single-cell analysis in high throughput. The implementation of bioinformatic (statistical and modeling) data analysis following individual sequencing poses huge challenges for biology and medicine, to which bioinformatics and data infrastructure are not yet extensively adjusted. Individual data analyses therefore require new and complex data capture and utilization processes for bioinformatic methods. New standards and reference data also need to be generated in this area to enable comparable interpretations. To process the growing volumes of data efficiently and make them accessible and usable for research, artificial intelligence and automated learning methods, such as machine learning (also referred to as deep learning) methods, are increasingly being used for the analyses, especially for complex process modeling. The application of single-cell data in clinical diagnostics will require a complexity reduction of single-cell data and their translation into key statements that are applicable for daily clinical usage.

Implications for specialist areas

Single-cell analysis technologies are developing at a rapid pace. The fast speed of technological innovation requires continuous technical adjustments to ensure that individual researchers and production centers remain internationally competitive. Moreover, intense education on the handling of such technologies must be given more attention, including knowledge of the application options and their limits in the respective fields of research. An important aspect is the growing influence of other disciplines for single-cell data interpretation such as mathematics, bioinformatics and computer science. In this context, it will be especially important to further expand training and the constructive and critical dialog beyond disciplinary boundaries. Alongside this specialist training, appropriate experimental and bioinformatic framework conditions are required so that single-cell data can be used in a sustainable manner.

There will be a wide range of applications and uses of single-cell analysis in life sciences, biotechnology and medicine. Whereas in life sciences the development and wide-ranging application of

various NGS technologies for basic research will remain in the foreground, in biotechnological applications and in medicine the focus will be more on the development and application of standardized processes. In biology, molecular processes can be analyzed in great depth and breadth for the first time, enabling principles of functional commonality and diversity between organisms to be grasped at a new molecular level. Aspects of biodiversity as well as of individual and ecological adaption can thus be determined much more precisely. In medicine, single-cell data will generate new possibilities for individualized molecular diagnosis (e.g., of various cancers) and will be indispensable in the research and application of cell-based processes (stem cells, regeneration, organoids). Single-cell analysis will also play an important role for quality assurance in the area of cell-based test and production methods in pharmacology.

Technology Assessment

As in all new biotechnological applications, it will be important to critically analyze the benefits and the application spectrum, but also the gray areas and limits of the new technology and to discuss these with a broader public. Single-cell analysis entails a range of ethical questions, which have also been discussed in the context of other biotechnologies and are highly relevant to society, above all the handling of sensitive medical data. It is necessary to investigate the extent to which existing rules for responsible handling and adequate data security and sovereignty need to be adjusted to the new possibilities. In the area of research, critical analysis of the data must be intensified in order to avoid misinterpretations and misjudgments.

The collation and joint analysis of genetic (genome) data and single-cell data will deepen the interpretation spectrum and open up new dimensions of certainty at the individual level. These possibilities need to be discussed with regard to their ethical and socio-political implications. The data sovereignty of potential test subjects and patients must be preserved unconditionally in the process. Knowledge of individualized single-cell data takes the question of individuality and the individual expression of the genetic basis to another level. The connection between genotype and phenotype could become detectable to an extent that enables phenotype prediction based on cellular features. Predictions of future diseases or disease progressions that are even more precise than before could then become possible, based on a biopsy, for example. This would mean an enormous knowledge gain compared to standard genetic tests. It is important here to ensure that findings be communicated in such a way that the person affected can understand them and evaluate what they mean. However, the spectrum of new findings regarding gene function and its cellular expression, which will emerge from this, are as yet unclear.

8.2 RECOMMENDATIONS FOR ACTION FOR THE HANDLING OF SINGLE-CELL TECHNOLOGY AND SINGLE-CELL DATA

- Single-cell analysis is a future and key technology for biology and medicine. Its significance will dramatically increase in the coming years. This technology should therefore be afforded a prominent position on research funding agendas.

- Germany has made a very good start (centers) in the use of single-cell technology and boasts proven competence in the bioinformatic processing of single-cell data. These strengths must be maintained and further expanded, for example through research initiatives such as LifeTime or Single Cell Omics Germany (SCOG) and infrastructures such as DFG sequencing centers.

- The clinical use of single-cell data requires the implementation of standardized processes. The "medical informatics initiative" can establish suitable frameworks for this purpose. Standards must be developed in order to prepare complex single-cell data for clinical application, which can be used for the development of new diagnoses and therapy processes. The single-cell data gathered in the clinical context must remain in a protected area and be effectively protected from unauthorized access.

- For a wide use of the generated data (as well as for the gathering of references with which patient samples can be compared), suitable data structures should be established that rely on uniform documentation standards in order to achieve optimal levels of compatibility. Suitable frameworks for data safety and data security must be created in the process (similar to genome data). This should also be reflected in the national research data infrastructure (NFDI).

- It will be important to establish informed consent processes for both single-cell diagnostics and research using personalized data.

- With regards to personal data, single-cell biology does not pose any fundamentally new legal or ethical questions. However, the analysis of individual genomes in single cells may lead to new findings, offering much scope for interpretation with a potential for personal stigmatization or discrimination. Related not only to incidental findings, the right not to know and the protection of personal rights need to be discussed again and more intensively in this context. Legal provisions to protect these rights must be enacted as necessary.

- It is of fundamental importance to ensure that the current hype regarding single-cell analysis does not lead to a reduction or negation of ethical standards that have been already established in other fields. The standards of good scientific, good clinical and good ethical practice must be upheld so that more data leads to more knowledge for the benefit of not just the individual but also society.

9. APPENDIX

9.1 AUTHORS

Dr. Hananeh Aliee
Postdoc at the Institute of Computational Biology, Helmholtz Zentrum München.

Dr. Anna C. Aschenbrenner
Postdoc at the Life & Medical Sciences Institute (LIMES), University of Bonn.

Prof. Dr. Heiner Fangerau
Professor for the History, Philosophy and Ethics of Medicine and Director of the Institute for History, Theory and Ethics of Medicine, Heinrich Heine University Duesseldorf. Member of the Leopoldina, the German National Academy of Science. Member of the IAG *Gene Technology Report*.

Prof. Dr. Boris Fehse
Professor for Cell and Gene Therapy and Head of the Research Department Cell and Gene Therapy within the Clinic for Stem Cell Transplantation, University Medical Center Hamburg-Eppendorf. President of the German Society of Gene Therapy. Member of the European Society of Gene and Cell Therapy and the American Society of Gene Therapy. Spokesperson of the IAG *Gene Technology Report*.

Dr. Nina Gasparoni
Postdoc at the Department for EpiGenetics, Saarland University.

Dr. Jürgen Hampel
Assistant at the Chair for Technological and Environmental Sociology, University of Stuttgart. Member of the Society for Risk Analysis of the European Federation of Biotechnology and staff member of TechnikRadar, a project of the Körber Foundation and acatech. Member of the IAG *Gene Technology Report*.

Prof. Dr. Ferdinand Hucho
Up to 2005 Professor for Biochemistry, Free University Berlin. Member of the German Society for Biochemistry and Molecular Biology (GBM) and of the Berlin-Brandenburg Academy of Sciences and Humanities. Deputy Spokesperson of the IAG *Gene Technology Report*.

Dr. J. Philipp Junker
Group Leader for Quantitative Developmental Biology at the Berlin Institute of Medical Systems Biology of the Max Delbrück Center for Molecular Medicine, Berlin.

Prof. Dr. Martin Korte
Professor for Cellular Neurobiology and Director of the Zoological Institute, Technische Universität Braunschweig. Founding member of the Junge Akademie. Member of the Berlin-Brandenburg Academy of Sciences and Humanities. Member of the IAG *Gene Technology Report*.

Yaroslav Koshelev
Research Assistant at the office of the IAG *Gene Technology Report*.

Dr. Lilian Marx-Stölting
Research Associate at the office of the IAG *Gene Technology Report*.

Dr. Elvira Mass
Head of the Lab for Developmental Biology of the Innate Immune System
at the Life & Medical Sciences Institute, University of Bonn.

Prof. Dr. Bernd Müller-Röber
Professor of Molecular Biology, University of Potsdam and Working Group Coordinator, Max Planck Institute for Molecular Plant Physiology for Signal Networks in Plants. Member of the German National Academy of Science and Engineering acatech and the Berlin-Brandenburg Academy of Sciences and Humanities. Member of the IAG *Gene Technology Report* until 2019.

Angela Osterheider
Research Associate at the office of the IAG *Gene Technology Report* and PhD candidate in communication studies at the Free University of Berlin.

Dr. Christian Popp
Science Project Manager, Program-Management Department at the Berlin Institute for Medical Systems Biology, at the Max Delbrück Center for Molecular Medicine, Berlin.

Prof. Dr. Nikolaus Rajewsky
Professor for Systems Biology and Founding Scientific Director of the Berlin Institute for Medical Systems Biology at the Max Delbrück Center for Molecular Medicine, Berlin

Prof. Dr. Jens Reich
Up to 2005 Research Group Coordinator for Bioinformatics, Max Delbrück Center for Molecular Medicine, Berlin. Up to 2004 university professor, Humboldt University Berlin. Member of the Berlin-Brandenburg Academy of Sciences and Humanities. Member of the IAG *Gene Technology Report*.

Marlen Reinschke
Research Assistant at the office of the IAG *Gene Technology Report*.

Dr. Anna Sacher
Head of Operations and Science Management at the Institute of Computational Biology,Helmholtz Zentrum München.

Hannah Schickl
Research Associate and Coordinator of the office of the IAG *Gene Technology Report* and research associate at the Chair of Systematic Theology II (Ethics), Friedrich-Alexander University Erlangen-Nuremberg.

Prof. Dr. Joachim L. Schultze
Head of the Department for Genomics and Immunoregulation at the Life & Medical Sciences-Institute (LIMES), University of Bonn. Director of the Platform for Single Cell Genomics and Epigenomics PRECISE at the German Center for Neurodegenerative Diseases and the University of Bonn.

Prof. Dr. Jochen Taupitz
Chair of Private Law, Law of Civil Procedure, Private International Law and Comparative Law as well as Director of the Institute for German, European and International Medical Law, Public Health Law and Bioethics, Universities of Heidelberg and Mannheim. Member of the German National Academy of Sciences Leopoldina. Member of the IAG *Gene Technology Report.*

Prof. Dr. Fabian J. Theis
Director of the Institute of Computational Biology and Group Leader on Machine Learning at the Helmholtz Zentrum Munich.

Prof. Dr. Jörn Walter
Professor for Genetics/Epigenetics at Saarland University, Germany. Member of the IAG *Gene Technology Report*.

Prof. Dr. Martin Zenke
Professor of Cell Biology at the Faculty of Medicine of RWTH Aachen University. Chairman and Founding Director of the Institute for Biomedical Engineering, Cell Biology at Aachen University. Member of the Central Ethics Committee for Stem Cell Research. Member of the IAG *Gene Technology Report*.

9.2 MEMBERS OF THE INTERDISCIPLINARY RESEARCH GROUP *GENE TECHNOLOGY REPORT*

Prof. Dr. Boris Fehse (Spokesperson of the IAG)
University Medical Center Hamburg Eppendorf, Clinic for Stem Cell Transplantation.

Prof. Dr. Ferdinand Hucho* (Deputy Spokesperson of the IAG)
Free University of Berlin, Institute of Chemistry and Biochemistry.

Prof. Dr. Heiner Fangerau
University of Düsseldorf, Department of the History, Philosophy and Ethics of Medicine.

Dr. Jürgen Hampel
University of Stuttgart, Institute of Social Sciences.

Prof. Dr. Martin Korte*
Technical University of Braunschweig, Institute of Cellular Neurobiology.

Prof. Dr. Stefan Mundlos*
Institute of Medical Genetics and Human Genetics of the Charité Berlin.

Prof. Dr. Bernd Müller-Röber*
University of Potsdam, Institute of Biochemistry and Biology.

Prof. Dr. Jens Reich*
Max Delbrück Center for Molecular Medicine, Berlin.

Prof. Dr. Silke Schicktanz
Universitätsmedizin Göttingen, Department of Medical Ethics and History of Medicine.

Prof. Dr. Jochen Taupitz
University of Mannheim, School of Law and Economics.

Prof. Dr. Jörn Walter
Saarland University, Institute of Biosciences.

Prof. Dr. Martin Zenke
University Hospital Aachen, Institute for Biomedical Engineering.

* Member of the Berlin-Brandenburg Academy of Sciences and Humanities

9.3 PUBLICATIONS OF THE INTERDISCIPLINARY RESEARCH GROUP *GENE TECHNOLOGY REPORT*

Books

Hucho, F. et al. (Hrsg.) (2018): **Vierter Gentechnologiebericht.** Bilanzierung einer Hochtechnologie. Nomos, Baden-Baden.

Zenke, M. et al. (Hrsg.) (2018): **Stammzellforschung.** Aktuelle wissenschaftliche und gesellschaftliche Entwicklungen. Nomos, Baden-Baden. Unter: https://www.nomos-elibrary.de/10.5771/9783845287720/stammzellforschung [03.07.2018].

Walter, J./Hümpel, A. (Hrsg.) (2017): **Epigenetik.** Implikationen für die Lebens- und Geisteswissenschaften. Nomos, Baden-Baden.

Müller-Röber, B. et al. (Hrsg.) (2015): **Dritter Gentechnologiebericht.** Analyse einer Hochtechnologie. Nomos, Baden-Baden.

Müller-Röber, B. et al. (Hrsg.) (2013): **Grüne Gentechnologie.** Aktuelle wissenschaftliche, wirtschaftliche und gesellschaftliche Entwicklungen. 3. neubearb. u. erg. Aufl. Forum W, Limburg.

Köchy, K./Hümpel, A. (Hrsg.) (2012): **Synthetische Biologie.** Entwicklung einer neuer Ingenieurbiologie? Forum W, Dornburg.

Fehse, B./Domasch, S. (Hrsg.) (2011): **Gentherapie in Deutschland.** Eine interdisziplinäre Bestandsaufnahme. 2. akt. u. erw. Aufl. Forum W, Dornburg.

Müller-Röber, B. et al. (Hrsg.) (2009): **Zweiter Gentechnologiebericht.** Analyse einer Hochtechnologie in Deutschland. Forum W, Dornburg.

Engelhard, M. et al. (2009): **Genetic Engineering in Livestock.** Springer, Berlin, Heidelberg.

Hucho, F. et al. (2008): **Gentherapie in Deutschland.** Eine interdisziplinäre Bestandsaufnahme. Forum W. Dornburg.

Schmidtke, J. et al. (Hrsg.) (2007): **Gendiagnostik in Deutschland.** Status quo und Problemerkundung. Supplement zum Gentechnologiebericht. Forum W, Limburg.

Müller-Röber, B. et al. (Hrsg.) (2007): **Grüne Gentechnologie.** Aktuelle Entwicklungen in Wissenschaft und Wirtschaft. Spektrum, München.

Wobus, A. M. et al. (Hrsg.) (2006): **Stammzellforschung und Zelltherapie.** Stand des Wissens und der Rahmenbedingungen in Deutschland. Supplement zum Gentechnologiebericht. Spektrum, München.

Hucho, F. et al. (Hrsg.) (2005): **Gentechnologiebericht.** Analyse einer Hochtechnologie in Deutschland. Spektrum, München.

Hucho, F./Köchy, K. (2003): **Materialien für einen Gentechnologiebericht.** Grundlagenforschung, Medizinische Anwendung, ökonomische Bedeutung. Spektrum, Heidelberg.

Köchy, K. et al. (Hrsg.) (2002): **Gentechnologie als Wirtschaftsfaktor.** Spektrum, Heidelberg, Berlin.

Booklets

IAG Gene Technology Report (Ed.) (2018): **Fourth Gene Technology Report.** Review of a High-tech Sector. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2018): **Stem Cell Research.** Current Scientific and Societal Developments. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2017): **Epigenetics.** Implications for the Life Sciences and the Humanities. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2015): **Third Gene Technology Report.** Analysis of a High-Tech Sector in Germany. Summary. BBAW, Berlin.

Reich, J. et al. (Ed.) (2015): **Human Genome Surgery.** Towards a Responsible Evaluation of a New Technology. Analysis by the Interdisciplinary Research Group Gene Technology Report of the Berlin-Brandenburg Academy of Sciences and Humanities. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2013): **Agricultural Biotechnology.** Current Scientific, Economic and Societal Developments. Supplement of the Interdisciplinary Research Group Gene Technology Report. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2012): **Synthetic Biology.** Is a New Kind of Biological Engineering Emerging? Supplement of the Interdisciplinary Research Group Gene Technology Report. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2011): **Gene Therapy in Germany.** An Interdisciplinary Survey. Supplement of the German Gene Technology Report. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2009): **Second Gene Technology Report.** Short Version. BBAW, Berlin.

Beier, H. et al. (2009): **Neue Wege der Stammzellforschung.** Reprogrammierung von differenzierten Körperzellen. Ed. by Berlin-Brandenburg Academy of Science and Humanities, Berlin. Available at: http://www.bbaw.de/service/publikationen-bestellen/manifeste-und-leitlinien/BBAW_Stammzellforschung.pdf [26.06.2018].

IAG Gene Technology Report (Ed.) (2008): **Gentherapie in Deutschland.** Eine interdisziplinäre Bestandsaufnahme. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2007): **Genetic Testing and Genetic Diagnosis in Germany.** Status Quo and Controversial Issues. Summary. BBAW, Berlin.

IAG Gene Technology Report (Ed.) (2005): **Gentechnologiebericht.** Analyse einer Hochtechnologie in Deutschland.Summary. BBAW, Berlin.

Other publications

Fehse, B. et al. (2018): **Debatte 19 – Die Gentechnologie in der Gesellschaft:** Von großen Versprechungen, hohen Erwartungen und Missverständnissen. Streitgespräche in den Wissenschaftlichen Sitzungen der Versammlung der Berlin-Brandenburgischen Akademie der Wissenschaften am 01. Dezember 2017. Ed. by Grötschel, M., Berlin. Available at: http://www.bbaw.de/publikationen/neuerscheinungen/pdf/Debatte-Heft-19 [26.06.2018].

Zenke, M. (Ed.) (2017): **Special Issue: Stem cells.** From Biomedical Research Towards Clinical Applications. In: Journal of Molecular Medicine 95(7).Available at: https://link.springer.com/article/10.1007/s00109-017-1548-y [03.07.2018].

Ropers, H. H. et al. (2013): **Stellungnahme zu den neuen Sequenzierungstechniken und ihren Konsequenzen für die genetische Krankenversorgung.** Ed. by Berlin-Brandenburg Academy of Science and Humanities, Berlin. Available at: [19.06.2018].

Gene Technology Report (Ed.) (2003): **Positionen der philosophischen Ethik zur Frage des Klonens.** Info Sheet. Berlin.

Gene Technology Report (Ed.) (2002): **Datenbanken zur Molekularbiologie und Genetik.** Info Sheet. Berlin.

ISBN: 978-3-939818-85-4