MARSILIUS KOLLEG

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

EURAT project group
"Ethical and Legal Aspects of Whole Genome Sequencing"

Position paper

ON THE RELEASE OF RAW GENOMIC DATA TO PATIENTS AND STUDY PARTICIPANTS

Heidelberg, December 2019

UNIVERSITÄTS KLINIKUM HEIDELBERG

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION
Research for a Life without Cancer

NCT NATIONAL CENTER FOR TUMOR DISEASES HEIDELBERG
supported by German Cancer Research Center (DKFZ) Heidelberg University Medical Center Hospital for Thoracic Diseases German Cancer Aid

NCT NATIONAL CENTER FOR TUMOR DISEASES PARTNER SITE DRESDEN UNIVERSITY CANCER CENTER UCC
Supported by: German Cancer Research Center University Hospital Carl Gustav Carus Dresden Carl Gustav Carus Faculty of Medicine, TU Dresden Helmholtz-Zentrum Dresden-Rossendorf

EMBL

Position paper

ON THE RELEASE OF RAW GENOMIC DATA TO PATIENTS AND STUDY PARTICIPANTS

TABLE OF CONTENTS

FOREWORD BY PROF. DR. MICHAEL BAUMANN

The latest sequencing methods for tumor genomes have not only revolutionized our understanding of how tumors form, but they have also provided us with information on concrete therapeutic approaches for developing and administering medication. DKFZ has established one of the largest genome sequencing units in Europe and provides it to a number of research projects. However, DKFZ is also aware that an essential prerequisite for this research is the willingness of patients to provide tumor samples for genome analysis. With a view towards reinforcing the trust patients place in our research, DKFZ scientists, as experts in their fields, are participating in the EURAT (Ethical and Legal Aspects of Whole Sequencing of the Human Genome) platform. The group aims to provide ethical and legal support for the innovative and dynamic field of genome research in the sense of sound leadership, in order, on the one hand, to keep the patients well informed and, on the other hand, dispel any uncertainties pertaining to regulatory and practical questions from the researchers that may hinder their work. The EURAT group already drew up templates for patient information in 2013 and 2015 as well as a code of conduct for genome researchers which is binding on DKFZ and its joint platforms with university medical faculties.

With this current statement on the release of raw genomic data, the EURAT group seeks to address a topic that is becoming increasingly important in research conducted in clinical, translational genome projects. Namely, the interest in and demand for genomic data on the part of patients and study participants. This is also due to the fact that information on the importance and the interest in the data content of the genome continues to grow among the population, whether it is for detecting disease risks at an early stage or simply identifying the genetic origins of one's own family in ancestry databases.

In this statement, the EURAT Group not only draws the conclusion after extensively examining the ethical and legal framework that, even in compliance with legal requirements, patients making a request do indeed

have a right to receive their raw genomic data. It also assesses the interests and possibilities for implementation in the research institutions concerned, specifically for research projects at the DKFZ and its partner institutions. EURAT proposes a procedure for releasing and returning raw genomic data that will enable patients and study participants to use their own raw genomic data responsibly. On the other hand, the information provided along with a written confirmation is also intended to protect the interests of the DKFZ. Concrete information, templates, and procedures have been developed for the practical implementation of a process for the release of raw genomic data.

I very much welcome this development as, with this statement, we are not only establishing responsible practice at the crossroads between research and clinics, but also sending a signal that we take patient participation and patients' rights at DKFZ very seriously.

Heidelberg, November 2019

Prof. Dr. Michael Baumann
Chairman and Scientific Director of
the German Cancer Research Center

FOREWORD BY PROF. DR. BERNHARD EITEL

The scientific and technological advances in life sciences and medicine are impressive and encouraging – for the people concerned and for society as a whole. Many people place big hopes in biomedical research. It is also for this reason that they contribute to scientific progress as patients and study participants. On the other hand, the results of life and medical science research also raise numerous ethical, legal and social questions that need to be considered in the specific fields of research or in the application of new knowledge and technologies. The University of Heidelberg as a research-oriented comprehensive university follows the premise that cutting-edge research goes hand in hand with the responsibility to consider the normative and societal effects, to recognize and name opportunities and risks, and to develop adequate regulatory guidelines.

The close integration of basic research, translation and reflection is part of the self-image of Heidelberg University and is therefore an integral part of its scientific and social mandate. In modern life sciences and medical research, social responsibility is particularly evident and sensitive, since new research results are particularly suitable for raising hopes for relief, healing and improving the quality of life of people with illnesses or disabilities. The project group EURAT (Ethical and Legal Aspects of Wholegenome sequencing of the Human Genome), which was created as part of a Marsilius project, is an excellent example of how this integration can be successful for the benefit of patients, doctors and researchers. With the 2013 statement “Key points for a Heidelberg practice of whole genome sequencing” (2nd, updated edition 2015), the project group presented a proposal for institutional self-regulation of whole genome sequencing.

Since then, this practical set of rules has had an impact far beyond Heidelberg and was subject of many public and scientific discussions. The position paper was grounded on the pooled expertise of life sciences, medicine, bioinformatics, law and ethics at Heidelberg as a center of scientific excellence.

The Marsilius College, as an interdisciplinary center for advanced studies at the Ruperto Carola, offered the framework to develop a scientific, well-founded, ethically and legally sound and medically practicable regulation.

The EURAT project group continued on this path with its latest position paper on the „Release of genomic raw data to patients and study participants“, which is devoted to the questions of informational self-determination in human genetics research and diagnostics. It puts the rights of patients and study participants in relation to the informational and informed consent obligations of involved responsible medical and scientific parties. Based on that evaluation it developed a balanced overall procedure. The statement once again proves that the intensive cooperation of scientists from different disciplines in cooperation with non-university research institutions at the Heidelberg research location brings immediate and practical benefits not only for research, but also for patients and study participants. The University of Heidelberg thanks all people and institutions involved in this statement and hopes that this Heidelberg contribution will in turn fertilize and inspire the national and worldwide discussion about the ethical, legal and social implications of genome research and its clinical applications.

Heidelberg, November 2019

Prof. Dr. Dr. h.c. Bernhard Eitel
Rector of Heidelberg University

CONTRIBUTING AUTHORS

Project spokeswoman

Prof. Dr. Dr. Eva Winkler EURAT spokeswoman,
Professor of Translational Medical Ethics
National Center for Tumor Diseases
Heidelberg (NCT),
Heidelberg University Hospital

Scientific coordinator

Dr. Irina Idler Scientific Coordinator of EURAT
National Center for Tumor Diseases (NCT),
German Cancer Research Center (DKFZ)
Heidelberg
Contact: irina.idler@nct-heidelberg.de
Tel.: +49 6221 5638695

Members of the working group

Dr. Katja Beck Department of Translational Medical Oncology
National Center for Tumor Diseases Heidelberg and German Cancer Research Center, Heidelberg

Prof. Dr. Benedikt Brors Applied Bioinformatics
German Cancer Research Center,
Heidelberg

Prof. Dr. Kai Cornelius Law Faculty
Heidelberg University

Dr. Nicola Dikow Institute of Human Genetics
Polyclinic, Heidelberg University Hospital

Dr. Henrike Fleischer Institute for German, European, and In ternational Medical Law, Health Law and Bioethics of the Universities of Heidelberg and Mannheim (IMGB)

Prof. Dr. Stefan Fröhling	Department of Translational Medical Oncology National Center for Tumor Diseases Heidelberg and German Cancer Research Center, Heidelberg
Prof. Dr. Hanno Glimm	Department of Translational Medical Oncology National Center for Tumor Diseases Dresden, University Hospital and German Cancer Research Center, Heidelberg
Prof. Dr. Esther Herpel	Department of General Pathology and Pathological Anatomy Institute of Pathology, Heidelberg University Hospital
Dr. Jan Korbel	Research unit of Bioinformatics, Genome Sequencing European Molecular Biology Laboratory (EMBL), Heidelberg
Prof. Dr. Peter Lichter	Department of Molecular Genetics German Cancer Research Center, Heidelberg
Dr. Daniela Richter	Department of Translational Medical Oncology National Center for Tumor Diseases Dresden German Cancer Research Center, Heidelberg
Dr. Christoph Schickhardt	Research Group in Translational Medical Ethics National Center for Tumor Diseases Heidelberg German Cancer Research Center, Heidelberg
Prof. Dr. Klaus Tanner	Systematic Theology and Ethics Faculty of Theology, Heidelberg University
Prof. Dr. Stefan Wiemann	Department of Molecular Genome Analysis German Cancer Research Center, Heidelberg

Editorial team

Irina Idler, Christoph Schickhardt, Kai Cornelius, Eva Winkler

All Members of the EURAT Project Group

Prof. Dr. rer. nat. Benedikt Brors
Prof. Dr. Kai Cornelius
Dr. med. Nicola Dikow
Prof. Dr. rer. nat. Roland Eils[1]
Prof. Dr. med. Stefan Fröhling
Prof. Dr. med. Hanno Glimm
Prof. Dr. med. Esther Herpel
Dr. rer. nat. Irina Idler
Prof. Dr. med. Christof von Kalle[2]
Dr. rer. nat. Jan Korbel
Prof. Dr. med. Andreas Kulozik[3]
Prof. Dr. rer. nat. Peter Lichter
Dr. rer. nat. Daniela Richter
Dr. phil. Christoph Schickhardt
Prof. Dr. med. Peter Schirmacher[4]
Prof. Dr. theol. Klaus Tanner
Prof. Dr. rer. nat. Stefan Wiemann
Prof. Dr. med. Dr. phil. Eva Winkler

[1] Professor Dr. Roland Eils, BIH Centre for Digital Health, Berlin Institute of Health (BIH) and Charité, University Medicine Berlin.
[2] Professor Dr. Christof von Kalle, Department of Translational Oncology, German Cancer Research Center (DKFZ), Heidelberg.
[3] Professor Dr. med. Andreas Kulozik, Clinic for Pediatric Oncology, Hematology, Immunology and Pulmonology, University Medical Center for Children and Adolescents, Heidelberg University Hospital.
[4] Dr. med. Peter Schirmacher, Department of General Pathology and Pathological Anatomy, Institute of Pathology, Heidelberg University Hospital.

PREAMBEL

- With awareness of the fact that we are entering unchartered ethical, legal, and organizational territory regarding the issue of releasing and returning raw genomic data
- With a commitment to respecting the rights of patients and study participants to make decisions, based on the idea of informational self-determination
- With the objective of enabling patients and study participants to use their raw genomic data responsibly
- In recognition of the need to keep the time and effort required for releasing the raw data in balance with the primary tasks of researchers and physicians
- With the intention of proactively designing a process of responsible handling of the release of raw data
- With the understanding that the persons involved rely on practical recommendations for the responsible implementation of the ethical and legal requirements

the EURAT group issues the following statement and practical recommendations.

1. OBJECTIVES

(1) This position paper aims to provide an overview of the legal framework and conditions of applicable legal requirements currently in place regarding the release and return of raw genomic data.

(2) It takes a comprehensive look at ethical aspects of the release of raw genomic data, addresses the different perspectives of the stakeholders involved, and assesses them in order to develop a basis for responsible action.

(3) Drawing on the legal and ethical analysis, recommendations are developed with practical relevance for handling requests for raw genomic data from patients and study participants. These recommendations include a manual on procedures, a template for documented information for patients and study participants and a further template for the written confirmation of receipt of raw data.

(4) These practical tools and provision of informative materials shall provide guidance to enable patients and study participants to make responsible decisions.

(5) Additionally, the written confirmation and information material for study participants within the context of the releasing procedure of raw data aim to safeguard the interests of the institution.

(6) The primary objective is to open up the discussion on the rights to the release of raw data and determine the framework for responsible practice at the interface between research and the clinic.

2. INTRODUCTION

In recent years the significant technological progress in the field of sequencing the human genome has led to a widespread application of sequencing technologies (NGS),[5] not only in basic research but also in clinical research and in the field of clinical diagnostics (see 2.1). Large volumes of raw genomic data are generated in whole genome sequencing (WGS) and whole exome sequencing (WES) for diagnostic and research purposes. Patients and study participants[6] whose genomes are sequenced in these studies are now increasingly showing interest in their own genome and more often express their desire to have their raw genomic data released to them.[7]

Until now, the ethical and legal discussion on the release of data from sequencing analyses to study participants has focused mainly on the primary and secondary findings from genomic analyses, their interpretation and validation, as well as the informed consent process necessary for the inclusion in studies. Accordingly, there has been a wide-ranging debate, with varying statements from several scientific organizations and specialist associations, on the release of results from genome research.[8,9,10] In its most recent paper on good scientific practice in genome sequencing, the EURAT group outlined its position that supports the release of validated findings in accordance with the preferences of study participants.[11]

In contrast, the issue of releasing raw genomic data to study participants (see 2.3.2.) has received little attention thus far as such requests represent a relatively new development. Therefore, with this recommendation on the handling of requests for the release and return of raw genomic data, we are

---

5 The term "next-generation sequencing" technology (NGS), also known as high-throughput sequencing, describes a number of different modern sequencing technologies. These make it possible to sequence DNA and RNA at a faster rate and at less cost than was possible with the previously used Sanger sequencing.

6 Whenever this statement refers to "study participants", this reference includes patients and study participants. For the sake of easier readability, the term "study participants" is used more frequently in this statement. In passages where a distinction between the terms is necessary for content or legal reasons, the terms are listed separately.

7 Middleton, Anna, et al. "Potential research participants support the return of raw sequence data." *Journal of medical genetics* 52.8 (2015): P. 571-574.

8 German Society for Human Genetics (GfH) – S1 Guideline on NGS Diagnostics: Molecular genetic diagnostics with high-throughput methods of the human genome, for example with next-generation sequencing. (Version: Sept.15, 2017).

9 National Committee on Health Research Ethics - Guidelines on Genomics Research (June 2018).

6 Biesecker, Leslie G. „ACMG secondary findings 2.0." *Genetics in Medicine* 19.5 (2017): p. 604.

11 For validation: EURAT, Statement. Position paper on the practice of whole genome sequencing in Heidelberg, 2nd ed. (2015): p. 23.

entering new territory, not only from a legal and ethical point of view but also in terms of logistical aspects.

First of all, raw genomic data must be clearly distinguished from primary and additional genomic findings, i.e. clinically relevant, validated information from sequencing analyses, interpreted by experts, which provides clearly delimited and communicative content. In contrast, by releasing raw genomic data study participants are given data from early processing stages (see 2.3.1). The informational content of this data is unspecified, since raw data are not evaluated in terms of the interpretable segments of the genome on the one hand, and on the other, some parts of the genome are not yet fully understood.

The question arises as to how the growing interest and the increased demand can be explained given the limited informational value of raw genomic data. In view of the significant progress and decreasing costs of sequencing technology, genome sequencing is increasingly being integrated into clinical diagnostics. As a result, more and more people are inevitably being introduced to the topic of genomics. This may explain the increased interest. Although most study participants will not be able to carry out the bioinformatic analysis required to obtain meaningful information on their own, it is nevertheless conceivable that they will approach an institution or a commercial analysis service of their choice for further processing.[12] The underlying motivation could be in obtaining a second opinion, further analyses, reanalysis, or subsequent interpretations of their raw genomic data.

It is also conceivable that some study participants would like to have a list of their uninterpreted genetic variants so that they can track the current state of research on these, either out of curiosity or to gain benefit for themselves from the information as it becomes available. At this stage, only clinically relevant variants are released to study participants who have agreed to the respective feedback after their validation within a finding. However, the dynamic increase in scientific knowledge concerning genomics (see 2.1) could still change the classifications and interpretations of certain variants, or new clinically significant variants could be added. It is therefore understandable that people want to monitor the evolving scientific knowledge independently. It has also been reported that study participants may seek to contribute their raw data to several research projects for altruistic reasons and may hand over the data themselves.

[12] Badalato, Lauren, Louiza Kalokairinou, and Pascal Borry. "Third party interpretation of raw genetic data: an ethical exploration." *European Journal of Human Genetics* 25.11 (2017): p. 1189.

2.1. Current applications of NGS technologies in clinical genomics

The immense increase in knowledge in basic medical research combined with advances in sequencing technologies have accelerated the inclusion of "next-generation sequencing" (NGS) technologies in diagnostics and thus patient care (clinical genomics).

In **human genetic diagnostics**, NGS investigations are currently limited to the use of various multi-gene panels, with the help of which several known genes associated with disease can be analyzed simultaneously in a single test.[13] The so-called panel approach represents a further development of the previous step-by-step diagnostic tool method of Sanger sequencing, with which diagnostically relevant genes related to the disease are investigated. A wide range of gene panels, each specific to a disease or phenotype, has become available for diagnosing various diseases, e.g. for congenital muscular dystrophies (CMDs),[14] limb-girdle muscular dystrophies or mitochondrial diseases,[15] cardiomyopathies,[16] and various cancer dispositions.[17,18] In addition, gene panels are used to characterize the human leukocyte antigen locus, which plays a significant role in typing HLA characteristics in transplantation and transfusion medicine.[19]

The more extensive the genomic investigations through parallel evaluation of many genes are, the greater the requirements for bioinformatic expertise for the subsequent analysis and interpretation of the many variants identified. Therefore, the multi-gene panel approach is still the method of choice for diagnosing rare diseases within the scope of standard care. More extensive methods, such as WES and WGS, are rarely used in routine diagnostics. The use of NGS panels up to a volume of <25 kilobases as "basic diagnostics" in the field of human genetics has, so far, been covered by the statutory health insurance providers for most indications.[20]

---

13 Rehm, Heidi L. "Disease-targeted sequencing: a cornerstone in the clinic." *Nature Reviews Genetics* 14.4 (2013): p. 295.

14 Valencia, C. Alexander, et al. "Assessment of target enrichment platforms using massively parallel sequencing for the mutation detection for congenital muscular dystrophy." *The Journal of Molecular Diagnostics* 14.3 (2012): p. 233-246.

15 Vasta V. et al. "Next generation sequence analysis for mitochondrial disorders" *Genome Med.* 1 No. 10 (2009), p.100.

16 Teekakirikul, Polakit, et al. "Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era." *The Journal of Molecular Diagnostics* 15.2 (2013): p. 158-170.

17 Pritchard, Colin C., et al. "ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing." *The Journal of Molecular Diagnostics* 14.4 (2012): p. 357-366.

18 Mitra, Anirban P., et al. "Generation of a concise gene panel for outcome prediction in urinary bladder cancer." *Journal of Clinical Oncology* 27.24 (2009): p. 3929.

19 Schöfl, Gerhard, et al. "2.7 million samples genotyped for HLA by next generation sequencing: lessons learned." *BMC genomics* 18.1 (2017): p. 161.

20 According to the German Medical Association Berlin, New Uniform Assessment Scale (EBM) from January 1, 2019 (2/24/2019), URL: https://www.kbv.de/html/ebm.php (Retrieved: 4/4/2019).

Gene panels are able to greatly shorten the diagnostic odyssey for some patients, but for rare diseases with very rare disease-causing genetic variants (rare mendelian disorders), many patients still remain without diagnosis. For this reason, some specialized institutes also provide more expansive analyses, such as exome sequencing, in justified individual cases to improve diagnostic power.

Exome sequencing is intended to help identify the causes of rare diseases in even very heterogeneous diseases or in patients with hitherto unknown syndromes. In pediatrics, in particular, exome sequencing is used to investigate developmental delays of unknown cause to facilitate diagnosis. [21,22,23]

Within the statutory insurance sector, more comprehensive NGS studies, i.e., larger panels (over 25 kb) and exome analyses, are included in the single assessment scale (EBM) catalog, but these require regulatory approval and must be applied for.[24] In statements by the Joint Federal Committee (G-BA), there is a current discussion on whether non-invasive prenatal diagnostics (NIPD) for determining the risk of autosomal trisomies 13, 18 and 21 in the mother's blood by means of a molecular genetic test (NIPT) for use in high-risk pregnancies (within the scope of maternity guidelines) should be included in the services covered by statutory health insurance providers. Testing is carried out using, among other ways, whole genome sequencing.

Pharmacogenomics is another branch of research in which the use of exome sequencing (WES) has already gained relevance. This field of research deals with the influence of hereditary dispositions on the effect and toxicity of medicinal drugs. It focuses on the stratification of patient groups according to therapy response to drugs based on the relevant genetic variants. Few pharmacogenetic studies have so far been put into practice.[25]

The most comprehensive form of NGS analysis, the sequencing of the whole genome (WGS), is still generally regarded as a research instrument and is being investigated in translational research projects for possible

---

21 Clark, Michelle M., et al. "Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases." *NPJ genomic medicine* 3 (2018).

22 Botstein, David, and Neil Risch. "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." *Nature genetics* 33.3s (2003): p. 228.

23 Ng, Sarah B., et al. "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." *Nature genetics* 42.9 (2010): p. 790.

24 According to the German Medical Association Berlin, New Uniform Assessment Scale (EBM) from January 1, 2019 (2/24/2019), URL: https://www.kbv.de/html/ebm.php (Retrieved: 4/4/2019).

25 Shi, Hubing, et al. "Melanoma whole-exome sequencing identifies V600E B-RAF amplification-mediated acquired B-RAF inhibitor resistance." *Nature communications* 3 (2012): p. 724.

added value in regular care. The term "translational" is used to describe how knowledge gained from research becomes part of new diagnostic, therapeutic, and/or preventive clinical applications.[26,27,28] In pediatric translational clinical trials, WGS is increasingly emerging as an effective method for diagnosing rare or unknown diseases.[29] As a result, WGS applications could steadily increase over the next few years and gradually become part of the standard treatment for certain cases.[30]

Oncology is the field in which the use of NGS technologies for patient care is particularly advanced. As such, understanding of the molecular basis has revolutionized the diagnostic classification and therapy for some tumor entities.[31,32,33] Investigations of gene mutations by gene panels in an increasing number of tumor entities are of primary importance and part of the pathological report, not only for diagnosis and prognosis, but also for the planning of tumor-specific targeted therapies.[34]

Within the scope of translational research projects, tumors are largely characterized molecularly by means of NGS technologies. New therapeutic targets can be determined based on the molecular properties of the tumors identified.

In the precision oncology programs in Heidelberg, for example, the patients' tumor material is sequenced and examined for molecular targets to find specific or immunotherapeutic treatment options.[35,36] Sequencing analyses include whole genome, exome, and transcriptome analysis. These research programs are primarily intended to reach a better understanding of tumor biology. However, the molecular information obtained forms

[26] Alizadeh, Ash A., et al. "Toward understanding and exploiting tumor heterogeneity." *Nature medicine* 21.8 (2015): 846.

[27] Joffe, Steven, and Franklin G. Miller. "Mapping the moral terrain of clinical research." *Hastings Center Report* 38.2 (2008): p. 30-42.

[28] Rosenberg, Roger N. "Translating biomedical research to the bedside: a national crisis and a call to action." *Jama* 289.10 (2003): p. 1305-1306.

[29] Saunders, Carol Jean, et al. "Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units." *Science translational medicine* 4.154 (2012): p. 154ra135-154ra135.

[30] Bick, David, et al. "Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases." *Journal of medical genetics* (2019): jmedgenet-2019.

[31] Louis, David N., et al. "The 2016 World Health Organization classification of tumors of the central nervous system: a summary." *Acta neuropathologica* 131.6 (2016): p. 803-820.

[32] Pajtler, Kristian W., et al. "Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups." *Cancer cell* 27.5 (2015): p. 728-743.

[33] Müller-Reible C., "Sequenzierung in Europa. Hochdurchsatzverfahren und Regelversorgung." *GID* 231 (2015).

[34] Kamel-Reid, Suzanne, et al. "Validation of KRAS testing for anti-EGFR therapeutic decisions for patients with metastatic colorectal carcinoma." *Archives of pathology & laboratory medicine* 136.1 (2012): p. 26-32.

[35] Perera-Bel, Júlia, et al. "From somatic variants towards precision oncology: Evidence-driven reporting of treatment options in molecular tumor boards." *Genome medicine* 10.1 (2018): p. 18.

[36] Horak, Peter, Stefan Fröhling, and Hanno Glimm. "Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls." *ESMO open* 1.5 (2016).

the basis for the continued translational program, which aims to provide the affected patients with access to new, individualized treatment options through a series of clinical trials.

In specialized sequencing programs such as those at the Centers for Personalized Medicine in Baden-Württemberg[37] or within the scope of the German Consortium for Translational Cancer Research (DKTK), the results of the comprehensive molecular characterization of the tumor for each individual patient are discussed in an interdisciplinary panel (molecular tumor board) consisting of oncologists, bioinformaticians, biologists, pathologists, and human geneticists. The therapeutic relevance of the determined variants and potential treatment options beyond the approved lines of therapy, as well as their possible inclusion in clinical trials are discussed, together with the respective physicians.[38]

2.2. Background and range of topics included in the statement

The positions on patients' side requests for raw genomic data, which are presented in practice as well as in the overseeable amount of literature on the topic, remain controversial. While some authors are critical of the release and return of raw data to study participants,[39] as these individuals are not able to do anything with it and would actually need specialists to conduct analyses and interpretations, other authors strongly and fully support the release and return of raw data.[40] Access to one's own raw data, which stem from the samples submitted for research projects or biobanks, increases reciprocity, it is argued, and gives people the opportunity to determine the way they use their own genomic data.[41] Still other authors link the release of raw data to conditions such as offers of assistance and information, and access to upstream and downstream genetic advice.[42] The authors who are in favor of releasing and returning raw data under the aforementioned conditions base their position on a study carried out by Middleton et al., in which the feedback preferences of different sections of the population (members of the public, without special genetic expertise, non-genetic health professionals and human geneticists) with a view

[37] https://www.aerzteblatt.de: "Baden-Württemberg bekommt vier Zentren für personalisierte Medizin." *Dtsch Arztebl news* (July 22, 2019).
[38] Merry, David, et al. "Trust and responsibility in molecular tumour boards." *Bioethics* 32.7 (2018): p. 464-472.
[39] Bredenoord, Annelien L., et al. "Disclosure of individual genetic data to research participants: the debate reconsidered." *Trends in Genetics* 27.2 (2011): p. 41-47.; Kaye et al., Medical Law Review 22 .1 (2014): p. 64, 73.
[40] Lunshof, Raw Personal Data: Providing Access. Science 343 (6169), (2014): p. 373-374.
[41] Ebd.
[42] Wright, Caroline F., et al. "Returning genome sequences to research participants: Policy and practice." *Wellcome open research* 2 (2017).

towards the different categories of additional findings from genome studies were recorded. This study shows that among all the groups analyzed the genetic counselors were the most cautious in terms of reporting knowledge of genomic risks. This study also asked how many "lay people" might be interested in their own raw genomic data as study participants and what might motivate such an interest. The majority of respondents (61%) stated that, as study participants, they would like unrestricted access to all of their personal genomic data. An independently sought analysis also seemed conceivable to the majority of respondents.[43] However, to our knowledge, no data on the frequency of raw genomic data actually released to study participants have so far been published.

The various points of view in the debate on the handling of requests for raw data indicate that, on the one hand, there is a need for a well-considered and normatively justified position on requests of this kind. On the other hand, there is an increasing need for specific guidelines that can be consulted for responses and implementation at the institutional level. This current statement by the EURAT group, which is based on the preparatory work from a BMBF-funded research project,[44] seeks to fulfill these requirements. In the following section, the content structure of the statement, including recommendations and objectives are introduced. The terms relevant to the topic are then discussed in detail.

To begin, Chapter 3 looks at whether study participants have a right to the release of their raw data. The discussion clearly shows how difficult it is in many cases to distinguish between the research and treatment contexts in order to define the applicability of the relevant legal norms. Due to the intended close interplay of clinical application and research in the translation of knowledge, it is difficult to clearly determine the applicable legal position in some cases. The current legal norms relevant to the release of raw data are the subject of the legal analysis of K. Cornelius, based on the preparatory work of H. Fleischer and C. Schickhardt (see Chapter 3).[45]

Drawing on the legal analysis, Chapter 4 addresses the ethical aspects that should be considered when releasing raw data. To begin, the moral rights of study participants to access their raw genomic data are established and examined to determine whether conflicts of interest and possible risks for

[43] Middleton, Anna, et al. "Potential research participants support the return of raw sequence data." *Journal of medical genetics* 52.8 (2015): p. 571-574.
[44] Fleischer, Henrike, et al. "Das Recht von Patienten und Probanden auf Herausgabe ihrer genetischen Rohdaten." *Medizinrecht* 34.7 (2016): p. 481-491.
[45] Ebd.

the various parties involved (medical staff, researchers, study participants, patients) can normatively justify a restriction of or deviation from the right to raw data. The justification behind a claim to raw data is that genomic data is a form of personal data that contains potentially valuable information about the study participants. They should be able to access and freely dispose of their personal data. Based on the notion of "caring liberalism", ignorance on the part of the study participants or possible erroneous assumptions about genomic information and the resulting risky handling of data are not counted as valid arguments against their release, but rather they constitute an obligation to inform the recipient about the nature and usefulness of their raw data. Ideally, study participants will be able to make an informed decision about whether and why they want to receive and use their raw data.

This analysis of the ethical and legal position has direct implications for the practical implementation at clinics and research institutions, since the responsible handling of the requests for raw data is a matter of procedure, logistics, and communicative requirements or challenges. Few authors have thus far dealt with the question of what should be included in a handout on the release of raw data to individual study participants.[46,47] An important suggestion, discussed in Chapter 5, is whether an additional authentification step on the part of the study participants is necessary regarding quality assurance to ensure that the "correct" genomic data are actually handed over to the "right" person.[48]

In Chapter 5, recommendations are developed for the process of releasing raw data that are based on the normative legal-ethical analysis and thus aim to ensure study participants are well-informed regarding the handling of their own raw genomic data. At the same time, however, the procedure for releasing the raw data should be carried out within an appropriate framework compatible with the primary tasks of physicians and researchers.

The implementation of the theoretical pre-consideration and the best practice recommendations for the procedure for releasing raw data are significantly helped by specific and practical documents and informational materials. These information materials must not necessarily be redeveloped at every location. Therefore, this position paper will be supplemented

[46] Wright, Caroline F., et al. "Returning genome sequences to research participants: Policy and practice." *Wellcome open research* 2 (2017).
[47] Thorogood, Adrian, et al. "APPLaUD: access for patients and participants to individual level uninterpreted genomic data." *Human genomics* 12.1 (2018): p. 7.
[48] Ebd. Wright, Caroline F., et al.

by templates for informing study participants as well as a form in which study participants "acknowledge" in writing the receipt of the data and information offered (see 7.2). Finally, Chapter 6 summarizes the practical recommendations of this position paper with a clear outline.

2.3. Definition of terms

2.3.1. Raw genomic data

The German Genetic Diagnosis Act (GenDG) defines "genetic testing results" as the results of a genetic analysis, including their interpretation, taking individual circumstances into consideration (Section 11 GenDG). In contrast to an actual, validated testing result (= finding), raw genomic data are not differentiated, specified, or interpreted regarding their specific medical and social significance for the individual participant. Raw data must therefore be clearly distinguished, on the one hand, from "results" or "findings" in the research context, and, in particular, from the final clinical stage of data processing, the quality-assured, validated findings. Based on this definition, it is clear that raw data cannot be seen as results, insofar as they do not allow for statements on genetic disposition without further analysis. Consequently, the study participants are not directly confronted with a genetic finding/result when raw genomic data is released to them.

Identifying and classifying variants from NGS data entails complex processing methods and several consecutive analysis steps. The bioinformatic data processing, which is usually done automatically, can be used to determine the data type from a sequencing analysis that can still be assigned to "raw genomic data".

In general, in this position paper the authors use raw genomic data as a collective term for very early (primary) and early (secondary) stages of bioinformatic data processing of a sequencing analysis (see Information box). Therefore, this statement describes all file formats stemming from the actual sequencing of a sample to the processing stage of the so-called variant call, before their interpretation and annotation, as "raw" (FASTQ, BAM, VCF files). Variant call lists of genetic variants that have not yet been annotated and interpreted are therefore also included in the "raw" data.

Although datasets of the so-called differential genome, which show the totality of all differences between the germline genome and, for example, the tumor genome of a human being or the totality of all deviations between

the germline genome of a human and the international reference germline genome, contain results from subsequent processing steps. These data have not yet been interpreted.

Information box

Generation of raw genomic data: primary, secondary, and tertiary data types

(1) The primary form of sequencing is storage intensive image-data; these images are translated into a text format with identified DNA/RNA bases (FASTQ files) on the control computer:
The most original raw data of the sequencing machines are image-data taken by CCD (charge-coupled device) chips. These are processed immediately, since these image data are too large and it would not make sense to store them permanently.

In an initial analysis step, the image-data are used to determine the base sequence of each sequenced section. This step is called "basecalling" and carried out on the computer connected to the sequencing machine. The original image-data are then automatically deleted. FASTQ files created in this process represent the pure sequence of DNA/RNA. From a technical perspective, FASTQ files could generally be considered as files of the sequencing results. However, this technical understanding of "result" does not correspond to the kind of "result", as laid out in this statement, as something of immediate importance to the people affected.

(2) The following alignment of the reads with the reference genome (SAM and BAM files) and the identification of variants (VCF files) can be summarized as a secondary form of data processing:
The human genome consists of about 3 billion base pairs, which are sequenced in a whole genome analysis. To sequence an entire human genome, a series of short reads (100 base pairs, depending on the sequencing platform) are usually generated and aligned with the reference genome. Each base of the genome is spanned by multiple reads. The number of reads at a point in the genome is also known as coverage. For example, an entire genome sequenced with 30x coverage means that on average, each base of the genome is covered by 30 sequencing reads. This high coverage is important to ultimately achieve a high quality of the resulting genome sequencing. Millions of short 100-base reads are generated, most of which are stored in FASTQ format. In addition to the letters of each base position

(also called base calling), these file formats also store a wide range of additional information (meta information), e.g., on the quality of sequencing. A typical FASTQ file therefore contains both the pure sequence of DNA/RNA and quality information. Their overall size is approximately 200 gigabytes for a whole genome analysis.

The generated data are then matched against a reference genome. By default, the result of mapping base sequences from the FASTQ files to the reference genome is stored in a Sequence Alignment/Map (SAM) file. To save disk space, the SAM files can be converted to binary Alignment/Map (BAM) files (approximately 100-150 GB) that require less disk space. The content is converted to binary code and can no longer be deciphered by humans. However, BAM files can be converted back to FASTQ format if necessary and are therefore suitable for long-term storage.

After the bases of the processed sequences have been identified, the resulting reads have been stored in the FASTQ files with the corresponding quality information, and they have been aligned with the reference genome, the resulting SAM files can be used to identify the variants. The genomes of two people differ by about 0.1% in terms of single-base variants (SNPs). This equates to about 3 million identifiable variants in an average human genome that can be detected per whole genome analysis. Additionally taking the structural variations into account, the genomes of two people differ by about 0.5-1%.

A list is created in the so-called Variant Call file format (VCF file), which contains all variants where the sequenced sample differs from the human "reference genome".

(3) Annotation, filtering, functional predictions, and the biomedical interpretation of variants can ultimately be defined as tertiary analyses. It is only at this stage that "result data" are produced in the proper sense which may influence the treatment of the respective patients or contain information on the predisposition of diseases: the term "variant call file" initially incorrectly suggests that the variant identification is sufficient to be able to identify, for example, tumor-relevant variants up to that point. However, subsequent processes such as annotations, filtering, and biological interpretation of the numerous variants and possibly other experiments may be necessary for classifying the variants. In order to determine, for example, the tumor-specific variants of an individual person, the variants that can also be found in that person's healthy tissue are

subtracted from the identified variants in a filtering process. As a result, the results include tumor-specific changes. However, not every identified tumor-specific change is necessarily relevant for therapy recommendations (e.g. passenger mutations). For this reason, variants are then interpreted to identify the meaningful variants that have a more likely effect on, for example, cell degeneration and/or therapy recommendation.

Table 1: Overview of the size and properties of file fomats of the initial sequencing steps from whole genome analysis

File format	Description	Required storage space for a genome (approx. 30x coverage)
FASTA/Q	· **FASTA:** a text-based format consisting of multiple DNA sequences, each with a description text. · **FASTQ:** similar to FASTA, additionally stores a quality rating for each sequenced base.	100–300 GB*
SAM (**S**equence **A**lignment **M**ap)	· Format for storing a long DNA sequence; · it is used to compare the reference genome and for the quality assessment	~500 GB
BAM (**B**inary **A**lignment **M**ap)	· Binary format of the SAM file; · a lossless compressed format for SAM; · it can be transformed back into the FASTQ format	~100 GB
VCF (**V**ariant **C**all **F**ormat)	· Text file format with a list of sequence variants that are different from the reference genome. · Variants are sorted by their position in the genome and usually annotated with their allele frequency.	~125 MB

*GB, Gigabytes; WES, **W**hole **E**xome **S**equencing (complete protein coding region - 50-60 million bases); WGS, **W**hole **G**enome **S**equencing (~ 3 billion bases).

2.3.2. The terms “patients” and “study participants”

To improve readability, this statement uses the term “study participants” whenever possible, which is intended to represent both patients and participants in clinical trials. Only in passages where a distinction between the terms is necessary for understanding the content or legal reasons are the terms “patients” and “study participants” distinguished.

2.3.3. Differentiation and transitions between treatment and research context

Traditionally, medical treatment in the context of a doctor-patient relationship is characterized by compliance with a recognized and established medical standard for the treatment of patients, without any expectation or intention to gain research knowledge from the treatment (Section 630a et seq. German Civil Code (BGB)).[49] Research to the benefit of third parties, on the other hand, does not aim to benefit a specific, individual patient, but rather gain knowledge for the purpose of exceeding and improving current medical standards.[50] In practice, however, it is assumed that there is a continuum, at the end of which measures which may be regarded as "pure treatment" and at the other end of which "pure" research activities are considered.[51] On the continuum between these two poles, there are different measures that have varying ratios and qualifiers for both treatment and research-typical characteristics. Particularly in translational, patient-oriented research, the aim is to have close interconnection between treatment and an increase in knowledge in a particular field. Given the corresponding difficulties in differentiating between these two poles, the question of releasing raw data in this statement is analyzed separately for the two poles, which ideally are seen separately as pure treatment on the one hand and pure research on the other.

2.3.4. The terms "genomic" and "genetic"

This position paper uses the term "genomic" and not "genetic" raw data. The literature applies both terms "genetic" and "genomic" when speaking of raw data. The term "genomic" refers to a wide range of genomic data that can be generated by high-throughput sequencing of the entire genome or parts of the genome, such as the exome. The term can be widely used and also describes both germline or purely somatic genome data. On the other hand, the term "genetic" is often used synonymously with "hereditary" and is thus limited to germline analyses. Since raw data may affect both somatic and germline data, this position paper uses the more general term "genomic".

---

49 See here Lipp, in: Laufs/Katzenmeier/Lipp, Arztrecht, 7. 2015 edition, XIII. Paragraph. 14.

50 Ehling/Vogeler, MedR 2008, 273; Bender, MedR 2005, 511; Lipp, in: Laufs/Katzenmeier/Lipp (Fn. 51) XIII. Paragraph. 41.

51 Taupitz, Jochen. Biomedizinische Forschung zwischen Freiheit und Verantwortung (2002): p. 42.; Ebd. "Schutzmechanismen zugunsten des Probanden und Patienten in der klinischen Forschung." *Forschung am Menschen*. Springer, Berlin, Heidelberg, 1999. p.13-32.

2.3.5. The term “release”

“The term release” is used here and in the following to distinguish this type of interaction from the similar terms “return” and “sharing”. The term “release” refers to the provision of a copy of the raw data if requested, while the original form of this data remains at the institution. „Return“ is oftentimes used in the context of results (not raw data) and therefore transports more the meaning of a diagnostic setting. The term “sharing”, on the other hand, implies the practice of sharing data for research and making it available to other investigators.

2.4. Scope of the statement

The following recommendations have been developed with a view towards the release of raw data to self-determined or mature persons of legal age. A transfer of the recommendations to self-determined and mature minors (young people) should be possible in principle. The position paper including its recommendations is expressly not intended for all other situations. This applies, in particular, to the release of raw data

(1) in all forms of prenatal human life to parents or third parties,
(2) of non-self-determining minors (children) to themselves, parents, or third parties,
(3) of non-self-determining adults to themselves, parents, or third parties
(4) of deceased persons to relatives or other third parties.[52]

The explicit limitation of the scope of this position paper to the standard case of releasing data to self-determining or mature adults is done with the understanding that the release of raw data in other constellations still has special specific aspects and other potential consequences.

[52] In such cases, GDPR does not apply. It remains to be clarified from a legal perspective to what extent such raw data can form part of a (so-called) “digital heritage”. In the doctor-patient relationship, there is a clear legal order of entitlement to access the patient data in which the deceased’s spouse comes first, followed by his or her children. An assessment must be made between the right of the deceased to keep his/her data under lock and key and the potential benefits for the health of the relatives.

3. LEGAL FRAMEWORK AND CONDITIONS

3.1. Rights of patients and study participants to information and the release of raw genomic data

The legal analysis differentiates between the rights to be informed and the rights to the release of raw data as granted to patients and study participants based on the different regulations. This is necessary as the clinically relevant results of the raw datafrom the research as they relate to treatment can also be included in medical records or hospital information systems, while data from a purely research context can be regularly included in external research databases.[53] However, since these are raw genomic data, any restrictions laid out in GenDG are first addressed before the patients' right to inspection as pursuant to Section 630g BGB and the general data protection provisions under Article 15 GDPR. Furthermore, the relationship of these statutes are taken into consideration.[54] In practice, the inspection of and/or information on patient data is usually governed by civil law, and this is supplemented here by the overarching regulations of the GDPR, which have been in effect since May 25, 2018.

3.2. Reporting the results of genetic examinations, Section 11 GenDG

Based on the premise of "genetic exceptionalism" adopted by law, it is specified precisely in Section 11 (1) GenDG how the results of genetic studies must be reported. However, GenDG only applies with regard to genetic characteristics inherited or acquired during fertilization or until birth. Conversely, genetic studies of genetic changes acquired after birth are not included.[55] Therefore, regardless of whether or not there are genetic tests, somatic genetic changes are governed by general medical law. Hereditary information that is not of human origin (e.g. HIV) is also not addressed by

[53] Fleischer, Henrike, "Rechtliche Aspekte der Systemmedizin: Der Umgang mit Gesundheitsdaten und -informationen in der Big Data-basierten Medizin unter besonderer Berücksichtigung des Gendiagnostikgesetzes." Vol. 18. *LIT Verlag Münster*, 2018.

[54] The special rights to information for patients, which also exist at national level, are not addressed in the following. For example, the right of information vis-à-vis those responsible for hospitals with ecclesiastical institutions, Section 4 No. 9, 17 KDG or Section 47 (3) of the Act Governing Regional Hospitals in Baden Württemberg (LKHG BW), which regulates the rights to access of relatives and visitors of the patient (but not the patient himself) to hospitals that are supported by public funds.

[55] Fleischer, Henrike, et al., *MedR* (2016), 34: 481-491.

GenDG. These are the consequences of infections and transmission of retroviruses, which includes the integration of viral genetic information into the human DNA. The interaction of the virus with DNA does not entail that this is human genetic data. As a result, data on tumor-specific mutations do not fall within the scope of GenDG, insofar as the study is limited to purely somatic genetic changes in the cells and not additionally to parts of the genome of a healthy body cell.[56]

Pursuant to Section 11 (1) and (2) GenDG, the result of the examination may only be communicated to the person concerned and only by the responsible medical doctor (i.e., the person who ordered the genetic examination). In the wide-ranging field of genetic testing, this medical doctor reservation is intended to serve the patient's right to self-determination as well as quality assurance.[57] The patient should only be confronted with a genetic finding within the scope of a doctor-patient relationship. However, according to the provision clearly laid out in Section 3 (1) No. 1 GenDG, such genetic examinations are always specific and require an analysis of the data. Such an analysis is not carried out with raw data. Consequently, the provisions of GenDG are not relevant to the question of releasing raw genomic data.[58] In particular, no further interpretation is necessary, since the release of the raw data does not lead to an immediate confrontation with a genetic finding due to the lack of analyses.[59]

3.3. Patients' right to inspection pursuant to Section 630g German Civil Code (BGB)

A right to inspection within the context of medical treatment (i.e., only for patients and not for study participants) results from the provisions laid out in Section 630g (1) of the German Civil Code (BGB) as the primary and special right to inspection of medical records.[60] This right is not exhausted in the right to inspection of medical records but rather extends to all patient data stored in the appropriate medical documentation systems (such as the hospital information system and PACS).[61] In addition, Section 630g (2)

---

56 Ebd.
57 Ebd.
58 Ebd.: p. 484 et seq.
59 Fleischer, Henrike. Rechtliche Aspekte der Systemmedizin: Der Umgang mit Gesundheitsdaten und -informationen in der Big Data-basierten Medizin unter besonderer Berücksichtigung des Gendiagnostikgesetzes. Vol. 18. *LIT Verlag Münster*, (2018): p. 248.
60 Fleischer, Henrike, et al., *MedR* (2016), 34: 481-491.
61 Ebd.: p. 485.

Sentence 1 BGB provides for the patient's right to receive a duplicate of the medical records. According to Sentence 2 of that provision, the costs incurred for this must be borne by the patient. The wording of Section 630g (2) Sentence 1 BGB expressly also provides for the right to receive duplicates in electronic form.[62] In principle, it is assumed that the inspection by and transfer of the duplicate to the patient must be done on site (see reference in Section 630g (1) Sentence 3 BGB to Section 811 BGB), which precludes a right to consignation.[63]

According to Section 630g (1) Sentence 1 BGB, the right to inspection can only be refused if there are *considerable* therapeutic grounds to the contrary, i.e., in particular if it is to be feared that the patient is seriously endangering him/herself (e.g., in the event of a risk of suicide)[64] or that personal information about third parties has been included in the information (for example, the explanatory memorandum to the legislation mentions the treatment of a minor child, which is carried out with the involvement of its parents).[65] However, since the raw genomic data will have not yet been analyzed, this restriction (as well as the immediate applicability of the limitations of GenDG) is excluded from the case of interest at issue here. Only an analysis of the raw data can lead to a finding that can give rise to considerable therapeutic concerns. The information on blood relatives necessarily contained in genome data does also not result in the applicability of the restriction in accordance with Section 630g (1) Sentence 1 BGB. It is true that GenDG is not applicable with regard to the raw data, since no genetic examination or analysis will have been carried out, Section 2 GenDG. However, the law indicates that, despite the necessary third-party reference to genetic data, it refers solely to the responsible medical person (see Section 8 GenDG), i.e., it does not include blood relatives. In this regard, it would also be contradictory, for example, to require consent only from the responsible medical person but then to refuse access on the basis of a third-party reference, particularly as no conclusions can be drawn in a specific case from raw data that have not yet been analyzed.

---

62 Rehborn/Gescher, in: Erman, BGB Kommentar, 15th ed. 2017, Section 630g BGB Margin note 16; it is disputed whether the right to electronic duplicate applies only if the original file is already available electronically or if a right to digitization of parts of the file can be derived from Section 630g (2) Sentence 1 (BGB), as such, based on the wording, Lafontaine correctly comments in: jurisPK-BGB, 8th ed. 2017, Section 630g BGB, Margin note 115 et seq.; 119 with further references; see also: Walter/Strobl, MedR (2018), 472-476); Wagner, in: MüKo, Commentary on BGB, 7th ed. 2016, Section 630g BGB Margin note 20, which in this respect is not convincing in its desire to assign the patient's right to choose between physical and electronic duplicate.

63 See. Bayer, Ärztliche Dokumentationspflicht und Einsichtsrecht in Patientenakten, p. 192 et seq.; another rejection of the right to consignation of copies is found here: Frankfurt Higher Regional Court from May 9, 2011 - 8 W 20/11, GesR 2011, 672 et seq.

64 Kensy, *MedR* 2013, 31 (12): 767-772.

65 BT-Drs. 17/10488, S. 27; Kensy, *MedR* 2013, 31 (12): 767-772.

3.4. Provisions of Article 15 GDPR

A general right of access (i.e., irrespective of whether treatment is being provided) is laid out by Article 15 (1) Half-sentence 2, Version 1 GDPR. The aim of this right of access is to enable the data subject to have a precise and comprehensible overview of the data available to the controller in accordance with Article 12 (1) Sentence 1 GDPR.[66] A right of access to originals is not granted by Article 15 (1) GDPR.[67]

Article 15 (3) Sentence 1 GDPR provides for the provision of a free copy of the data stored by the controller.[68] In principle, the data should be made available for printing by means of remote access in the sense of online retrieval.[69] This fact is expressed by law in Recital 63 (4) GDPR.[70] Otherwise, Article 15 (1) GDPR assumes a request from the data subject and can therefore also be described as a "passive obligation of transparency".[71] It is not a question of only passively allowing the production of a copy being made by the data subject at the place where the data are provided. Article 12 (1) Sentence 2 GDPR, which refers to "any communication under [...] Articles 15 to 22", assume "transmission" of the data.[72]

In accordance with Article 15 (3) Sentence 1 GDPR, insofar as no safe remote access has been set up or this is not accepted by data subjects, the first copy should be provided at no cost on a suitable and secure storage medium (USB stick or hard drive), given the large amounts of data.[73]

---

66 Bäcker, in: Kühling/Buchner, DSGVO/BDSG Kommentar, 2nd ed. 2018, Article 15 GDPR Margin note 32.

67 Stollhoff, in: Auernhammer, DSGVO/BDSG Kommentar, 6th ed. 2018, Article 15 GDPR Margin note 29; Franck, in: Gola, DSGVO Kommentar, 2nd ed. 2018, Article 15 GDPR Margin note 30; see also Schaffland/Holthaus, in: Schaffland/Wiltfang, DSGVO/BDSG, published 11/2018, Article 15 GDPR Margin note 59, which presumably fails to take into consideration the fact that Article 15 (3) GDPR does not contain any provision comparable to Section 34 (9) Sentence 1 et seq. of the German Data Protection Act (BDSG).

68 In part, the claim laid out in Article 15 (3) Sentence 1 GDPR is understood as a special form of information, according to Schmidt-Wudy, in: BeckOK Datenschutzrecht, 25th Edition 2018, Article 15 GDPR Margin note 87.3 or as a regulation on the form of information due in accordance with (1), Paal states in: Paal/Pauly, DSGVO/BDSG, 2nd Ed. 2018, Article 15 GDPR Margin note 33; Kamlah, in: Plath, DSGVO/BDSG, 3rd ed. 2018, Article 15 GDPR, Margin note 16; in addition, (3) is also seen as an addition to (1), whereby (1) may also contain descriptions of the data, according to Bäcker, in: Kühling/Buchner DS-GVO Art. 15 GDPR Margin note 39 et seq.

69 Fleischer, Henrike. "Rechtliche Aspekte der Systemmedizin: Der Umgang mit Gesundheitsdaten und -informationen in der Big Data-basierten Medizin unter besonderer Berücksichtigung des Gendiagnostikgesetzes." Vol. 18. *LIT Verlag Münster*, 2018: p. 237.

70 See Ehmann, in: Ehmann/Selmayr DSGVO Kommentar, 2nd ed. 2018, Article 15 GDPR Margin note 15; Schantz, in: Schantz/Wolff, *Das neue Datenschutzrecht*, Chap. F Margin note 1193.

71 Also Veil, in: Gierschmann/Schlender/Stentzel/Veil, GDPR, 2018, Article 15 GDPR, Margin note 6, Article 15, Maring note 19; Contrary to Article 15 GDPR, the rights of the data subject under Articles 13 and 14 GDPR do not assume a request and could therefore be described as "active transparency obligations", see also Dix, in: Simitis, BDSG, 8th ed. 2014, Section 33 BDSG et seq. Margin note 3.

72 Walter/Strobl, MedR (2018), 36 (7): 472-477. Quaas, in: *BeckOK Datenschutzrecht*, 25th Edition 2017, Article 12 GDPR Margin note 28; Franck, in: Gola, DSGVO Kommentar, 2nd ed. 2018, Article 15 GDPR Margin note 28.

73 See Franck, in: Gola, DSGVO Kommentar, 2nd Ed. 2018, Article 15 GDPR Margin note 28; *Quaas*, in: BeckOK Datenschutzrecht, 25th Edition 2017, Article 12 GDPR Margin note 28.

Article 89 (2) GDPR provides for the possibility of flexibility clauses in national law which allow exemptions from the rights of data subject (such as the right of access) within the context of scientific research, provided these rights seriously impede or render impossible the specific purpose and the exceptions are necessary for the achievement of the purposes of the research. German Federal Law made use of this flexibility clause in Section 27 (2) BDSG (new).[74] With regard to any possibility of restricting the right to information or release of raw data, the same regulatory content as in the previous legal situation (until the time GDPR takes effect) must be assumed.[75] The release of a copy of the raw data can only be refused if enforcement of this right renders the research project concerned impossible or leads to a serious impairment of the research project concerned.

In addition, Article 20 GDPR provides for a right to data portability. Accordingly, data subjects may, by consent or on the basis of a contract (Article 6 (1) lit. b GDPR), obtain the personal data which they made available to a controller in a structured, commonly used, and machine-readable format and transmit these data to another controller (Article 20 (1) GDPR) or have it transmitted directly by the controller (Article 20 (2) GDPR). The patients or study participants have indeed provided the material for obtaining the raw data but not the raw data themselves. These data were created only by sequencing. Therefore, even with the broadest interpretation (of Article 29 of the German Data Protection Working Party),[76] the raw data are not data provided by the data subjects themselves but rather derived data generated by the controllers.[77] In this respect, the right to data portability does not apply to the case of interest at issue here.

3.5. Relationship between civil and data protection regulations

Beyond the doctor-patient relationship, only the right to information and access as regards data protection (including the free provision of a copy)

---

74 Fleischer, "Rechtliche Aspekte der Systemmedizin", 2018, p. 247 et seq. on the restrictive interpretation of Section 27 (2) BDSG, which is required by European law.

75 See Fleischer et al., *MedR* (2016), 34: 481-491.

76 Article 29 German Data Protection Working Party, Guidelines on the right to data portability, WP 242 rev.01 (Version: April 5, 2017), p. 11, available at https://www.bfdi.bund.de/SharedDocs/Publikationen/DokumenteArt29Gruppe_EDSA/Guidelines/WP242_DataPortabilityDE.html; if dealing with "raw data", these refer to data that are processed directly by meters (e.g. trackers) resulting from the user's observations. This should not be confused with the concept of "raw data" used here.

77 This applies even more in a narrower interpretation, see Strubel, ZD 2017, 355, 360; Jülicher, Medizininformationsrecht, p. 127 et seq., which argues in favor of the applicability of Article 20 within the context of agreements on treatment but does not elaborate on the restrictive requirement for the release of data; according to Fleischer, Rechtliche Aspekte der Systemmedizin, 2018, p. 238, "only those data that have been generated for treatment purposes and which are used for research purposes with the consent of the data subject (such as the sequence data of the responsible medical person, which in the context of treatment are not further analyzed, and are not currently stored in the medical records or in the hospital information system), although it is not clear to what extent such "generated" data should have been "made available" by the responsible medical persons themselves."

under Article 15 GDPR applies to the study participants. Within the context of a doctor-patient relationship, the relationship between a civil claim under Section 630g BGB and the rights of the data subject according to data protection law must be clarified in accordance with Article 15 GDPR.

In principle, EU law takes precedent over any conflicting national law.[78] Nevertheless, the national standard, including regulations and obligations, can complement the GDPR. In such cases, national law and EU law mutually apply. In the event of an imminent conflict of legal statutes, the possibility of interpreting national law in conformity with EU law must also be examined.[79] In addition, derogations may be permitted under a flexibility clause laid out in Article 23 GDPR.

As the above has shown, there are only partial differences between the (national) civil and (European) data protection legislation. For example, the refusal of a duty to consignation as pursuant to Section 630g (2) Sentence 1 BGB stands in contrast to Article 15 (3) GDPR. However, in order to preserve national legislation, an interpretation in conformity with EU law is possible in the light of the similarly regulated Article 15(3) GDPR, since the wording of Section 630g (2) Sentence 1 BGB also allows an interpretation towards consignation.[80] However, regarding the provision laid out in Section 630g (2) BGB on the obligation to pay costs, it is not possible to interpret this provision in conformity with EU law.[81] This is contradicted both by the clear wording of the standard and by the intention of the law.[82] An interpretation contra legem (contrary to the clear intention of the law) is excluded.[83] The application of a flexibility clause under Article 23 GDPR is also more likely to be rejected.[84]

Whether the primacy of application of Article 15 GDPR will apply in this regard can be left open in view of the current foreseeable developments. In the future, there will be a (national) claim by patients to make available the data collected from the medical person in an electronic health record or in the electronic medical record pursuant to Section 291a (5) Sentence

---

78 Permanent Rspr, the ECJ and BVerfG with different grounds instead of many decisions by ECJ from 7/15/1964 – Rs 6/64 (Costa/E.N.E.L.), BeckRS 1964, 105086; ECJ from 3/9/1978 – Rs 106/77 (Simmenthal), NJW 1978, 1741; BVerfG from 10/12/1993 - 2 BvR 2134, 2159/92, BVerfGE 89, 155 (190) (Maastricht).

79 Breyer, ZD 2018, 302 (302 et seq.) for the ratio of conflicting provisions between the German Telemedia Act and GDPR.

80 See the recitals in Bayer, Ärztliche Dokumentationspflicht und Einsichtsrecht in Patientenakten, p. 192 et seq.

81 However, see also Rybak, talk given at gevko Symposium in Berlin, 9/14/2016, p. 4 et seq., https://www.gevko.de/de/symposium/2016/2_Tag_Komplett.pdf, retrieved on: 11/8/2018.

82 BT-Drs. 17/10488, p. 27; Walter/Strobl, *MedR* 2018, 477.

83 See Wißmann, in: Erfurter Kommentar zum Arbeitsrecht, 19th ed. 2019, Preamble to AEUV Margin note 37; BVerfG of 12/10/2014 - 2 BvR 1549/07, NZA 2015, 375 (378).

84 See in detail Walter/Strobl, *MedR* 2018, 472, (476).

9 German Social Code Book V (SGB V). Since patients will also be granted access to their own electronic medical record beginning in 2021, the government's draft of the Appointment Service and Supply Act provides for the future merging of electronic health record and electronic medical record (Section 291a (3) Sentence 1 No. 4 SGB V).[85] The future electronic medical record will consist of server files from copied original documents and be based on the (documentation) record as administered by doctors.[86] In doing so, patients' rights (under social law) to gain access to these data, which pursuant to Section 291a (3) Sentence 1 No. 5 in conjunction with Section 5 Sentence 9 and Section 4 Sentence 2 SGB V does not comply with an obligation according to social law, will not be subject to a fee; this stands in contrast to Section 630g (2) Sentence 2 BGB.[87] It can be assumed that this innovation (under social law) will defuse the question of (remote) access law under data protection law, as already provided for in Recital 63 GDPR.

3.6. Summary

Study participants and patients are entitled to have their raw genomic data released to them. This claim is excluded only if the information or release leads to a serious impairment of the research project, which ultimately jeopardizes the implementation of the entire project.

Article 15 (3) GDPR grants a right to transfer (see Article 12 (1) GDPR) of a copy of the data stored by the responsible body at no cost (!). A final decision on whether, in the context of the treatment, the data protection claim under Article 15 GDPR supersedes the civil claim under Section 630g BGB in this respect must not be made here, since the previously foreseeable regulations for the granting of access to the data in the electronic medical record will also be free of charge for the patients.

However, as long as direct remote access is not possible, it will be necessary for study participants and patients to bear some of the costs in the

[85] Draft by the Cabinet of the German Federal Government, p. 164; BT-Drs. 19/3528, p. 5; see also Scholz, in: BeckOK Sozialrecht, 50th edition 2018, Section 291a SGB V, 3c.

[86] See "Telematik-Kuddelmuddel", E-Health-COM, June 13, 2018, https://e-health-com.de/details-news/telematik-kuddelmuddel/f87bfbf1f9b203cd8526a7776529a8b5/, retrieved on: 11/6/2018; "Patientenakte: Mobiler Zugriff soll zügig gesetzlich geregelt werden", aerzteblatt.de, June 21, 2018, https://www.aerzteblatt.de/nachrichten/95976/Patientenakte-Mobiler-Zugriff-soll-zuegig-gesetzlich-geregelt-werden, retrieved on: 11/6/2018.

[87] Bayer, Thomas. Ärztliche Dokumentationspflicht und Einsichtsrecht in Patientenakten: Eine Untersuchung zu den §§ 630f und 630g BGB mit Bezügen zum nationalen sowie europäischen Datenschutzrecht. Springer-Verlag, 2018. p. 200.

case of the necessary purchase of physical storage media, such as hard drives or USB sticks as pursuant to Article 12 (5) Sentence 1 and Article 15 (3) Sentence 1 GDPR if they wish to continue using the storage media.

4. ETHICAL ASPECTS

4.1. Preliminary considerations

The question of the release of raw genomic data to study participants and patients requires a careful ethical analysis, which must take into consideration the generally high sensitivity and complexity of genomic data, the importance of the issue of releasing raw data for the various persons and parties concerned, as well as the practical challenges.[88]

Genomic data are sensitive personal data that contain potentially significant information about the data subjects (and their biological relatives). Therefore the handling of genomic data is closely linked to the data subject's right to privacy or the informational self-determination. However, the issue of the provision of raw genomic data to study participants[89] also concerns other persons and parties: the researchers and physicians who were responsible for producing the raw data and to whom a request for release is directed, the institutions concerned, or the relatives of the study participants. Therefore an appropriate ethical analysis and evaluation of the issue of making raw data accessible to study participants must also consider the persons and parties whose interests and rights may be affected by access to and provision of raw data.

The ethical-normative basis for analysis and evaluation as well as for the development of the recommendation on the release of raw genomic data is a position that can be described as "caring liberalism". This position builds on the classical liberal tradition of political philosophy and legal philosophy (John Locke, Immanuel Kant, Wilhelm von Humboldt, John Stuart Mill) by giving priority to respecting the freedoms of the individual. Equal respect for individuals means, above all, giving each person a system of freedoms that allows them to lead their lives autonomously and individually according to their own ideas of a good and happy life, while respecting the same freedoms of all others. In addition to the classical position of liberalism, caring liberalism explicitly acknowledges that respect for people requires more than just renouncing interference in the (formally guaranteed) freedoms

88 Schickhardt, Christoph et al., "Do patients and research subjects have a right to receive their genomic raw data? An ethical and legal analysis" *BMC Medical Ethics* (2020).

89 Whenever this statement refers to "study participants", this reference includes patients and study participants. For the sake of easier readability, the term "study participants" is used more frequently in this statement. In passages where a distinction between the terms is necessary for content or legal reasons, these are listed separately.

of individuals; caring liberalism requires that individuals be actively supported, e.g. with regard to necessary means and abilities, in the use of their formal rights, so that they are able to live according to their own ideas of a good life.[90] Humans are not purely rational and self-sufficient beings. When it comes to interpreting themselves, their needs, and their ideas of a good or happy life, they depend on interacting and communicating with other people. In order to thrive and to exercise their freedoms, they depend on specific means and abilities to be made available by others. This is particularly true in the field of medicine and biomedical research. Study participants and patients must be considered in their various dimensions, including their vulnerabilities, fears, and hopes as well as their dependencies on others. Patients are particularly vulnerable: they have a potentially serious illness and find themselves in an asymmetric relationship with their physician in terms of dependencies, needs and competences. Study participants also find themselves in a de facto unequal relationship with researchers in terms of biomedical and technical knowledge. In these unequal circumstances, patients and study participants are particularly dependent on being protected from possible harms and burdens as well as on being supported in the individual exercise of their self-determination (empowerment).

4.2. The right to informational self-determination and the release of raw genomic data

Raw genomic data belong with a clarity to personal data like few other form of data, since each person's genome is unique and therefore inherently identifiable. Raw genomic data therefore fall within the scope of the person's informational self-determination. The term (of the right to) informational self-determination is often used in a way similar to the concept of (the right to) privacy, so that the question arises as to how the two concepts relate to each other. In the following, we only use the term informational self-determination and consider it to include the right to privacy. With regard to data and the age of digitization, the concept of informational self-determination seems more suitable, as it goes beyond merely talking about protecting privacy and instead emphasizes that it is not only a right of defense, but also an individual's right to make decisions freely, actively, and continuously regarding his/her personal data. The right to informational self-determination grants a person the freedom to determine whether and, if so, by whom and how personal data are collected, used,

[90] Rawls, John0. "Eine Theorie der Gerechtigkeit." *Frankfurt am Main: Suhrkamp*, 1979. p. 126 et seq.

processed, or passed on by third parties. The fundamental value of informational self-determination can be seen, on the one hand, intrinsically in and of itself. According to this approach, it is per se good and valuable to informationally determine oneself in a free and competent way and, for example, to evade the observation of others. On the other hand, the value of informational self-determination can also be seen in its instrumental value for other things that we consider to be valuable and respectable in the context of liberal value systems and democracy, e.g. as a condition for autonomy, wellbeing, intimate or confidential social relationships, dignity, personal flourishing, or even civil liberties and equality.

There are several reasons for granting the release of raw data regarding the importance of informational self-determination in the data age. In general, providing a copy of personal data collected by others is a prerequisite for the competent and concrete exercise of the right to informational self-determination, since it is only in this way that one is really able to know which personal data are available to third parties. The right to a copy of personal data must therefore be seen as part of the right to informational self-determination. This also applies, in principle, to the release of raw genomic data – although the individual making the request may still not receive all the information that third parties have at their disposal. Through analyses and interpretations of the raw data, third parties may, for example, generate or have generated knowledge that is not necessarily also made available to the data subjects when their raw data are handed over to them. Furthermore, the raw data collected continue to belong to the study participants in an irrevocable and unchangeable sense, as they continue to be related to these individuals and thus have a very personal connection to them.[91] Not only are they traceable, but they may also contain predictive information that could be relevant to the way the bearers of the genome live their lives. Each person should be able to access and freely dispose of the data on their own. This does not imply that the data subject is the only person who legitimately possesses these data; for example, a person may have granted and continue to grant researchers certain rights of use of the data. As a last reason, the release of personal data is essential for reasons of equality and transparency. In general, the release of the raw data may not be adequate for fully compensating for the informational asymmetry between the data subjects themselves and the third parties who have access to the data. Nevertheless, it can contribute to reducing this asymmetry and strengthening equality and transparency.

[91] Kaye, Jane, et al. "Can I access my personal genome? The current legal position in the UK." *Medical Law Review* 22 (1) (2014): p. 64- 73.

4.3. Protecting and enabling study participants

Most lay persons have a rather low understanding of genetics in general and of raw genomic data in particular. Understanding the potential information of health relevance implied in raw data and identifying variants that pertain to diseases is complicated, resource consuming and requires highly specialized skills. Each human being has numerous genomic variants in their genome. According to the current state of knowledge, some variants are associated with an increased risk of disease. However, many of these variants are thought to be neutral and not associated with diseases or the susceptibility to diseases. Other variants are not yet well understood and of uncertain clinical significance. Even well-researched and well-understood variants usually just provide evidence for indications of a susceptibility to disease leading to a probabilistic evaluation of the probability of developing a disease. It should be noted that communication and discussion of risks to allow patients to assess these appropriately themselves and to develop coping strategies are general challenges in medicine.[92]

Given this complexity, there is ample cause for concern that study participants may take on erroneous assumptions about the nature and usability of raw genomic data, which could lead to disappointment in terms of expected benefits and even expose them to the risk of burdens or harm. For example, it may happen that the benefit expected by the study participant cannot be realized with the raw data or that this benefit is difficult to achieve. We have reasons to believe, for example, that study participants rely on erroneous assumptions if they, as lay persons, hope to achieve the goal of learning more about their own genetic predispositions to a disease by presenting the raw data to their family doctor or studying it themselves at home.[93]

Concerns that some study participants associate the raw data with unrealistic hopes or uses, while at the same time possibly creating potential social, psychological, and economic burdens and risks for them and their families appear, in principle, to be justified. It is now crucial for the ethical analysis and evaluation of the issue of the release of raw data to adequately deal with these legitimate concerns and to draw appropriate conclusions from them. For a balanced assessment, it is first important to recognize

[92] Wegwarth, O., and G. Gigerenzer "Risiken und Unsicherheiten richtig verstehen lernen: Risikokommunikation." *Deutsches Ärzteblatt* 108.9 (2011): p. 448-451.

[93] Such an approach, for example, was considered by a significant proportion of lay people who were asked the hypothetical question in an online survey about what they would do with genomic data. See Middleton, Anna, et al., *Journal of Medical Genetics* 52 8 (2015): p. 571-574.

not only one-sided risks and burdens but also the possible benefits. There are very obvious reasons why study participants would want access to their raw data: Above all, it is possible that, despite all ethical and legal concerns and the proven quality problems in the use of private-sector analysis tests (direct-to-consumer tests), the study participants will have a medically valuable indication of certain "actionable" disease dispositions.[94] Second, the use of sequencing technologies in clinical diagnostics that go beyond the testing of defined genes in gene panels is a relatively recent field with very dynamic growth in the knowledge and assessment of the significance of certain genomic variants and continuously optimized processes. Accordingly, variants are regularly reclassified.[95] To this end, the American College of Medical Genetics and Genomics and the Association for Molecular Pathology recently published a series of guidelines to identify variants as "pathogenic", "likely pathogenic", "uncertain significance", or "nonpathogenic".[96] However, using these recommendations is not mandatory; even when applying the standards, sequencing sites may come to different conclusions given the room for interpretation. Occasionally, variants initially classified as nonpathogenic are later considered risk factors for diseases once data on larger cohorts are grouped together.[97] For these reasons, it is understandable that study participants might want a complete list of their genetic variants in order to follow the evolving scientific knowledge on these variants. A third potential motive would be to share data within certain patients-driven initiatives, such as "patients like me", in order to independently investigate the diagnosis of their own disease.[98]

In addition to the aforementioned medical reasons, study participants may also seek access to their raw data for reasons of a different nature, e.g., for the purpose of genealogy, the search for relatives or similar genetic profiles, for educational or entertainment purposes or even just to keep their future options open.[99,100]

---

94 For details on published weaknesses with regard to the validity of the published findings, see Tandy-Connor, Stephany, et al. "False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care." *Genetics in Medicine* 20.12 (2018): p. 1515.

95 Walsh, Michael F., et al. "Genomic biomarkers for breast cancer risk." Novel Biomarkers in the Continuum of Breast Cancer. Springer, Cham, 2016. p. 1-32; Kalia Sarah S., et al., *Genetics in Medicine* 19 2 (2016): p. 249-255.

96 Richards, Sue, et al. "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." *Genetics in medicine* 17.5 (2015): 405.

97 Williams, Elizabeth, et al. "Diagnostic yield from reanalysis of whole exome sequencing data." Poster presented at: 2016 *ACMG Annual Clinical Genetics Meeting*. 2016.

98 Patients like me, URL: https://www.patientslikeme.com (retrieved on: 7/31/2019).

99 Middleton, Anna, et al. "Potential research participants support the return of raw sequence data." *Journal of medical genetics* 52.8 (2015): 571-574.

100 Sanderson, Saskia C., et al. "Psychological and behavioural impact of returning personal results from whole-genome sequencing: the HealthSeq project." *European Journal of Human Genetics* 25.3 (2017): p.280.

Just as one should not unilaterally take into consideration possible risks and burdens, one should not assume that all study participants who request the release of their raw data have lacking or severely flawed understanding of the nature and usability of raw genomic data. This would be an unjustified generalization. The mere possibility that some people may be disappointed or even burdened or harmed by the (limited) usability of the raw data due to a lack of competence and understanding is not supported by real world experience and does not seem to be a sufficient reason to deny study participants the right to the release of their raw data. Interpreting possible lack of competence and potential risks for individuals as grounds to deny individuals or all study participants their right to the release of their raw data goes against the very notion of caring liberalism. Instead, a potential lack of competence or understanding as well as possible erroneous assumptions and risks should be addressed by helping study participants who wish to have access to their raw data to ideally develop an adequate understanding of the nature and usability of their raw data; an appropriate basic understanding should enable them to adequately assess the opportunities and risks for them associated with the raw data in principle and their ability to appropriately use the raw data. To this end, it is essential to explore the motivation behind their requests in dialog with the study participants and, if necessary, to provide specific and pertaining information in a comprehensible manner. Based upon our approach of caring liberalism and focusing by now solely on study participants themselves, we draw the ethical conclusion that study participants have a right to the release of their raw genomic data and that they should be supported in the process surrounding their request by means of providing information, including optional individual discussion.[101]

Recognizing the moral right of study participants to have their genomic data released, including an informative offer of support, can now be challenged by the fact that it is in contrast or contradiction to the rights or interests of other persons or parties concerned. To examine this, the following section assesses how the release of raw data to study participants affects the issuing researchers, doctors, and research institutions as well as the participants' relatives, and whether there are reasons for rejecting or restricting the right to the release of raw data to study participants.

[101] The provision of information and personal discussion must not be confused with individualized genetic counseling.

4.4. Physicians and researchers: Claims and reasonableness in releasing data

4.4.1 Primary right of use for the sequencing body for research purposes

Physicians and researchers may be under the impression that the data they collect and produce for diagnostic or research purposes are owned by them. However, the patients or study participants only grant a right to use their data for purposes that correspond to the relationship with the physician (treatment) or the researchers (research benefits), which contradicts any general ownership claims by physicians or researchers. In these relationships, neither physicians nor researchers nor their institutions acquire a right to the exclusive and complete use of the data. However, the time and effort required to conduct the sequencing, which is currently still carried out within the scope of research projects, can give rise to a claim by "primary researchers" or those generating the data vis-à-vis third-party researchers that they have a primary right of use over all other researchers who want to conduct secondary research with these data[102]. Such a primary right of use, limited to a certain period of time, granted to researchers/persons generating the data to use the data for research purposes is relevant in view of the fact that released raw data could otherwise be used by potentially "competing" researchers for their own research purposes and publications without previous agreement with the data producers. This scenario is conceivable, on the one hand, if patients, on their own initiative, request access to the raw data in order to make it available to another research project. On the other hand, it is conceivable and has actually taken place in practice that other research groups directly approach the research participants and encourage them to request the sequenced data in order to then transfer them to the research group for scientific use. Such use and the subsequent publication of the raw data without the knowledge and consent of the primary researchers/persons generating the data would in both cases clearly violate the rules of "good scientific practice". This unspoken rule considers the right of use of the data to lie with those who generate and collect them.[103]

[102] A primary right of use is not referred to here in a legal sense; instead, we define it as the possibility for research institutions to be the first entity to use the raw genomic data for scientific purposes. The secondary use of these data for research purposes by other researchers should be coordinated with the institution that primarily generates the data.

[103] See Code of Conduct of the DFG Research Foundation, Guidelines for Safeguarding Good Scientific Practice, Guideline 10 (2019) – after approval by the Senate March 28, 2019, the DFG's Annual General Meeting in Rostock approved this guideline on July 3, 2019.

The affected researchers could also argue that the release of raw data contradicts their freedom of research. However, it should be stated that the core principles of the freedom of research (free choice of questions, topics, methods) is by no means violated by the release of a copy of raw data to study participants. The data can still be used freely for research purposes. For the physicians or researchers affected by a request from a study participant, as well as for their institutions, the release of raw data and the scope of information and consultations undoubtedly means an increase in time and effort and commitment to resources, but also an increased responsibility.

4.4.2. Time and effort

The costs associated with the responsible release of raw data (in terms of financial resources and workload, etc.) could, in fact, represent a valid argument for researchers against such a release, as it could prevent them from conducting their work if, for example, the additional time and effort would jeopardize the success of their research projects. Lunshof et al. argue that the costs of the data storage media continue to fall, but they do not take into account that the release should meet certain criteria, which requires a certain amount of human and logistical resources.[104,105] The time and effort required may be high for the first requests, but are expected to decrease over time as the availability of tools, models, and best practice recommendations increases.[106] Based on our experiences, we assume that relying on increasing routine in implementing our recommendations the technical steps, coordination, and communication will equate to approximately 2 working hours spent on these tasks.

Researchers usually have both the right and the duty to conduct research. If the time and effort involved in releasing the raw data seriously impairs their ability to conduct research, this would be grounds not to comply with requests for the release of raw data in specific cases. The burden of proof in this case lies with the researchers. They would have to demonstrate that the time and effort required are unreasonable for the concerned project. With an approximate time and effort of around 2 working hours, we can estimate that requests by singlestudy participants for the release of their raw data do not constitute a serious impairment to a research

[104] Middleton, Anna, et al. "Potential research participants support the return of raw sequence data." *Journal of medical genetics* 52.8 (2015): 571-574.
[105] Lunshof, Science 343 (6169) (2014): p. 373,373 et seq.
[106] Thorogood, Adrian, et al. "APPLaUD: access for patients and participants to individual level uninterpreted genomic data." *Human genomics* 12.1 (2018): p. 7.

project, are reasonable and thus to be respected. However, if the trend increases significantly and study participants regularly and in larger numbers seek the release of their raw data, the financing of the processing of such requests must be revisited. First and foremost, government bodies and funding agencies should fund the necessary infrastructure to support the release of raw data – for example, by way of quality-assured data platforms.[107]

4.5. Protecting the institutions

It is in the interest of researchers and research institutions to protect themselves in advance against potential, unjustified allegations and accusations in the case of negative consequences resulting from study participants' handling of their raw data. In order to clarify the ethical and legal responsibilities and prevent possible damage to their reputation, researchers or research institutions should oblige participants to sign a written confirmation ("receipt") in which the participants declare i) that the data have been made available to them on their own request, ii) that they were provided information on the general nature and possible implications of handling the raw data; iii) that they take responsibility for the consequences of their handling of the raw data both for themselves and third parties. In view of the significance of the interests of the entity releasing the data and the small cost to the study participant, we consider it justified to ensure that this brief statement by study participants becomes a compulsory condition for the release of raw data.

4.6. Consideration of relatives

A person's genomic data always also relate to the (biological) relatives of that person. Genetic first-degree relatives have about 50% of the same genome. Genomic data from study participants can therefore also contain information on disease dispositions of relatives or be used to identify relatives.[108] The raw genomic data of a person therefore undoubtedly also affect the informational self-determination of their genetic relatives. A special situation for study participants and their relatives may arise when study

[107] Shabani, Mahsa, Danya Vears, and Pascal Borry. "Raw genomic data: storage, access, and sharing." *Trends in Genetics* 34.1 (2018): p. 8-10.

[108] Guerrini, Christi J., et al. "Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique." *PLoS biology* 16.10 (2018): e2006906.

participants learn of an inheritable disease risk, most likely also affecting relatives, based on an analysis of their raw data. If the study participants share their own risk constellation with a relative, this may violate the relative's right not to know and may also entail uncertainties and burdens. It is also possible that the study participants do not adequately understand the genetic characteristics themselves or that they do not successfully communicate the genetic risk, so that the relatives may have an inappropriate understanding of the facts or feel unreasonably troubled. On the other hand, it is of course also possible for a relative to receive medically useful information. Against the backdrop of potential concerns and implications for relatives, the question arises how the concerns of relatives should be appropriately taken into account during the process of releasing raw data. For many reasons, it seems to us that individual information, informed consent or a right to veto on the part of all close relatives are not ethically justified nor practically feasible. However, it seems appropriate and necessary to explicitly inform the study participants of the sensitivity and potential implications of the data for the close relatives.

4.7. An outlook on society

The issue of releasing raw genomic data from the contexts of public clinical care or research points to further, fundamental ethical questions that are relevant to society as a whole. Speaking very roughly and simplifying, two fundamentally different developments are conceivable. On the one hand, citizens can monetize (the right of use of) their data from medical care and biomedical research in the future and potentially make a profit from this. They could obtain copies of these data (free of charge) to pursue personal financial interest and sell them (or the right to use them) to private companies. Both the individual citizen and the company could be motivated by personal profit, and there would be a risk that a change in the data would also shift the usefulness of the data from the publicly funded research sector to the private sector. An opposite effect to this scenario would be that citizens and public biomedical institutions consider the data from medical care and academic research as part of the common good and make them (under determined conditions) accessible to public institutions to benefit future patients and society as a whole. However, this latter approach to biomedical data with a view on the common good would require systemic changes in order to meet the legitimate interests of all stakeholders, e.g., the right of primary researchers/persons generating the data to a reliable and adequate scientific recognition of data production.

In addition, the free release of raw genomic data could lead to unfair competition, market distortion, and misplaced incentives for potential study participants. If a publicly funded biomedical institution passes on raw data to the responsible subjects free of charge, this could have a negative impact on private companies that offer customers the sequencing and analysis of their own genome as a commercial service. However, such an effect does not seem very realistic thus far, partly because private companies such as 23andMe[109] not only limit their service to sequencing the genome but also offer genetic analysis and interpretation which are explicitly not part of the release of raw data within the meaning of this position paper. However, concern about market distortions and unfair competition could be one reason why publicly funded genomic research projects should refrain from attracting research participants by offering and advertising the possibility of the release of raw data free of charge.

4.8. Summary

Study participants have a moral right to the release of their raw genomic data. They also have the right to be informed about general and elementary characteristics and implications of the raw data, including individual discussion (not genetic counseling) in order to gain a basic understanding as basis for their decisions as to whether they really want to have their raw data and how to use them. With regard to the effects of the release of raw data for third parties, certain rights and interests of third parties, which might be affected in some way by the release of raw data, have been identified. In general, however, there are not sufficient grounds to reject, restrict, or deny the right of study participants to the release of their raw data. The interests of concerned third parties nevertheless deserve to be taken seriously and considered responsibly.

Among other things, requiring study participants to sign a brief "receipt" when the data are released appears to be justified. Furthermore, it seems important to us to point out the primary right of use of the researchers who generate the raw data in a research context. Furthermore, study participants' awareness for potential concerns and implications for biological relatives should be raised by means of general information and optional individual discussion, where appropriate.

[109] 23 and me. URL: https://www.23andme.com/en-int/ (most recent retrieval: July 5, 2019).

5. RECOMMENDATIONS FOR THE RELEASE OF RAW GENOMIC DATA – SPECIFIC PROCEDURE AND CONSULTATION PROCESS

5.1. Description of a multilevel process for releasing raw data

5.1.1. Objectives of the release process

The multilevel release process has two objectives:

I. The primary purpose is to guarantee the basic right of the study participant to obtain a copy of their raw genomic data.[110]

II. On the other hand, appropriate consideration should be given to the interests of the study participant, other persons affected by the release of the data, and the institution involved. The concerns that should be protected include:

a) the **protection of the study participant from harm** and **the ability of the study participant** to use their own raw data in an autonomous manner that is of use to him/her;

b) the **interest of the institution in** protecting its reputation among study participants and the public as well as transparently communicating the responsibilities of study participants for possible negative effects and harm stemming from the receipt and use of raw data;

c) encouraging study participants to take the interests of third parties concerned into account, since the potential information stemming from the raw data could also affect the **informational interests of the (biological) relatives of the study participants**. Although there is no connec-

[110] In the interests of better readability, this statement uses the term "study participants". This refers to both patients and test subjects. If it is necessary to make a distinction, this will be specifically mentioned.

tion in treatment to any blood relatives, their interests and rights may be affected as a result of the handling and release of the raw genomic data, given that they are related to the study participants, e.g., if their germline sequence data is made accessible with a personal reference. This process could, for example, pose the specific risk of genetic discrimination;

d) the **interest of the research institution and the researchers** to limit the costs, time and effort involved in the release of the data;

e) the **interest of the researchers** in evaluating and publishing the raw data first or being asked if other researchers want to work with the data.

5.1.2. Initial consultation with the requesting study participant

The first step after receiving the request involves a face-to-face consultation between the custodian[111] or an authorized person and the study participant making the request. In this consultation, which can also be conducted by telephone, the authorized representative of the institution should provide initial information on the release procedure (see below) and may get a first impression of the level of competence and the motives of the person making the request.

In **the initial consultation**, the study participants making the request should receive information on the following aspects:

on the release procedure at the relevant institution:

a) the institution recognizes the right of study participants to the release of their raw data.
b) the institution provides a specific procedure for the release of the data which includes general written information and an offer for an in-depth consultation.
c) the study participants must not contribute to the costs of the first copy of the raw data. Insofar as no safe remote access has been set up, in accordance with Article 15 (3) Sentence 1 GDPR, the first copy should

[111] The role of the custodian: The identifying data of the study participants are usually kept strictly separate from the genomic data sets. Only authorized persons (data custodians) have access to the identifying data of the study participants via the sample identifier of the genomic data, a so-called pseudonymization number, and are thus able to associate all the de-identified records to a specific person (re-identification). This statement defines a custodian as a person/institution authorized to make a connection between the pseudonymization number of the study participant and the personal data and to enable the identification of the person.

be provided at no cost on a suitable and secure storage medium (e.g., a hard drive) given the large amounts of data.

d) before the data is published, the study participants are required to sign a statement in which they declare themselves responsible for their data after the release. This responsibility can only be transferred to the study participants for the copy of the raw data. The institution, of course, will still be responsible for the data remaining at the institution.

e) in the event that study participants wish to provide their raw data to other research projects, it will be determined that other researchers may only re-use the raw data after consulting the management of the issuing institution.

on the properties of the raw genomic data:

f) insofar as the raw data originate from a research context, they are often not created in full in certified conditions. It should therefore be clarified whether the data can be used for clinical diagnostics without further validation.

g) information that is not directly related to medical and health conditions can be derived from the raw data for lay persons but also for non-specialist doctors. However, precisely because it will be easier to analyze raw genomic data via evaluation programs offered commercially in future, it is all the more important to provide qualitative assurance of evaluations and interpretations.

h) in order to obtain information relevant to health conditions from the amount of raw data, further special analyses and interpretations, such as those carried out by experts and specialized laboratories and also by commercial suppliers, are necessary.

i) Analyses of the raw data can result in very sensitive information about diseases and hereditary systems of the study participants themselves as well as their close relatives and children. In particular, commercial providers of sequence analyses could also have their own interests in the data, which go beyond the desires and questions of the study participants. The study participants must be fully informed about the risk of misuse.

Finally, study participants should be offered the opportunity to first reflect on this information and to decide at a later stage how to proceed. In order to assist study participants in this process, they should be offered an informative written list that is easy to understand and lists the most important points; they should be able to receive this by email, fax, or letter (see 7.1).

At the end of the first step, the study participant has **two options**: He/she may withdraw the request if he/she realizes that the desire to have their raw data released was based on misconceptions of the raw data and their use. Second, he/she may request the release of the raw data. If they wish to proceed with the release, the following points will become relevant and written information material will be provided.

5.1.3. Written information material

The next step is to provide the study participants with written information on the general nature and implications of raw genomic data (see 7.1). It is important to note that this is an **offer of information**; it cannot be assumed that the study participants are required to review the information materials or confirm that they have read and understood the information.

In principle, the study participants have the right to obtain a copy of their individual raw data, even if they do not confirm that they have reviewed the information offered. However, from the point of view of this statement, it should be mandatory that the study participant sign the declaration on the provision of information regarding the release of the raw data (see 7.2).

On the general nature of the information to be obtained from the raw data as it relates to medical and health issues:

a) most information on genetic predispositions to a disease are by nature based on probabilities and only express the probabilities of the onset of disease at a later date.

b) for some genetic predispositions, there are preventive measures or treatment measures. For other predispositions, including serious or incurable diseases, no proven prevention or treatment measures are known.

c) in addition to the possible benefits for prevention or treatment, knowledge of the carrier of genetic predispositions to the disease can also pose a psychological and social burden.

d) in principle, the provisions of GenDG state that for genetic examinations for medical purposes the analysis and interpretation of genomic data and its relevance to health issues may only be carried out by medical specialists who rely on professional advice, e.g., from human geneticists or specialists with additional genetic qualifications.

On risks and threats related to data protection:

e) Genomic data inherently identify the carrier. Even if they are stored without a name or other personal data, they can be assigned to the carrier under certain circumstances. There is an increased risk of the study participant being unintentionally identified by third parties, for example, if study participants themselves or their close relatives have given their genomic data along with personal data in freely accessible places, e.g., in databases that are used for ancestry research or (other) social networks on the internet.

f) If the raw data were to land in the wrong hands, the possibility that they may be used by third parties – even illegally – to the detriment of the study participant or their relatives cannot be ruled out.

g) Medical findings generated by a subsequent analysis of the raw data for medical purposes (independently or with the help of third parties) may have to be disclosed to other bodies, e.g., before purchasing a (life) insurance policy. This applies to life and disability insurance and long-term pension schemes with benefits of more than EUR 300,000 or more than EUR 30,000 per year (Section 18 (1) Sentence 2 GenDG).

5.1.4. Individual consultation

In a third step, the study participant submitting the request is offered an individual consultation. The basis for the consultation will be the written form provided in step 2. This consultation should not be considered as a (medical) genetic counseling and must be clearly and explicitly distinguished from such a genetic counseling scenarium in a medical sense. The subject of the individual consultation are the nature and potential implications of raw data in general. It should be noted that the consultation does not cover the individual genetic characteristics of the study participant. It is therefore not necessary for the institution to have human geneticists or doctors with additional genetic training perform the consultation. The aim is to give the study participants the opportunity to ask individual questions and discuss aspects of particular interest to him/her. If the representative of the institution gets the impression from statements of the study participant that their ideas or plans regarding the raw data pose particular risks for the aforementioned concerns that should be protected as pertains to the study participants themselves, their relatives, or the institution, this will be expressly

addressed.[112] The consultation should be conducted in person and only in exceptional cases can it be carried out by telephone, in particular out of consideration for special difficulties for study participants, for example, due to an unreasonably long journey or a health-related weakness.

5.1.5. Written confirmation of the study participant

In the fourth step, the study participants are required to confirm the receipt of the raw data in writing before the raw data are released, and the following points are addressed:

a) with regard to step 1 of the procedure: the study participant was informed by telephone as to the essential aspects of the type of the raw data;

b) with regard to step 2 of the procedure: detailed written information has been received;

c) with regard to step 3 of the procedure: there was individual consultation, or the study participant was offered one;

d) it is understood that the raw data are sensitive data concerning both the study participants themselves and their biological relatives;

e) the study participant assumes responsibility for the handling of the raw data handed over to him/her and the consequences for him/her and third parties arising from his/her handling of the data.

5.1.6. Release of the raw data to study participants

In the fifth step, a copy of the requested raw data is safely released to the study participants by a person authorized in the role of custodian or by an institutional advisory board.[113]

If the study participant submits a request regarding the release of his/her raw data, this is to be regarded as implied consent to the restoration of the

[112] Even in circumstances in which representatives of the institution have reason to assume that, for example, a patient is unable to make an adequate assessment of the risks due to his/her current situation, the right of the study participant to have the research institution release their raw data remains intact. The representatives of the institution have no other options or rights than to express and clearly address their concerns and assessments in the consultation with the study participants and possibly document this process internally (as proof).

[113] In this sense, "safe" can be, for example, by way of personally releasing the raw data by an authorized person or by sending a password-protected data storage medium or via encrypted data transmission to the study participant, whereby the latter must identify him-/herself upon receipt of the shipment and the password must be sent in a separate mailing.

personal reference by a person entitled to do so, provided that these raw data have previously been pseudonymized.

5.2 Is it advisable to provide additional verification of the genetic identity of the study participants for correctly associating the raw data?

Within the context of the release of raw data to study participants, it is necessary to ensure that they also receive the correct raw genomic data that belongs to their person. This is also connected to the question as to whether the existing quality assurance measures (QA) are sufficient to avoid confusion or whether another identification verification between the sequenced data and study participants should take place before data release. Such a verification could be performed by analyzing a second sample of the participant either via a second sequencing analysis or a less expensive alternative, e.g., by means of array-based genotyping.

The question of the likelihood of an incorrect identification of clinical samples is ultimately a question of the effectiveness of the QA set up at each specific site and the avoidance of errors (confusion of samples). The pathological institutions have a long history of QA developments. There are process and function descriptions for each step and function as well as corresponding error management, which means that the rate of errors occurring is kept low (<1%). This could be considered as a benchmark/ standard for effective QA in terms of avoiding confusion of samples.

For the relatively new translational application of NGS technologies, standardized QA measures are being developed. Within the context of translational oncology in Heidelberg, QA measures are in place that immediately after the samples have been entered into the sequencing device help to effectively prevent confusion of samples: In the Heidelberg core sequencing facility of the German Cancer Research Center (DKFZ) and NCTs, for example, each sample to be sequenced is assigned a unique barcode (segments of a phage DNA) before any further processing in order to detect a possible mix up of samples. The hypothetical possibility of an error source after the analysis has been completed in which an incorrect sample would be incorrectly assigned to an already performed sequencing run is excluded, due to the uniqueness of the added sequenced barcode.

However, before the actual input of the sample into the sequencing devices, it is quite conceivable that a confusion of samples could occur: The actual

sample collection, which is often taken from the study participants in the form of tumor and control samples, could potentially be a source of error before the actual sequencing analysis takes place. These samples are rarely taken at the same time; the tumor tissue is extracted, for example, during an operation, while the control sample is taken from healthy tissue (e.g., blood) cells of the same patient, which usually takes place at a different time, e.g., during a hospital stay or during an outpatient visit. The risk of confusion would be significantly lower if the control and tumor samples were taken at the same time and place and the samples were labelled simultaneously. In fact, the separate time and location of sampling can be a potential source of error if samples are manually labeled incorrectly. Based on the experience in the Heidelberg oncological sequencing programs, this type of error occurs with an estimated frequency of less than 1%. We are not aware of validated surveys with these error rates.

In the subsequent step of processing, the actual sequencing analysis, thismix-up would be noticeable since the comparison of the sequences of tumor genome and germline/control genome would produce contradictory and conspicuous results (e.g., regarding the total number of variants between tumor and germline sample). These inconsistencies would, in turn, initiate clarification measures (e.g., resampling and repetition of the analyses).

As explained above, mix ups do occur, but the quality assurance measures implemented in this example would effectively prevent an incorrect assignment of raw genomic data to the respective study participants.

However, a confusion of samples could not be ruled out if both the control and tumor samples were labelled incorrectly at the time of acceptance of the sample material. Then, there would be no inconsistency between the samples noticeable in the subsequent sequencing analysis and the mistake would not be detected.

Compared to pathological institutions, the translational applications of NGS technologies clinical applications are a relatively recent development. It is therefore recommended that questions concerning quality assurance are dealt within the context at the respective sequencing sites including all project participants. Parameters such as the frequency of incoming requests and frequencies of detected sample confusions, should be documented and lead to specific, appropriate adaptations of existing quality assurance measures on site. Overall, the risk of incorrectly associating raw genomic

data within research contexts known to us is very low (<1%) and does not justify additional sampling with the corresponding additional financial and logistical burdens of an additional genetic analysis as confirmation.

If study participants seek a secondary opinion of their own raw data, then the responsibility for the use of the raw data and thus also for the validation of sequencing and identity matching should also lie with the study participant once the raw data has been handed over to them.

6. SHORT VERSION – PROCEDURE FOR THE RELEASE OF RAW GENOMIC DATA (PRACTICAL IMPLEMENTATION OF THE RECOMMENDATIONS)

6.1. Objective

This guideline proposes a uniform approach to releasing raw genomic data to study participants.

6.2. Scope

The guideline has been drawn up with a view to releasing raw genomic data to adult, consenting study participants or patients. In the case of enquiries that call for the release of raw data to parents of children, to persons unable to give consent, or even to relatives of deceased persons, further consideration is needed, as other legal and ethical framework conditions apply.

6.3. Responsibilities and tasks

Site-specific planning of the process of releasing should determine how and from whom the raw genomic data is transmitted to the respective study participants.

Table 2: Responsibilities and tasks

Responsibilities	Tasks
Role of custodian	1. **The custodian** may delegate activities to knowledgeable, qualified persons responsible for the project but retains overall responsibility. 2. **Protecting the identity of study participants:** The custodian protects the identity of the study participants by securely managing the pseudonymization code that links the sequenced data with the identifying personal data (e.g., medical record, name). 3. **Decoding** is therefore only permitted by the (few) custodians and the person directly entrusted with this activity within the project. 4. **Transfer function:** A direct transfer of the copied raw data is done either by the custodian or an internal member of project who has been commissioned by the custodian.
Person responsible for the project	1. Performs activities delegated to him/her by the custodian: · **Release/provision of the information offered:** Information material and consultation · **Obtaining written confirmation:** Clarification and confirmation of responsibility by the study participant. · **Release of the copy of the raw data** to the study participants.
Data archive	1. **Storage of pseudonymized raw data** on a suitable medium. 2. **Releasing of the copy of the raw data to the custodian.**

6.4. Prerequisite – Personnel

6.4.1. Custodian

The custodian (e.g., principal investigator) bears the **overall responsibility** for the procedure. Since the custodian is responsible for the correct assignment and de-pseudonymization, the principal investigator him-/herself or a person from the study team is eligible for this position. The individual activities within the release procedure may be delegated via the custodian to knowledgeable, qualified persons responsible for the project.

6.4.2. Person responsible for the project

The person responsible for the project should be familiar with the research project or the corresponding clinical program and should be aware of the importance and potential use of raw data and the legal framework of releasing the raw data.[114] The person must not necessarily have medical training.

[114] National Ethics Council. Biobanks for research: Statement. National Ethics Council, 2004: p. 68 f; Morr, Ulrike. Zulässigkeit von Biobanken aus verfassungsrechtlicher Sicht, 2005: p. 143; Söns, Udo. Biobanken im Spannungsfeld von Persönlichkeitsrecht und Forschungsfreiheit: eine Gefahr für Selbstbestimmungsrecht und Datenschutz?, 2008: p. 162; Antonov, Katrin. Der rechtliche Rahmen der Zulässigkeit für Biobanken zu Forschungszwecken, 2006: p. 205; Damm, Reinhard. "Gesetzgebungsprojekt Gentestgesetz—Regelungsprinzipien und Regelungsmaterien." *MedR* Medizinrecht 22.1 (2004): p. 1-19.

Furthermore, the person should be able to communicate with lay people in a comprehensible manner regarding complicated medical concepts.

Since the consultations are not to be confused with genetic counselings by trained human geneticists, it is not necessary for trained human geneticists or doctors with additional genetic training to conduct them. Nevertheless, it is necessary for the information-transfering person to familiarize him-/herself with the topic of raw genomic data to the extent that he/she can provide information about the possibilities and risks associated.

6.5. Procedure steps of releasing raw data

6.5.1. Who should be allowed to receive the raw genomic data?

The individual raw genomic data will be handed over to the person whose genome has been sequenced and who has confirmed in writing that they have received information on the properties of raw genomic data, the risks that could be associated with using raw genomic data and that they have received the copied data.

6.5.2. What data/data formats should be provided?

In principle, the consultations should also address the objectives and ideas of the study participants in order to determine the appropriate data format. A BAM file format is generally recommended for creating a complete copy of the raw data, as this requires less space than the FASTQ format and a reverse transformation is possible if necessary (see Table 1: Overview of the size and properties of files of the initial sequencing steps from whole genome analysis).[115]

6.5.3. Implementation

The recommended multilevel procedure is clearly summarized in Table 3. The procedure includes both an initial informative consultation, usually by telephone, and a second, face-to-face consultation. The interviews must be certified by a written declaration (see 7.2) on the side of the study participant, since by the signature provided also transfer responsibility for the future handling and use of the raw data.

[115] File formats commonly used for raw genomic data are FASTQ, BAM, CRAM, or VCF files. VCF files with non-interpreted variants are the latest stage of bioinformatic processing which by definition is still assigned to the term "raw data" (see Definitions in Chapter 2.2.1).

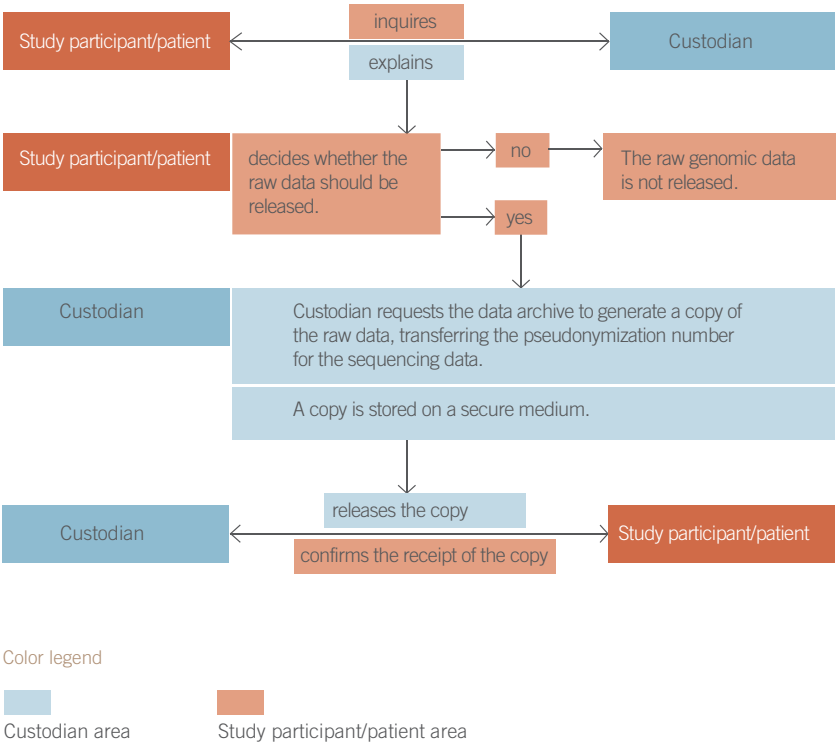
6.5.4. Short summary that serves as the basis for a standard operating procedure (SOP)

Table 3: Short summary

Initial interview (by phone or in person)	The study participant/patient is informed that the right to be handed over their raw data is recognized in principle and that there is an institutionally established procedure for the release of the raw genomic data. **Description of the procedure:** - there are no costs for the study participants for creating the first copy of the data. - the offer of information not only includes a short initial interview but also information material and a second consultation. - the general characteristics, risks, and opportunities of raw genomic data are addressed. - both the offer of information on the possible risks and the transfer of responsibility for using the raw data should be confirmed in writing by the person making the request. - the release of the copy of the data must be confirmed in writing.
Initial interview	**Important characteristics of raw genomic data:** - the data are obtained in a non-standardized research workflow that cannot normally be used for medical or diagnostic purposes without further validation. - In order to derive information from the raw genomic data, further analysis by experts is required. - These additional analyses may identify variants that are relevant to health issues. Before any further steps are taken, the relevant variants should be validated, and a prior counseling with human genetic training personnel should be carried out. - The analyses may reveal sensitive information about the study participants and their biological relatives. It should be pointed out that there are potential risks of misuse. - If the sequenced data comes from the research context, the first right of use should be granted to the person generating the data. The data should only be passed on to other research groups with the informed consent of the person generating the data.
Handing over of written information	Time allotted to read and think after the initial interview and the handing over of the information material (see 7.1).
Second interview (Personal consultation)	**Topics of the written information** Answers to the study participant's questions. With a view to selecting the appropriate format for the raw data, the motivation for the wish to have the raw data should be discussed.
Decision by the study participant	After having a period of time for reflection, the study participant decides whether the copy of the data should be handed over.
Written confirmation	If a direct releasing is desired, a written declaration on the transfer of responsibility for the future use of their own raw data is to be signed by the respective study participant (see 7.2).

Order send to data archive	The custodian (or a person commissioned by the custodian) places an order with the data archive for a copy of the raw data to be generated. The necessary pseudonymization number of the study participant is transmitted.
Creation of the copy (storage medium, e.g. hard drive)	The raw data are copied on a secure storage medium.
Handover of the copy of raw genomic data	The raw copy of the raw data will be transferred via a custodian, a person directly commissioned by the custodian, or an institutionally established advisory board and handed over to the study participant who will present proper verification of their identification.
Written confirmation	The study participant confirms in writing that the raw data copy has been released (see 7.2).

6.5.5. The chain of action for releasing raw genomic data



7. ANNEX

7.1. Template for an information leaflet on releasing raw genomic data

Raw genomic data - Answers to your questions

You expressed your wish to have your raw genomic data released to you. For this purpose you may have many questions about the properties and possible uses. The following information is intended to help you to understand the characteristics of raw genomic data and to help you make informed decisions regarding the handling of your raw genomic data.

This information document was written at the suggestion of doctors and researchers and is to be understood as an additional offer beyond the personal consultations.[116]

How are genomic data generated?

When a genome is sequenced, a long molecule called "deoxyribonucleic acid" (DNA) is decoded. You can imagine DNA as a chain of four different building blocks. The blocks are represented as letters: The four letters contained in DNA are adenine (A), thymine (T), cytosine (C), and guanine (G). As a result of the succession of the building blocks (=sequence), the building instructions for proteins are encoded. Proteins are vital for the normal function of cells.

Methods that determine this letter sequence are described as sequencing technologies. The so-called next-generation sequencing (NGS) technology is a DNA sequencing technology that allows billions of DNA molecules to be decoded in parallel. First, the short sequence sections (150 letters) of the DNA are sequenced, which are then saved.

These short sequence sections are saved in a so-called **FASTQ format.**

From smaller parts, a long molecule can be reconstructed by means of overlaps between the short sequence sections. Consequently, the long DNA sequence of a human genome can be assembled on a computer using these shorter sequence sections.

The reconstruction of a genome is carried out using a so called human reference genome, on which the decoded sequences are mapped to. This human reference genome is composed of the sequences of the genomes of several persons and used uniformly internationally.

The alignment process on the reference genome generates a **Binary Alignment Map (BAM)** fileand a corresponding Binary Alignment Index (BAI) file.

[116] Note: If these raw genomic data were generated as part of a research project, further use in other research projects may only take place after consultation with the responsible researcher of the research project. This is in line with the rules of good scientific practice. We ask you to please take note of this.

After the short sequence sections have been aligned with the human reference genome, the differences between the sequenced genome and the reference genome can be identified.

This process is called a "variant call" and creates files in the **Variant Call Format (VCF).**

At this "raw" processing stage, the data do not make sense or yet yield a recognizable "meaning".

Though, if the raw genomic data continues to be evaluated, information can be obtained that can be useful for humans. However, the processing is complicated and requires expertise.

Various raw data formats may be obtained: FASTQ, BAM, or VCF files.

What are "variants"?

Variants are differences in the DNA that have been discovered between a genome and a reference genome. In human history, the genome has been copied so often over the generations that minor differences between different genomes are normal and natural. These differences are what make us unique.

Millions of such variants can be expected, especially with raw genomic data out of a whole genome sequencing.

Only a fraction of these variants is associated with disease. Many of the variants are therefore to be considered normal or healthy.The possible significance of many-variants is unclear.

Why is it important to distinguish between "tumor-specific" and "inheritable" variants?

When studying cancer, attempts are made to find variants related to cancer growth. For this purpose, tumor material and also a control sample (blood or tissue) is taken. This double sampling is important in order to find out which of the found variants are only present in the tumor and not also present in normal body tissue.

The variants that occur only in the tumor are referred to as "tumor-specific". These variants can serve as therapeutic targets and are therefore important for planning cancer treatment.

By evaluating the control samples, in addition to the tumor-specific variants, hereditary variants can also be found. These hereditary variants can be important for cancer as well as for completely unrelated diseases.

The hereditary form of variants may be important for family members.

What is particularly important about genomic data?

Genomic data are like your fingerprint. Your genomic data are unique to you and can therefore inherently **identify you.**

Genomic data may contain personal information about susceptibility to certain characteristics or an increased risk of a particular disease.

You share **"common" genetic material** with your family. Due to heritage, results from your sequencing may not only affect you but also blood-related relatives.

What is particularly important about raw genomic data?

In contrast to evaluated genomic data, raw data do not yet contain anything recognizable in this unprocessed state.

The sequence of raw sequenced data can also have errors that are only detected by experts. Undetected, such errors could lead to incorrect conclusions.

There are privacy risks based on personal information contained in genomic data. A further analysis of the raw data to obtain meaningful content is complicated and requires well-trained experts. Such an analysis could identify genetic variants with possible health significance. In this case, we recommend a re-examination of the sequence as well as arranging additional human genetic counseling.

What risks may arise from using raw genomic data?

· The risks stemming from inadequate analysis of raw genomic data

Especially whole genome sequencing has the potential to to identify changes (variantswhose biological and medical implication are still unclear.

This also means that interpreting these data requires very complex analytical and bioinformatic processes as well as close cooperation between treating doctors and the researchers involved.

Without the combination of experience and expertise in analyzing the complex data, there is a risk of incorrect results or insufficient interpretations that could be of harm.[117]

The risk of genetic discrimination

In order to avoid genetic predisposition from leading to discrimination, patients in Germany have been protected from "discrimination on the basis of genetic characteristics" by the German Genetic Diagnostics Act (Section 1 GenDG) since 2010. The German Genetic Diagnostics Act prohibits healthinsurance companies and employers from discriminating against people on the basis of their genetic characteristics. However, in the case of insur-

[117] A private genetic testing provider mixed up customers' samples. The users received incorrect results. MacArthur D., Sample Swap at 23andMe: A Cautionary Tale (July 6, 2010), URL: https://www.wired.com/2010/06/sample-swaps-at-23andme-a-cautionary-tale/ (Retrieved on April 26, 2019).

As part of an investigation into the accuracy of analyses by private genetic test providers, identical samples were sent from customers to various private providers. Different results were obtained in about one third of the analyses. Ng, P., Murray S., *An Agenda for Personalized Medicine, in: Nature* 641 No. 7265 (2009), p.724-26.

In a recent study on the reliability of health-related results from DTC-GT companies, misinterpreted and false-positive results were issued to consumers in over 40% of cases. Tandy-Conner S. et al., False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care, in: *Genet Med.* 20 No.12 (2018), p. 1515-1521.

ance companies, the statutory provisions (Section 18 of the GenDG) allow for the use of genetic analyses that have previously been carried out if "life insurance, disability insurance, and pension insurance (...) include benefits of more than EUR 300,000 or more than EUR 30,000 annual pension per year".

Risks to informational self-determination with genomic data

Personal information about you can be learned by further analyzing raw genomic data. Therefore, the use of raw data poses a risk to privacy.

Not every patient or study participant wants third parties to receive information about a disease assessment without their consent.

Is there a 100% protection against genetic privacy?

Despite the protection measures enshrined in law, it remains difficult to ensure 100% protection of genetic privacy, since your genome is like a personal fingerprint.

The more possibilities there are to link your genetic data, e.g., from public databases for genealogy, with other entries e.g. on social media, the easier it is to assign genomic data to your person.

What does linking data mean?

An example in the United States shows how third parties can use genomic data from publicly accessible databases and additional information to find an individual. In this example, the FBI was able to identify a long-sought serial killer, the "Golden State Serial Killer", through a publicly available ancestry database. In principle, such databases can be used to obtain information about a person's own background, ancestors, and previously unknown family members. Users have their DNA sequenced, evaluated, and then upload the DNA profile onto the public database.

In the case of the Golden State Killer, a DNA profile/genetic fingerprint of the perpetrator was created using DNA traces from the crime scenes. Using this DNA fingerprint, a publicly accessible genealogy database was used to search for DNA profiles similar to the DNA profile of the perpetrator. In this specific case, a search of the almost 1 million profiles in the database revealed that several of the persons registered in the database were related to the perpetrator. As it turned out, the DNA matches were from the killer's third and fourth cousins. Additional information, such as a reconstructed the family tree, the approximate age of the perpetrator, and the locations of the crime scenes eventually helped investigators to narrow the suspects down. A re-examination of this man's sample confirmed that his DNA profile matched 100% to that of the perpetrator.[118]

The suspect himself had never undergone a DNA test. But the reconstruction of the family tree was actually the link that made it possible to solve the crime with additional information.

[118] Syndercombe Court DForensic genealogy: Some serious concerns, in: Forensic Sci Int Genet Nr. 36 (2018), p. 203-204.

How can you protect your genetic privacy?

Genome data from research are stored in access-controlled databases, and their evaluation is only permitted for specified purposes. When conducting research, genomic data are usually stored under a pseudonym. Therefore,a link between you as a person and your genomic data can only be established by knowing the code and with access authorization. Access to your data is only granted in a research context and after a thorough examination of the person requesting access.

The risk of being reidentified may increase due to:

A. Storage of your raw genomic data either in publicly accessible databases or with (online) companies.
B. Additional, publicly available, personal information about you in other databases or in social media posts.

We hope that this information has provided you with answers to your questions about the opportunities and risks of sharing your genome data.

Before evaluating and using the raw genomic data, however, you could ask yourself the following questions:

· Who has long-term control over your raw genomic data?

· What information about you may be disclosed and who may obtain it?

We would be pleased to answer these and possible further questions in a personal consultation.

7.2. Template for the declaration of the study participant/patient on the provision of information and release of raw genomic

a) Name of the study participant/ patient:

_____

b) I have been informed by telephone and/ or in writing (please delete as applicable) about general aspects of the raw data type that are connected to me and which I have requested access to. I have received detailed written information on the general nature of raw genomic data and on the information relating to health issues which could potentially be obtained from these raw data.

- I am aware that the data stem from a scientific context. Without further clinical, diagnostic validation, these results should not be used for clinical diagnosis.

- I am aware that information that can be generated from the raw data may also pose a psychological or social burden.

- I have also taken note of the information on risks and harm related to data protection (identification based on raw genomic data, risk of misuse of data, relevant disclosure obligations).

c) I have taken advantage of the individual consultation offered to me.
or
I have chosen not to make use of the individual consultation offered to me.
(Please delete as applicable)

d) I understand that the raw data are sensitive data concerning both myself and my biological relatives;

e) By releasing the raw data to me, I take responsibility for the further handling and security of the transferred raw data and for any possible consequences that result from my handling of the raw data for third parties or myself.

_____
Location, date

_____
Signature of the study participant/patient

IMPRINT

Publisher:
Marsilius-Kolleg der Universität Heidelberg
Im Neuenheimer Feld 130.1
69120 Heidelberg
Germany

Design and Layout:
Designbüro Waldpark

DOI: 10.11588/fmk.2020.0.76125
ISSN: 2196-2839

Bibliographic Information of the German National Library:
The German National Library catalogs this publication in the German National Bibliography; detailed bibliographic information can be found on the Internet website: http://dnb.d-nb.de.