

COMMENTARY

Open Access

The impact of cross-kingdom molecular forensics on genetic privacy



Eran Elhaik^{1*} , Sofia Ahsanuddin², Jake M. Robinson^{3,4}, Emily M. Foster⁵ and Christopher E. Mason^{6,7,8,9*}

Abstract

Recent advances in metagenomic technology and computational prediction may inadvertently weaken an individual's reasonable expectation of privacy. Through cross-kingdom genetic and metagenomic forensics, we can already predict at least a dozen human phenotypes with varying degrees of accuracy. There is also growing potential to detect a "molecular echo" of an individual's microbiome from cells deposited on public surfaces. At present, host genetic data from somatic or germ cells provide more reliable information than microbiome samples. However, the emerging ability to infer personal details from different microscopic biological materials left behind on surfaces requires in-depth ethical and legal scrutiny. There is potential to identify and track individuals, along with new, surreptitious means of genetic discrimination. This commentary underscores the need to update legal and policy frameworks for genetic privacy with additional considerations for the information that could be acquired from microbiome-derived data. The article also aims to stimulate ubiquitous discourse to ensure the protection of genetic rights and liberties in the post-genomic era.

Keywords: Genomics, Genetic privacy, Metagenomics, Next-generation sequencing, Forensics, Genetic discrimination, Microbiomics

Background

DNA sampling and sequencing are now routinely applied in several areas of research and practice. These include criminal investigations, research on public parks and subways, and multidisciplinary genomics projects. The collection of genomic and metagenomic profiles across the world has led to a greater understanding of the world's genetic diversity. However, it also raises an array of ethical and legal questions [1–4]. Indeed, privacy expectations are reduced once personal materials are in the public realm. A person's genetic information (e.g., acquired from their hair, skin cells, or microbiome) could conceivably be collected in a public setting. This could have a considerable impact on privacy, and the

ability to genetically discriminate could be utilized for nefarious means.

In the USA, the Genetic Information Non-discrimination Act 2008 (GINA) made it illegal for companies with over 15 employees to use genetic discrimination for health insurance and employment purposes [5]. However, a person and their family's genetic profiles can still be utilized to deny or adjust life insurance, long-term care insurance, and disability insurance [6]. Survey data show that 6 years after its enactment, over 80% of US adults were unaware of GINA, and 30% of those who were informed reported deep concerns about genetic discrimination [5]. The National Institute of Health's "Precision Medicine Initiative" [7] and genome-guided medical care also raise concerns regarding privacy and safety. This is due to the risks of data sharing when congenital haplotypes found in the data may be used to discriminate against patients and their family members. In addition, the depth of information gained from genetic remnants on public surfaces has implications for individual

* Correspondence: eran.elhaikj@biol.lu.se; chm2042@med.cornell.edu

¹Department of Biology, Lund University, 22362 Lund, Sweden

⁶Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10021, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

rights to genetic privacy under GINA and the criminal justice system.

In 2019, an investigation was launched into the supply of 1500 DNA samples from the Crumlin children’s hospital in Ireland to a DNA collection company without authorization from the patients [8]. This may represent a breach of the European Union’s general data protection regulation (GDPR) Article 9, which requires proper consent for the processing of DNA data [9]. However, many countries implement their own interpretation of this regulation, thus leading to calls to develop a cross-border code of conduct for genomic data sharing [10]. Importantly, explicit considerations for microbiome-derived information are not part of the genetic privacy narrative that informs this regulation. In March 2018, a Parliamentary Joint Committee (PJC) inquiry into the Australian life insurance industry recommended an immediate ban on the use of predictive genetic test results [11]. A recent study revealed illegal genetic discrimination by Australian life insurance companies, representing a broader concern [11].

In criminal law, the US legal system’s notion of privacy under the Fourth Amendment is based on a person’s reasonable expectation of privacy: “what a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection [12].” The prevailing question is should humans, who constantly shed their genetic (host and microbiome) materials, have a reasonable expectation that these materials will not be collected from public surfaces? In other words, should genetic material be treated differently than fingerprints or material property? As we enter the era of “ubiquitous sequencing [2],” people must now assume that their DNA can be collected, sequenced, annotated, and interpreted by other people, particularly

if retrievable from a public place. In that respect, another important question is when a person leaves DNA or RNA behind on a public surface, what personal information could be revealed?

Main text

Molecular signatures in the post-genomic era

Technological advances in genomics have created the potential for forensic methods beyond DNA, leading us to a “post-genomic” forensic era with identifiable information in the RNA and epigenetic states. Information that can contribute to identifying human individuals could potentially be found within the vast diversity of microorganisms that are in, on, and around us, collectively known as the microbiome [13, 14]. Next-generation sequencing (NGS) has been used to localize these organisms to a unique part of the human body, forming a “molecular cartography” of an individual [15]. Beyond microorganisms, recent metagenomics studies have enabled a cross-kingdom examination of life, including urban genetic maps [16] in worldwide cities [17]. The availability of these new datasets and methods can enable inferences of at least a dozen phenotypes across several categories. These include human DNA, human RNA, epigenetics, epitranscriptome, and metagenomes (Fig. 1). Furthermore, the advent of “Big Data” allows the intersection of these identifiers to provide additional revelatory information.

In the following 12 sections, we summarize the phenotypes that could be discerned from DNA or RNA deposited on public surfaces, along with methods to acquire this information. Some of the suggestions are still in development, but given the ubiquity of genomic tools and possibilities for advancements, it is still important to highlight areas with potential.

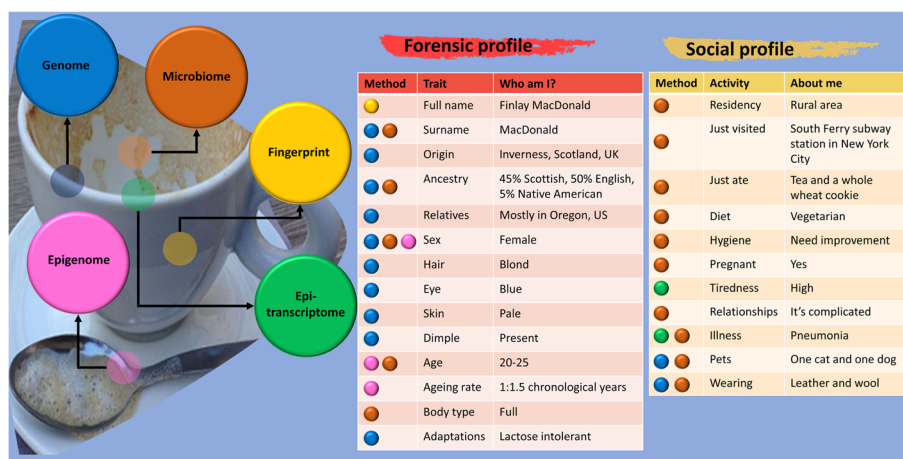


Fig. 1 Cross-kingdom methods of forensics. Categories of multi-omic and multi-kingdom measurements can create both forensic (left table) and social (right table) profiles for a person, based on their epigenome (pink), epitranscriptome (green), fingerprint (yellow), microbiome (orange), and genome (blue). Categories for types of inferred information are detailed in each table by the trait/activity and the revelatory information

Identity

Criminal cases have been solved by comparing DNA found at the crime scene with the DNA of individuals or their families who have submitted their genomic data to genealogy databases [18]. However, a supposedly “anonymous dataset” may also reveal a person’s identity using the Y-chromosome’s short tandem repeats (STRs), which can be linked to surnames [19, 20], or by querying genealogy databases [21], which arguably is constitutional under the Fourth Amendment [22]. The widespread use of DNA profiling with Combined DNA Index System (CODIS) markers for anyone who is arrested allows law enforcement to use STR markers to find a relative who committed a crime through “familial DNA searches” [23]. Assuming a database of gut metagenomes from individuals, variation in gut microbiota can contribute towards genomic fingerprinting [24], which can be crossed with consumer DNA that companies like FamilyTreeDNA are sharing with the FBI [25]. Studies have shown that the skin microbiome may exhibit a degree of inter-individual variability and can be recovered from objects such as computers even if left untouched for 2 weeks [26]. The microbiome has also been linked to a person’s phone and shoes [27]. Many of the studies in the microbial forensics realm have moderate model accuracies, and improvements are required before highly accurate information can be elucidated. However, as technology continues to improve, there is potential to identify individuals through microbial profiles with a much higher accuracy level in the near future.

Age

Each time a cell replicates, a chromosomal loss can occur. Measuring these molecular alterations in the Y- and X-chromosomes [28], telomeres [29], and DNA methylation [29] allows the molecular age of the body to be estimated. Applications of the latter method have been reported to predict chronological age [29] closely. The skin microbiome also has the potential to be a predictor of age with high accuracy (mean \pm standard deviation, 3.8 ± 0.45 years of chronological age) [30]; however, the results vary by tissue, gender, and age. Further studies are needed to demonstrate the applicability of this approach.

Biological sex

This can be inferred through DNA, RNA, and epigenetics [31]. The microbial communities of pubic hairs [32] and the lower gastrointestinal tract [33] also differ between the sexes. Moreover, Luongo et al. [34] studied airborne microbial diversity in university dormitory rooms ($n = 91$). Through relative abundance analysis, machine learning techniques could predict the biological sex of occupants with 79% accuracy [34]. As microbial

relative abundance data are a completely different form of data to that of host DNA, it could contribute towards gaining personal information in the absence of quality host DNA samples.

Facial features

Several facial features, like hair color and texture/thickness, eye color, and skin tone, can be genetically predicted with varying levels of accuracy [23], and this is expected to improve as genotypic and phenotypic databases continue to increase.

Ancestry and biogeography

Differences between continental groups are observable in different phenotypes, such as genomic sequences [35] and the vaginal microbiome [36]. Genetic variation between human populations is engraved in ancestry informative markers (AIMs) [37], which can be used to predict the geographic region of the DNA’s ancestral origins with an accuracy of a few hundred kilometers [38] or less [39, 40]. However, caution should be practiced since biases may arise primarily due to the inaccuracy of ancestry estimation tools and subpopulation undersampling [41].

Geospatial localization

Various techniques are available to ascertain concurrent localization information, including microbial DNA of soil [42], microbial DNA of surfaces [14], pathogenic viruses [43], and metagenomes [44]. To geo-localize a sample, these approaches require global city-wide reference data, such as those collected by the MetaSUB consortium for metagenomes [14]. Although studies that aim to localize samples typically suffer from modest sample sizes or prioritize classification over prediction, there is undoubtedly potential to develop more rigorous methods and improve the confidence of geospatial inferences.

Zooculture

Metagenomes could provide valuable evidence regarding interactions with pets and wildlife. This information could conceivably be evidence of criminal behaviors such as illegal fishing or poaching [45]. Additionally, the exchange of microbial communities between a pet and the owner can provide information on the length of their association [46]. There are pathways to trace microbial profiles from other non-human species to a given environment or pet ownership. These ideas are theoretical at the moment, but with methodological refinement, there is potential to connect an individual with a particular location through their shared microorganisms with animals [47].

Cell type

The Human Microbiome Project (HMP) [13] and the assignment of bacteria to primary areas of the human body prompted the identification of body parts that may have been in contact with public surfaces in subways [16]. As each tissue also has unique gene expression [48], RNA [49], and epigenetic [50] measurements, different molecules could potentially be used for cross-validation in the future. This is currently unrealistic outside of lab conditions, but given the rapidity of technological advancements and model accuracies, the potential is considerable.

Obesity

Microbiome profiles of obese and non-obese people can differ dramatically [51].

This could not only allow the stratification of individual profiles [52], but also potentially provide insight into valuable information on digestion or the eating habits of the individual [53]. Although host genetics are more reliable for host identification, there is potential to acquire some useable information in the absence of sufficient quality and quantity of DNA from somatic or germ cells.

Pregnancy

The maternal gut microbiome begins similar to that of non-pregnant women [54] but changes throughout the pregnancy [55]. While the statistical relationships in these studies can be weak and multifactorial, they could, eventually, be used to identify an individual's gestational status accurately.

Circadian rhythm

Although this application is presently only speculative, the coupling of RNA modifications with circadian rhythm could be developed as epitranscriptomic markers to determine if someone was sleeping or awake based on the cells left in their bed [56]. However, areas for advancement would need to include optimizing the processing of low biomass samples, which face additional challenges of degradation and environmental contamination.

Disease and infection

The carriage of oral pathogens is not only an indicator of periodontal health [57], but also an indicator of more complex disorders like pancreatic cancer [58]. Assuming RNA is preserved on a surface, modifications to RNA can contribute towards understanding if a person's body is responding to a human immunodeficiency virus (HIV) infection, as well as the type and severity of the HIV infection [59]. It is unrealistic to assume that information can currently be used with a high level of accuracy. Still, it does highlight another potential privacy issue that could have important implications in the future.

Ubiquitous measurements and potential counter-measures

Many recent advances in genomics and biomedicine are made possible by technological advancements in cameras and imaging. These tools have enabled us to peer inside single cells with unparalleled resolution. The affordability and high resolution of contemporary cameras have also enabled their adaptations to forensics. A "time machine" of movement around a city can be created with the use of drones that maintain a geosynchronous location above a city through daily image collection. A private company, Persistent Surveillance, has used manned aircraft equipped with high-powered cameras to track and catch criminals in Camden (New Jersey, USA) and Ciudad Juárez (Mexico) [60]. Location information can also be gained through cell phone tracking using the multilateration of radio signals between different cell towers of the network and the phone or even more easily using GPS. In 2019, the existence of a file containing 50 billion data points recording the movements of 12 million American cell phone users, including the US President and his guests, was revealed [61]. In a landmark US Supreme Court case, it was determined that "the time-stamped data provides an intimate window into a person's life, revealing not only his particular movements but through them his 'familial, political, professional, religious and sexual associations'" and is thereby a violation of the Fourth Amendment to the US Constitution to gain access to these data without a search warrant [62]. Personal information, however, continues to be collected, analyzed, and cross-referenced, as has been demonstrated when a teenager's consumer metadata revealed her pregnancy even before her parents were made aware [63].

The new abilities to collect personal information have severe implications for people's expectations of privacy. DNA evidence, the key to the conviction or exoneration of suspects of various types of crime, from theft to rape and murder, can be fabricated and planted in the crime scene to implicate any known genetic profile [64]. If travel history is detectable, it can not only be mined for criminal or other investigations but could potentially be used in marital disputes, employment discrimination, and other tracking purposes. Combined with data mining methods employed by corporations, surveillance tools are increasingly eliminating the traditional notion of privacy. Indeed, if DNA sequencers were truly ubiquitous, not only could one's identity be matched with their location, but their actions and associations may also be inferred through a "genetic time machine" (Fig. 1) and cross-referenced with corporate and online information.

Due to these potential encroachments on privacy and ample molecular means to link deposited cells and molecules to phenotypes, new approaches for obfuscation

have been developed and even patented [65]. It is possible to order and spray synthetic oligonucleotides to mask one's historical presence in a room [64]. However, unless such "genetic camouflage" encompasses the cross-genome spectrum (Fig. 1), the deception could be detected. Nonetheless, the potential to assemble such a precise molecular match to a person's genetic and molecular identity, which can be obtained from public databases [21], may create new opportunities in forensics and considerable problems. A technology for accurate and specific matches at the genetic, epigenetic, RNA, epitranscriptome, and microbial levels would also augment the ability to frame a person in a criminal context. Moreover, although this is currently speculative, with precise methods for trans-differentiation and re-differentiation of cells, it might be possible to convert the skin cells left behind at the scene of a crime into erythrocytes and imply that a bloody incident had occurred. With easy-to-use CRISPR systems [66] and epigenetic modification mechanisms [67], such precise and potentially malicious genetic manipulations can feasibly be developed.

Given these developments, we call for renewed and widespread discussions on genetic privacy, the legal implications of these new technologies, and updated statutory protections. We also call for explicit considerations for microbiome-derived data in any standards and regulations designed to protect genetic privacy.

Conclusions

Just as the amount of traceable electronic data for people in the modern era only grows with time, the likelihood of a person's biological remnants being found in the environment also increases with time, concomitant with an ever-increasing ability to extract and interpret these data. Many of the tools are theoretical at present, and molecular methods and interpretation are unlikely ever to be 100% accurate. Nonetheless, given the number of tests that can be used across all different data types, metrics, and kingdoms of life, a new "cross-kingdom forensic landscape" has emerged that eviscerates previous notions of privacy. The speed and availability of these tools, algorithms, and multi-layered mechanisms for detection and re-programming of identity from genomics and post-genomic data are increasing. This will likely create challenges for judicial enforcement for those who violate the law. It also raises critical questions about who has access to the metadata, genetic material, and the unintended revelatory information encased in microscopic particles left on every surface.

Given all the identifiable information that is present in a sample and all the metadata about people that are being collected, a new risk of discrimination is now an issue. Unfortunately, legal frameworks, like GINA [6]

and equivalent US state statutes, do not prevent life insurance underwriters from changing their premiums based on genetic markers—even if the markers were taken from genetic material left on a drinking cup (Fig. 1) or the saliva under the stamp or envelope that was mailed to the insurance company [68]. Moreover, any information about family members could also *legally* be used as a basis for altered eligibility, coverage, or premiums on life, disability, or long-term care insurance. By extension, any of the forensics mechanisms described above could potentially be used to change, deny, or alter coverage for a person or their relatives.

These legal frameworks, designed to safeguard worker's DNA and genetic information, are frozen by the definitions provided by the legislature and thus do not apply to the epigenome, microbiome, or metagenome information. Specifically, the GINA statute says that "a genetic test means an analysis of human DNA, RNA, chromosomes, proteins, or metabolites, that detects genotypes, mutations, or chromosomal changes" [5]. In other words, all other biologic non-genomic personal identifying information may be used to achieve what GINA attempted to prevent: health insurance and employment discrimination. Consider, for example, the case of *Lowe v. Atlas Logistics Group Retail Services* [69], where the company's employees were asked to undertake a genetic Short tandem repeats (STR) test to identify the mysterious "devious defecator" who violated their warehouse. Plaintiffs Lowe and Dennis filed a lawsuit under GINA and won; however, had the plaintiffs been asked to submit a gut microbiome sample, GINA would not have been protected them. Concerningly, even the Patient Protection and Affordable Care Act (ACA) of 2010, which prohibits health insurance companies from using genetic information to establish the rules or terms of an individual's eligibility, never defined "genetic information" [70].

It is noteworthy that the GINA and ACA only set the minimum bar of protection against genetic discrimination, and US state laws can set up stricter protections. For example, in 2011, the California Legislature passed the California Genetic Information Nondiscrimination Act" (CalGINA), yet even this more stringent act [71] maintained GINA's definition for genetic information. Given this loophole, an updated GINA and similar frameworks should account for these non-human markers and the battery of molecular signatures described above. Only more inclusive policies that guard any "personally-identifying molecular signature" to be exempt from use in insurance and employment decisions and options for people, as well as their relatives, can guarantee against genetic discrimination. An example for such an elaborated definition can be found in the recently issued California Genetic Information Privacy Act ("GIPA") aimed to regulate the

privacy and security aspects of genetic testing and testing companies. The act defines “Genetic data” as “any data, regardless of its format, that results from the analysis of a biological sample from a consumer, or from another element enabling equivalent information to be obtained, and concerns genetic material. Genetic material includes, but is not limited to, deoxyribonucleic acids (DNA), ribonucleic acids (RNA), genes, chromosomes, alleles, genomes, alterations or modifications to DNA or RNA, single nucleotide polymorphisms (SNPs), uninterpreted data that results from the analysis of the biological sample, and any information extrapolated, derived, or inferred therefrom” [72]. Such updated language can also inform which of these elements could be controlled, utilized for the public good, or patented. Although privacy may be hard to keep in a world of ubiquitous genetic, molecular, and data profiling, the statutes and laws can be updated as the methods and tools advance.

Abbreviations

GINA: Genetic Information Non-discrimination Act; GIPA: Genetic Information Privacy Act; ACA: Affordable Care Act; GDPR: General Data Protection Regulation; STRs: Short Tandem Repeats; CODIS: Combined DNA Index System; AIMS: Ancestry Informative Markers; HMP: Human Microbiome Project; PJC: Parliamentary Joint Committee

Acknowledgements

We would like to thank Yaniv Erlich, Eric Schadt, and Paul Bertone for discussions during the writing of this manuscript.

Authors' contributions

EE, JMR, and CEM wrote this manuscript. SA and EMF assisted in the research. All the authors read and approved the manuscript.

Funding

We would like to thank the Epigenomics Core Facility at Weill Cornell Medicine as well as the Starr Cancer Consortium grants (I9-A9-071) and funding from the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts, Bert L. and N. Kuggie Vallee Foundation, the WorldQuant Foundation, The Pershing Square Sohn Cancer Research Alliance, NASA (NNX14AH50G, NNX17AB26G), the National Institutes of Health (R25EB020393, R01AI125416, R01ES021006), the National Science Foundation (grant no. 1120622), the Bill and Melinda Gates Foundation (OPP1151054), and the Alfred P. Sloan Foundation (G-2015-13964), NIH (U01DA053941), and STARR (I13-0052). Support was also provided by the Tri-Institutional Training Program in Computational Biology and Medicine and the Clinical and Translational Science Center (Jeff Zhu). We would also like to thank the Crafoord Foundation, the Swedish Research Council (2020-03485), and Erik Philip-Sörensen Foundation (G2020-011). The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lund, partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Availability of data and materials

Not applicable

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

E.E. consults the DNA Diagnostics Center. The rest of the authors declare that they have no competing interests.

Author details

¹Department of Biology, Lund University, 22362 Lund, Sweden. ²Department of Medical Education, Icahn School of Medicine at Mount Sinai, New York, USA. ³The Department of Landscape Architecture, University of Sheffield, Sheffield S10 2TN, UK. ⁴The Healthy Urban Microbiome Initiative (HUMI), Adelaide 5005, South Australia. ⁵Columbia, USA. ⁶Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10021, USA. ⁷The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, NY 10021, USA. ⁸The Feil Family Brain and Mind Research Institute (BMRI), New York, NY 10021, USA. ⁹The Information Society Project, Yale Law School, New Haven, CT 06511, USA.

Received: 8 March 2021 Accepted: 7 April 2021

Published online: 20 May 2021

References

- Shamarina D, Stoyantcheva I, Mason CE, Bibby K, Elhaik E. Communicating the promise, risks, and ethics of large-scale, open space microbiome and metagenome research. *Microbiome*. 2017;5(1):132. <https://doi.org/10.1186/s40168-017-0349-4>.
- Erlich Y. A vision for ubiquitous sequencing. *Genome Res*. 2015;25(10):1411–6. <https://doi.org/10.1101/gr.191692.115>.
- Mason CE, Porter SG, Smith TM. Characterizing multi-omic data in systems biology. In: Maltsev N, Rzhetsky A, Gilliam TC, editors. *Systems Analysis of Human Multigene Disorders*. New York: Springer New York; 2014. p. 15–38.
- Hawkins AK, O'Doherty KC. “Who owns your poop?”: insights regarding the intersection of human microbiome research and the ELSI aspects of biobanking and related studies. *BMC Med Genomics*. 2011;4(1):72. <https://doi.org/10.1186/1755-8794-4-72>.
- Green RC, Lautenbach D, McGuire AL. GINA, genetic discrimination, and genomic medicine. *N Engl J Med*. 2015;372(5):397–9. <https://doi.org/10.1056/NEJMp1404776>.
- Pub. L. No. 110–233, 122 Stat. 881. In: <https://www.govinfo.gov/content/pkg/PLAW-110publ233/pdf/PLAW-110publ233.pdf> (Accessed 4 Nov 2019); 2008.
- Collins FS, Varmus H. A New Initiative on Precision Medicine. *N Engl J Med*. 2015;372(9):793–5. <https://doi.org/10.1056/NEJMp1500523>.
- Edwards E. Hospital investigates release of DNA samples to research firm. In: *The Irish Times*. <https://www.irishtimes.com/news/health/hospital-investigates-release-of-dna-samples-to-research-firm-1.3773529> (Last Accessed 15 Feb 2021); 2019.
- The European Parliament and the Council of the EU: Regulation (EU) 2016/679 of the European Parliament of the Council. In: *Official Journal of the European Union*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e2051-1-1> (Last Accessed 15 Feb 2021); 2016.
- Molnár-Gábor F, Korbel JO. Genomic data sharing in Europe is stumbling—Could a code of conduct prevent its fall? *EMBO Molecular Medicine*. 2020;12(3):e11421. <https://doi.org/10.15252/emmm.201911421>.
- Tiller J, Morris S, Rice T, Barter K, Riaz M, Keogh L, et al. Genetic discrimination by Australian insurance companies: a survey of consumer experiences. *Eur J Hum Genet*. 2020;28(1):108–13. <https://doi.org/10.1038/s41431-019-0426-1>.
- Katz V. *United States*, 389 U.S. 347; 1967.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JL. The human microbiome project. *Nature*. 2007;449(7164):804–10. <https://doi.org/10.1038/nature06244>.
- Danko DC, Bezdan D, Afshinnikoo E, Ahsanuddin S, Alicea J, Bhattacharya C, Bhattacharyya M, Blekman R, Butler DJ, Castro-Nallar E et al: Global genetic cartography of urban metagenomes and anti-microbial resistance. *Cell* (in Press) 2021. <https://doi.org/10.1016/j.cell.2021.05.002>.
- Bouslimani A, Porto C, Rath CM, Wang M, Guo Y, Gonzalez A, et al. Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci USA*. 2015;112(17):E2120–9. <https://doi.org/10.1073/pnas.1424409112>.
- Afshinnikoo E, Meydan C, Chowdhury S, Jaroudi D, Boyer C, Bernstein N, et al. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems*. 2015;1(1):72–87. <https://doi.org/10.1016/j.cels.2015.01.001>.
- Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017;551(7681):457–63. <https://doi.org/10.1038/nature24621>.

18. Shapiro E: How a DNA database's new policy is changing police access and could hinder solving cold cases. In: <https://abcnews.go.com/US/leave-murderer-running-streets-dna-databases-policy-changing/story?id=63150489> (Last Accessed 5 Nov 2019): ABC News; 2019.
19. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science*. 2013;339(6117):321–4. <https://doi.org/10.1126/science.1229566>.
20. Claerhout S, Roelens J, Van der Haegen M, Verstraete P, Larmuseau MHD, Decorte R. Ysurnames? The patrilineal Y-chromosome and surname correlation for DNA kinship research. *Forensic Sci Int Genet*. 2020;44:102204.
21. Ney P, Ceze L, Kohno T. Genotype extraction and false relative attacks: security risks to third-party genetic genealogy services beyond identity inference. In: *Network and Distributed System Security Symposium (NDSS)*; 2020.
22. Ortyl E. DNA and the Fourth Amendment: would a defendant succeed on a challenge to a familial DNA search? *Am J Law Med*. 2019;45(4):421–42. <https://doi.org/10.1177/0098858819892746>.
23. Mason-Buck G, Graf A, Elhaik E, Robinson J, Pospiech E, Oliveira M, et al. DNA based methods in intelligence-moving towards metagenomics. Preprints. 2020;2020020158.
24. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. *Nature*. 2012; 493:45.
25. Haag M: FamilyTreeDNA admits to sharing genetic data with F.B.I. In: *The New York Times*. <https://www.nytimes.com/2019/02/04/business/family-tree-dna-fbi.html> (Last Accessed 17 Feb 2021); 2019.
26. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. *Proc Natl Acad Sci USA*. 2010;107(14):6477–81. <https://doi.org/10.1073/pnas.1000162107>.
27. Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, et al. Forensic analysis of the microbiome of phones and shoes. *Microbiome*. 2015;3(1):21. <https://doi.org/10.1186/s40168-015-0082-9>.
28. Guttenbach M, Koschorz B, Bernthaler U, Grimm T, Schmid M. Sex chromosome loss and aging: in situ hybridization studies on human interphase nuclei. *Am J Hum Genet*. 1995;57(5):1143–50.
29. Naue J, Hoefsloot HCJ, Mook ORF, Rijlaarsdam-Hoekstra L, van der Zwalm MCH, Henneman P, et al. Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. *Forensic Sci Int Genet*. 2017;31:19–28. <https://doi.org/10.1016/j.fsigen.2017.07.015>.
30. Huang S, Haiminen N, Carrieri A-P, Hu R, Jiang L, Parida L, et al. Human skin, oral, and gut microbiomes predict chronological age. *mSystems*. 2020;5(1): e00630–19.
31. Liu J, Morgan M, Hutchison K, Calhoun VD. A Study of the Influence of Sex on Genome Wide Methylation. *PLOS ONE*. 2010;5(4):e10028. <https://doi.org/10.1371/journal.pone.0010028>.
32. Tridico SR, Murray DC, Addison J, Kirkbride KP, Bunce M. Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. *Investig Genet*. 2014;5(1):16. <https://doi.org/10.1186/s13323-014-0016-5>.
33. Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. *PloS one*. 2016;11(5):e0154090. <https://doi.org/10.1371/journal.pone.0154090>.
34. Luongo JC, Barberán A, Hacker-Cary R, Morgan EE, Miller SL, Fierer N. Microbial analyses of airborne dust collected from dormitory rooms predict the sex of occupants. *Indoor Air*. 2017;27(2):338–44. <https://doi.org/10.1111/ina.12302>.
35. Elhaik E. Empirical distributions of F_{ST} from large-scale human polymorphism data. *PLoS One*. 2012;7(11):e49837. <https://doi.org/10.1371/journal.pone.0049837>.
36. Fettweis JM, Brooks JP, Serrano MG, Sheth NU, Girerd PH, Edwards DJ, et al. Consortium tVM, Jefferson KK, Buck GA: Differences in vaginal microbiome in African American women versus women of European ancestry. *Microbiology*. 2014;160(10):2272–82. <https://doi.org/10.1099/mic.0.081034-0>.
37. Elhaik E, Yusuf L, Anderson AJ, Pirooznia M, Arnellos D, Vilshansky G, et al. The Diversity of REcent and Ancient huMan (DREAM): a new microarray for genetic anthropology and genealogy, forensics, and personalized medicine. *Genome Biol Evol*. 2017;9(12):3225–37. <https://doi.org/10.1093/gbe/evx237>.
38. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, et al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun*. 2014;5:1–12.
39. Marshall S, Das R, Pirooznia M, Elhaik E. Reconstructing Druze population history. *Sci Rep*. 2016;6(1):35837. <https://doi.org/10.1038/srep35837>.
40. Das R, Wexler P, Pirooznia M, Elhaik E: The origins of Ashkenaz, Ashkenazic Jews, and Yiddish. *Front Genet* 2017, 8(87).
41. CCarress H, Lawson D, Elhaik E. Population genetic considerations for using biobanks as international resources in the pandemic era and beyond. *BMC Genomics*. 2021. <https://doi.org/10.1186/s12864-021-07618-x>.
42. Habtom H, Pasternak Z, Matan O, Azulay C, Gafny R, Jurkevitch E. Applying microbial biogeography in soil forensics. *Forensic Sci Int Genet*. 2019;38: 195–203. <https://doi.org/10.1016/j.fsigen.2018.11.010>.
43. Hellmér M, Paxéus N, Magnus L, Enache L, Arnholm B, Johansson A, et al. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl Environ Microbiol*. 2014; 80(21):6771–81. <https://doi.org/10.1128/AEM.01981-14>.
44. Reese AT, Savage A, Youngsteadt E, McGuire KL, Koling A, Watkins O, et al. Urban stress is associated with variation in microbial species composition—but not richness—in Manhattan. *ISME J*. 2016;10(3):751–60. <https://doi.org/10.1038/ismej.2015.152>.
45. Arenas M, Pereira F, Oliveira M, Pinto N, Lopes AM, Gomes V, et al. Forensic genetics and genomics: Much more than just a human affair. *PLoS Genet*. 2017;13(9):e1006960. <https://doi.org/10.1371/journal.pgen.1006960>.
46. Trinh P, Zaneveld JR, Safranek S, Rabinowitz PM: One health relationships between human, animal, and environmental microbiomes: a mini-review. *Front Public Health* 2018, 6(235).
47. Robinson JM, Pasternak Z, Mason CE, Elhaik E: Forensic applications of microbiomics: a review. *Front microbiol* 2021, 11(3455).
48. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015;348(6235):648–60. <https://doi.org/10.1126/science.1262110>.
49. Edqvist P-HD, Fagerberg L, Hallström BM, Danielsson A, Edlund K, Uhlen M, et al. Expression of human skin-specific genes defined by transcriptomics and antibody-based profiling. *J Histochem Cytochem*. 2015;63(2):129–41. <https://doi.org/10.1369/0022155414562646>.
50. Polak P, Karlič R, Koren A, Thurman R, Sandstrom R, Lawrence MS, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature*. 2015;518(7539):360–4. <https://doi.org/10.1038/nature14221>.
51. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJM, et al. Identifying personal microbiomes using metagenomic codes. *Proc Natl Acad Sci USA*. 2015;112(22):E2930–8. <https://doi.org/10.1073/pnas.1423854112>.
52. Mar Rodríguez M, Pérez D, Javier Chaves F, Esteve E, Marin-García P, Xifra G, et al. Obesity changes the human gut mycobiome. *Sci Rep*. 2015;5(1):14600. <https://doi.org/10.1038/srep14600>.
53. Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, et al. Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metab*. 2015;22(2):320–31. <https://doi.org/10.1016/j.cmet.2015.07.001>.
54. Santacruz A, Collado MC, Garcia-Valdes L, Segura M, Martin-Lagos J, Anjos T, et al. Gut microbiota composition is associated with body weight, weight gain and biochemical parameters in pregnant women. *Br J Nutr*. 2010; 104(1):83–92. <https://doi.org/10.1017/S00007114510000176>.
55. Edwards SM, Cunningham SA, Dunlop AL, Corwin EJ. The maternal gut microbiome during pregnancy. *MCN Am J Matern Child Nurs*. 2017;42(6): 310–7. <https://doi.org/10.1097/NMC.0000000000000372>.
56. Wang C-Y, Yeh J-K, Shie S-S, Hsieh I-C, Wen M-S. Circadian rhythm of RNA N6-methyladenosine and the role of cryptochrome. *Biochem Biophys Res Commun*. 2015;465(1):88–94. <https://doi.org/10.1016/j.bbrc.2015.07.135>.
57. Chen C, Hemme C, Beleno J, Shi ZJ, Ning D, Qin Y, et al. Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *ISME J*. 2018;12(5):1210–24. <https://doi.org/10.1038/s41396-017-0037-1>.
58. Fan X, Alekseyenko AV, Wu J, Peters BA, Jacobs EJ, Gapstur SM, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. *Gut*. 2018;67(1):120–7. <https://doi.org/10.1136/gutjnl-2016-312580>.
59. Lichinchi G, Gao S, Saletore Y, Gonzalez GM, Bansal V, Wang Y, et al. Dynamics of the human and viral m6A RNA methylomes during HIV-1 infection of T cells. *Nature Microbiol*. 2016;1(4):16011. <https://doi.org/10.1038/nmicrobiol.2016.11>.
60. Pena A: Company uses aerial footage technology to fight crime. In: *CBS News*. <https://www.cbsnews.com/news/company-uses-aerial-footage-technology-to-fight-crime/> (Last Accessed 8 Nov 2019); 2015.

61. Thompson SA, Warzel C: Twelve million phones, one dataset, zero privacy. In: <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html> (Last Accessed 5 Mar 2020); The New York Times; 2019.
62. Carpenter v. United States, 585 U.S. (2018). <https://supreme.justia.com/cases/federal/us/585/16-402/>. Accessed 10 May 2021.
63. Duhigg C: How Companies Learn Your Secrets. In: New York Times. <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> (Last Accessed 4 Nov 2019); 2012.
64. Frumkin D, Wasserstrom A, Davidson A, Grafit A. Authentication of forensic DNA samples. *Forensic Sci Int Genet.* 2010;4(2):95–103. <https://doi.org/10.1016/j.fsigen.2009.06.009>.
65. Hillis WD, Myhrvold NP, Wilson R. System for obfuscating identity. In: Google Patents; 2015.
66. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science.* 2014;346(6213):1258096. <https://doi.org/10.1126/science.1258096>.
67. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol.* 2015;33(5):510–7. <https://doi.org/10.1038/nbt.3199>.
68. Barbaro A, Cormaci P, Teatino A, La Marca A, Barbaro A. Anonymous letters? DNA and fingerprints technologies combined to solve a case. *Forensic Sci Int.* 2004;146:S133–4. <https://doi.org/10.1016/j.forsciint.2004.09.039>.
69. Lowe v. Atlas Logistics Group Retail Services. In: F Supp 3d. vol. 102: Dist. Court, ND Georgia; 2015: 1360.
70. 111th Congress: Patient Protection and Affordable Care Act. In: Compilation of Patient Protection and Affordable Care Act. <http://housedocs.house.gov/energycommerce/ppacacon.pdf> (Last Accessed 16 Feb 2021); 2010.
71. California State Senate: SB-559 Discrimination: genetic information. In. http://www.leginfo.ca.gov/pub/11-12/bill/sen/sb_0551-0600/sb_559_bill_20110906_chaptered.pdf (Last Accessed 7 Nov 2019); 2011.
72. California State Senate: SB-980 California Genetic Information Privacy Act ("GIPA"). In. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=2019202005B980 (Last Accessed 17 Feb 2021); 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

