



OPEN LETTER

# Returning genome sequences to research participants: Policy and practice [version 1; referees: 2 approved]

Caroline F. Wright <sup>1</sup>, Anna Middleton <sup>1</sup>, Jeffrey C. Barrett <sup>1</sup>, Helen V. Firth<sup>1,2</sup>, David R. FitzPatrick<sup>3</sup>, Matthew Hurles<sup>1</sup>, Michael Parker <sup>4</sup>

<sup>1</sup>Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

<sup>2</sup>East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

<sup>3</sup>MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU, UK

<sup>4</sup>The Ethox Centre, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK

**v1** First published: 24 Feb 2017, 2:15 (doi: [10.12688/wellcomeopenres.10942.1](https://doi.org/10.12688/wellcomeopenres.10942.1))  
Latest published: 24 Feb 2017, 2:15 (doi: [10.12688/wellcomeopenres.10942.1](https://doi.org/10.12688/wellcomeopenres.10942.1))

## Abstract

Despite advances in genomic science stimulating an explosion of literature around returning health-related findings, the possibility of returning entire genome sequences to individual research participants has not been widely considered. Through direct involvement in large-scale translational genomics studies, we have identified a number of logistical challenges that would need to be overcome prior to returning individual genome sequence data, including verifying that the data belong to the requestor and providing appropriate informatics support. In addition, we identify a number of ethico-legal issues that require careful consideration, including returning data to family members, mitigating against unintended consequences, and ensuring appropriate governance. Finally, recognising that there is an opportunity cost to addressing these issues, we make some specific pragmatic suggestions for studies that are considering whether to share individual genomic datasets with individual study participants. If data are shared, research should be undertaken into the personal, familial and societal impact of receiving individual genome sequence data.

## Open Peer Review

Referee Status:

	Invited Referees	
	1	2
<b>version 1</b> published 24 Feb 2017	 report	 report

1 **Bartha Maria Knoppers**, McGill University, Canada

2 **Alison Hall**, PHG Foundation, UK

## Discuss this article

Comments (0)

**Corresponding author:** Caroline F. Wright ([cw9@sanger.ac.uk](mailto:cw9@sanger.ac.uk))

**Competing interests:** No competing interests were disclosed.

**How to cite this article:** Wright CF, Middleton A, Barrett JC *et al.* **Returning genome sequences to research participants: Policy and practice [version 1; referees: 2 approved]** Wellcome Open Research 2017, 2:15 (doi: [10.12688/wellcomeopenres.10942.1](https://doi.org/10.12688/wellcomeopenres.10942.1))

**Copyright:** © 2017 Wright CF *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [HICF-1009-003], a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute [098051]. The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health. Publication in Wellcome Open Research does not imply endorsement by Wellcome. The DDD study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC; GEN/284/12 granted by the Republic of Ireland REC).  
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**First published:** 24 Feb 2017, 2:15 (doi: [10.12688/wellcomeopenres.10942.1](https://doi.org/10.12688/wellcomeopenres.10942.1))

## Context

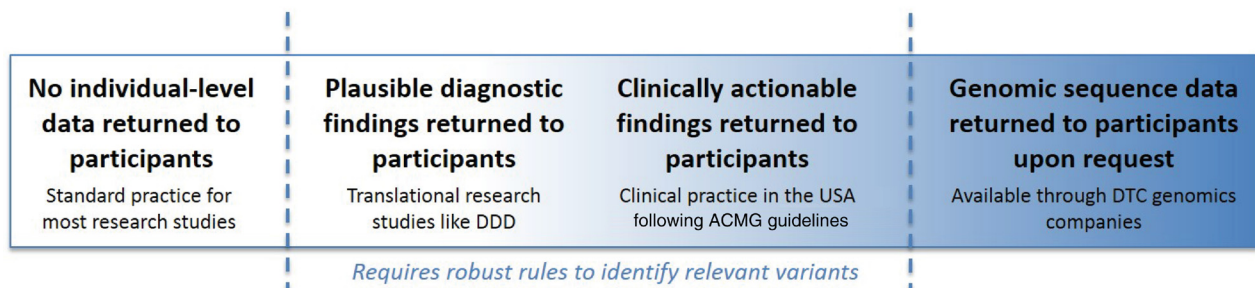
As DNA sequencing becomes cheaper and more widely available, the question of what genomic data could or should be returned to an individual research participant has become important (Figure 1). To date, most discussion in the literature has focused on the relative merits of returning specific health-related findings that are either pertinent or incidental to the purpose of sequencing<sup>1</sup>. Strong arguments have been made both for and against returning actionable findings<sup>2–4</sup>, and numerous clinical and research laboratories have developed robust systems for identifying, assessing and reporting variants that might fall into this category<sup>5</sup>.

The option of returning individual ‘raw’ genome sequence data – BAM or VCF files (see Table 1) – has thus far received less attention, and until now the best practice has been not to share sequence data with individual participants<sup>6</sup>. Since the vast majority of the public (and indeed most researchers) would be unable to meaningfully interpret large and complex genomic datasets, the value of returning such data is unclear. However, although there is currently no legal avenue for participants to request their data<sup>7</sup>, and no expectation upon researchers to share data if it negatively affects their research<sup>8</sup>, the public are widely supportive of the idea and anecdotal examples exist of such requests being made<sup>9</sup>. When surveyed, most people indicate that they plan to seek out an interpretation of their data, while others wish to access their data simply because it is ‘theirs’<sup>9</sup>. Coupled with the potential benefits – supporting autonomy, empowerment, health prevention, reciprocity, and improved trust between researchers and society<sup>10</sup> – this suggests researchers and research institutions have an obligation to consider whether, how and specifically what individual sequence data could reasonably be returned to study participants. Moreover, as genome sequencing becomes more integrated into mainstream healthcare,

scientists may come under increasing pressure to accede to requests to provide individual research participants with their data. Anticipating this development, the UK 100,000 Genomes Project Protocol (Genomics England, January 2017, Section 9.1.1; <https://www.genomicsengland.co.uk/information-for-gmc-staff/rare-disease-documents/>) already states that “patients will have the right to request that their whole genome sequencing (WGS) data be made available to them.”

Given that such requests are likely to increase, here we outline the issues that need to be considered when formulating a policy on returning genome sequence data to individual study participants. Specifically, we address two questions: what are the resource implications of returning sequence data and, if it were cost-neutral (or specifically funded within the research proposal), should a mechanism be provided to facilitate individual-level data sharing?

As a case study to illuminate the issues, we use the UK Deciphering Developmental Disorders (DDD) Study, a translational research project involving whole exome sequencing of nearly 14,000 clinically-ascertained families (>32,000 exome sequences) to find the genetic causes of severe developmental disorders<sup>11</sup>. The DDD study has pioneered a proportionate data sharing policy<sup>12</sup>, which returns likely diagnostic variants to families via DECIPHER<sup>13</sup> and their local clinical teams, and enables *bona fide* researchers to access anonymised whole exome sequence data from the European Genome-phenome Archive under a managed access model<sup>14</sup>. However, the DDD study does not currently return whole exome sequence data to individual participants, a decision that was reached after careful consideration of the issues outlined below and how best to deploy the study’s finite resources.



**Figure 1. Options for returning genome sequence data to participants.** DDD, Deciphering Developmental Disorders study; ACMG, American College of Medical Genetics and Genomics; DTC, direct-to-customer.

**Table 1. Genomic sequence data file types and sizes.** \*GB, gigabytes; WES, whole exome sequencing (i.e. all the protein coding regions of the genome, ~50–60 million bases); WGS, whole genome sequencing (~3 billion bases, assumes 30X coverage).

Name	Description	WES file size (GB)	WGS file size (GB)
BAM	<i>Binary Alignment/Map format</i> : nucleotide sequence with corresponding quality scores mapped to the reference genome, derived from raw data files using an alignment algorithm	~5–15	~150–250
VCF	<i>Variant Call Format</i> : files for storing variant bases relative to the reference genome, which are derived from sequence files using a variant calling algorithm; usually annotated with allele frequency and predicted consequence	~0.02	~0.2

## Logistical challenges

Even in cases such as the DDD study, where exome sequencing data have been generated and there is a direct link (via a clinician) to each individual participant, there are still substantial resources required to ensure that whole exome data could be safely returned to individuals<sup>7</sup>.

*Confirming the data belong to the requestor:* the laboratory and data management systems utilised by researchers are not generally required to meet the rigorous quality standards of accredited diagnostic laboratories. Therefore, it is imperative to confirm that any data generated through a research pipeline actually belongs to the individual requesting the data. Sharing person A's unique genomic data with person B, without their knowledge or consent, would be a negligent and egregious abuse of trust and, depending on the use to which the data were put, could constitute a major clinical risk. To solve this problem, a second DNA sample could be compared to the dataset using a DNA fingerprint. Resources would be needed to support secondary sampling, alongside processes to manage sample logistics, DNA extraction, genotyping and data analysis.

*Providing bespoke informatics support:* although some research participants may have some degree of personal bioinformatics expertise, who may request their data for a variety of reasons, providing the data in an accessible format would likely require some informatics support. Research datasets are frequently generated and annotated using experimental bioinformatics pipelines that are constantly evolving, and annotations can change substantially over time. Additionally, care must be taken to remove annotations common in such files that might reveal sensitive information about other members of the cohort. It would therefore be necessary to develop a specific bioinformatics pipeline to clean any research datasets prior to release to participants. Data return options vary from a simple encrypted hard drive to a user-friendly secure data access portal, which would require substantial resources to set-up and manage, but would offer greater equity of access. Either option would require some technical management to ensure that the correct data are uploaded and accessed by the correct individual.

## Ethico-legal considerations

In addition to the resource-dependent issues identified above, there are a number of other issues that need to be considered in studies involving patient populations, such as families in the DDD study.

*Returning data to family members:* since the majority of genetic variants are shared between close biological relatives, any returned data may have implications for other (un-sequenced) members of the family<sup>15</sup>. Although a single participant may have a right to their own genome data, regardless of whether it is "shared" with family members, the familial implications should be made clear, and in cases where an entire family's data are requested, consent should be sought from all individuals or proxies upon whom data are being requested. In addition, paediatric studies, such as DDD, must address the thorny issue of returning a child's genome sequence to

their parents<sup>16</sup>, where it is complex to determine who has rights of control and to what extent as the child approaches and passes the age of majority. There are already well rehearsed arguments that attempt to balance an individual's right to privacy with the desire to protect their relatives from harm<sup>15</sup>, but this issue is even more complex within the context of a long-term research study, during which time – unbeknownst to the researchers – an individual may gain (or lose) the capacity to make a decision about their own data. Researchers should be aware of the legal status of the data within their jurisdiction, particularly in relation to individuals who lack capacity to give consent.

*Mitigating against unintended consequences:* much of the previous literature discussing return of individual results from genome sequencing studies has focused on the utility of the data and its limitations<sup>17–19</sup>. Our empirical data on attitudes towards the return of genome sequence data has shown that while research participants say they would like access to their 'raw data', many have unrealistic expectations of what they can do with it, e.g. 43% (n=1844) said they would take their raw DNA data to their General Practitioner<sup>9</sup>. Although numerous genome interpretation services exist, unlike returning specific health-related findings where the onus lies with the research team to develop robust processes for identifying and verifying relevant variants, researchers could not be held responsible for any participant-initiated or third party analyses of their data. Nonetheless, there are legitimate concerns about the moral and legal responsibilities of researchers and research institutes for any harm that might befall participants or members of their family as a result of data return. There are no regulated standards for genomic analysis and very limited consensus between existing interpretation services for personal whole genome data<sup>20</sup>. There could also be a substantial impact on clinical services, due to interpretation requests and follow-on work, creating a significant and potentially unmanageable healthcare demand. Data that are analysed in a clinical context could potentially result in litigation for negligence due to a different interpretation of results, an incorrect diagnosis, or even a 'missed' life-saving diagnosis. To mitigate against these risks, the procedure for data generation and processing should follow best practice (to the extent that it is defined) and be clearly documented, along with known limitations as part of the terms and conditions of receiving data. Researchers must be clear that they are not providing a clinical service, but participants should be signposted to appropriate sources of further information to help them decide what to do next. Within the wider ecosystem, development of quality-assured (e.g. "kite-marked") websites are needed, which both clinical geneticists and genomic scientists should agree are useful for the public.

*Ensuring appropriate governance:* careful consideration and oversight of the issues outlined above is essential for any study wishing to return genomic sequence data to individual participants. There needs to be a formal and transparent process in place for making decisions, which involves a number of stakeholders involved both in making the initial decision to disclose data and in maintaining on-going oversight and ensuring good data governance. This might include the principal investigators, clinical and

research study sponsor(s), professional bodies (including legal and ethics representation) and patient representatives.

## Conclusions

In the context of the open data movement and the rapid growth of genome sequencing, it is likely that researchers will increasingly face requests from study participants to access their own data, for various reasons. Research studies could potentially face reputational risk, loss of trust or litigation either from withholding data without good reason, or from inappropriate or reckless data return. Researchers have a duty to society and their funders to make the best use of finite resources, and must balance their moral responsibilities to individual participants with broader responsibilities to the whole cohort and the creation of generalizable knowledge (note, these duties are often different to the duties of clinicians). Moreover, erroneous diagnoses arising from over-interpretation of the data and poor implementation of genomics could result in real harm to patients, particularly in translational studies involving vulnerable clinical groups. There is a pressing need for a rigorous public health evaluation of the significance of identifying actionable variants in asymptomatic individuals with no family history.

We conclude that unless the return of genomic sequence data to individual participants is specifically and appropriately resourced, potentially by research funders themselves, to do so could do more harm than good. How can we reconcile this with mounting ethical arguments in favour of data sharing, and strong empirical support for receiving individual genome sequence data? Where resources are available to support robust sample and data management<sup>21</sup>, as well as good data governance, returning individual genome sequence data to research participants is logistically achievable and may also be morally desirable. For studies considering whether to share individual genomic datasets with individual study participants, we make the following pragmatic suggestions:

- (1) A second sample should be obtained from all requestors to verify the provenance of the data;
- (2) Informatics support should be provided to generate appropriately processed datasets and enable individuals to access their data;
- (3) Research budgets should explicitly include funding to support return of individual-level data;

- (4) The research team should document their protocols, processes and data limitations, and signpost participants to quality sources of information;
- (5) Appropriate governance structures should be put in place to ensure data access requests are handled correctly, including signed consent forms from all requestors;
- (6) Research should be undertaken into the personal, familial and societal impact of receiving individual genome sequence data.

---

## Author contributions

All authors contributed on behalf of the DDD study. CFW drafted the manuscript with advice from MP; CFW, JCB, DRF, HVF, MEH and MP jointly discussed and decided the data sharing policy with respect to the DDD study and AM led the associated social science study gathering empirical evidence. All authors discussed and agreed to the final content of the manuscript.

## Competing interests

No competing interests were disclosed.

## Grant information

The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [HICF-1009-003], a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute [098051]. The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health. Publication in Wellcome Open Research does not imply endorsement by Wellcome. The DDD study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC; GEN/284/12 granted by the Republic of Ireland REC).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Acknowledgements

The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

## References

1. Wolf SM, Lawrenz FP, Nelson CA, *et al.*: **Managing incidental findings in human subjects research: analysis and recommendations.** *J Law Med Ethics.* 2008; **36**(2): 219–48, 211.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Bredenoord AL, Kroes HY, Cuppen E, *et al.*: **Disclosure of individual genetic data to research participants: the debate reconsidered.** *Trends Genet.* 2011; **27**(2): 41–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Angrist M: **You never call, you never write: why return of 'omic' results to research participants is both a good idea and a moral imperative.** *Per Med.* 2011; **8**(6): 651–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Knoppers BM, Zawati MH, Senecal K: **Return of genetic testing results in the era of whole-genome sequencing.** *Nat Rev Genet.* 2015; **16**(9): 553–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Green RC, Berg JS, Grody WW, *et al.*: **ACMG recommendations for reporting**

- of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013; **15**(7): 565–74.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Jarvik GP, Amendola LM, Berg JS, *et al.*: **Return of genomic results to research participants: the floor, the ceiling, and the choices in between.** *Am J Hum Genet*. 2014; **94**(6): 818–26.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  7. Kaye J, Kanellopoulou N, Hawkins N, *et al.*: **Can I access my personal genome? The current legal position in the UK.** *Med Law Rev*. 2014; **22**(1): 64–86.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  8. Middleton A, Morley KI, Bragin E, *et al.*: **No expectation to share incidental findings in genomic research.** *Lancet*. 2015; **385**(9975): 1289–90.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  9. Middleton A, Wright CF, Morley KI, *et al.*: **Potential research participants support the return of raw sequence data.** *J Med Genet*. 2015; **52**(8): 571–4.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  10. Hallowell N, Hall A, Alberg C, *et al.*: **Revealing the results of whole-genome sequencing and whole-exome sequencing in research and clinical investigations: some ethical issues.** *J Med Ethics*. 2015; **41**(4): 317–21.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  11. Wright CF, Fitzgerald TW, Jones WD, *et al.*: **Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data.** *Lancet*. 2015; **385**(9975): 1305–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  12. Wright CF, Hurles ME, Firth HV: **Principle of proportionality in genomic data sharing.** *Nat Rev Genet*. 2016; **17**(1): 1–2.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  13. Bragin E, Chatzimichali EA, Wright CF, *et al.*: **DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation.** *Nucleic Acids Res*. 2014; **42**(Database issue): D993–D1000.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  14. Lappalainen I, Almeida-King J, Kumanduri V, *et al.*: **The European Genome-phenome Archive of human data consented for biomedical research.** *Nat Genet*. 2015; **47**(7): 692–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  15. Dheensa S, Fenwick A, Lucassen A: **'Is this knowledge mine and nobody else's? I don't feel that.' Patient views about consent, confidentiality and information-sharing in genetic medicine.** *J Med Ethics*. 2016; **42**(3): 174–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  16. Knoppers BM, Rioux A, Zawati MnH: **Pediatric research 'personalized'? International perspectives on the return of results.** *Per Med*. 2013; **10**(1): 89–95.  
[Publisher Full Text](#)
  17. Gliwa C, Berkman BE: **Do researchers have an obligation to actively look for genetic incidental findings?** *Am J Bioeth*. 2013; **13**(2): 32–42.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  18. Christenhusz GM, Devriendt K, Dierickx K: **To tell or not to tell? A systematic review of ethical reflections on incidental findings arising in genetics contexts.** *Eur J Hum Genet*. 2013; **21**(3): 248–55.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  19. Wolf SM: **The past, present, and future of the debate over return of research results and incidental findings.** *Genet Med*. 2012; **14**(4): 355–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  20. Corpas M, Valdivia-Granda W, Torres N, *et al.*: **Crowdsourced direct-to-consumer genomic analysis of a family quartet.** *BMC Genomics*. 2015; **16**: 910.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  21. Warren E: **Strengthening Research through Data Sharing.** *N Engl J Med*. 2016; **375**(5): 401–3.  
[PubMed Abstract](#) | [Publisher Full Text](#)



# Open Peer Review

Current Referee Status:  

---

## Version 1

Referee Report 09 March 2017

doi:[10.21956/wellcomeopenres.11795.r20527](https://doi.org/10.21956/wellcomeopenres.11795.r20527)



**Alison Hall**

PHG Foundation, Cambridge, UK

This excellent paper addresses the challenges that might be raised if entire genome sequences were released to individual research participants. I particularly like the fact that the authors have described the logistical and bioinformatics support in tandem with ethico-legal considerations. As the authors point out in the conclusions section of the paper, research funders and researchers will need to pro-actively plan for whole genome sequence release, well in advance of its taking place. It seems increasingly likely that policies will need to take account of this possibility - and this paper deserves to be widely read.

Title and abstract: apposite, clear and concise

Article Content: One very minor point could be added if the paper is amended: this would be to highlight in a few words in the ethico-legal considerations discussion that unintended consequences might include unexpected knowledge about familial relationships (such as mis-attributed paternity or consanguinity - especially if multiple genomes from trios are returned).

Conclusions: A helpful set of recommendations to guide future practice.

**Competing Interests:** Caroline Wright (the first author) used to work at the PHG Foundation (my current employer) but left more than 5 years ago. I do not consider myself to have a personal relationship with her that might compromise or bias my review.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 06 March 2017

doi:[10.21956/wellcomeopenres.11795.r20667](https://doi.org/10.21956/wellcomeopenres.11795.r20667)



**Bartha Maria Knoppers**

Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine, McGill University, Montréal, QC, Canada

“Returning genome sequences to research participants: Policy and practice” is an excellent article!  
Approved as is

Title and Abstract: The title of the article and abstract are appropriate

Article Content: The “Logistical Challenges” section is a novel (necessary and practical!) addition to traditional ELSI discussions.

Conclusions: The conclusions highlighting the resource/infrastructure issues as fundamental to realizing a return of results policy – bravo!

P.S. SVP put the words “...no legal avenue in the UK for participants...” – p. 2 Context section (2<sup>nd</sup> parag, 3<sup>rd</sup> sentence)

***Competing Interests:*** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---