Enjeux de confidentialité à l'ère des sciences « omiques » et de la recherche ouverte (open science)

Semaine de la recherche responsable 3e édition

Faculté de Médecine et des Sciences de la Santé (FMSS) de l'Université de Sherbrooke

CIUSSS de l'Estrie-CHUS et Centre de recherche Charles-Le Moyne

7 mai 2024

Charles Dupras, PhD

Professeur adjoint et responsable, Programmes de bioéthique

Département de médecine sociale et préventive (DMSP)

École de santé publique de l'Université de Montréal (ESPUM)

charles.dupras@umontreal.ca

THE AMERICAN JOURNAL OF BIOETHICS
2021, VOL. 21, NO. 12, 46–64
https://doi.org/10.1080/15265161.2020.1863516

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases

Charles Dupras[a] and Eline M. Bunnik[b]

[a]McGill University; [b]Erasmus MC

ABSTRACT

While the accumulation and increased circulation of genomic data have captured much attention over the past decade, privacy risks raised by the *diversification* and *integration* of omics have been largely overlooked. In this paper, we propose the outline of a framework for assessing privacy risks in multi-omic research and databases. Following a comparison of privacy risks associated with genomic and epigenomic data, we dissect ten privacy risk-impacting omic data properties that affect either the risk of re-identification of research participants, or the sensitivity of the information potentially conveyed by biological data. We then propose a three-step approach for the assessment of privacy risks in the multi-omic era. Thus, we lay grounds for a data property-based, 'pan-omic' approach that moves away from genetic exceptionalism. We conclude by inviting our peers to refine these theoretical foundations, put them to the test in their respective fields, and translate our approach into practical guidance.

KEYWORDS

Confidentiality & privacy; genetic research; human subjects research; IRB (Institutional Review Board); research ethics

INTRODUCTION

Over the past twenty years, the number and size of biobanks and databases set up for the purposes of biological and health research have increased exponentially. Their content has not only exploded quantitatively; it has also diversified qualitatively. Today, both public and private repositories contain massive amounts of genomic data about individuals from various countries and ranging from single nucleotide polymorphism to whole-genome sequenc-

To study associations and causal relationships between different omics, researchers not only collect these complementary data types, they also routinely merge them into multi-omic databases and computation systems, allowing them to perform increasingly sophisticated integrative analyses (Creanza et al. 2015; Blekhman et al. 2015; Hasin, Seldin and Lusi 2017; Yang et al. 2018; Karimi et al. 2018). Integrative single-cell analysis, for instance, aims to study different omics systems simultaneously to provide more accur-

Exceptionalisme génétique

Oui, l'information génétique possède certaines propriétés qui méritent une attention particulière. Mais rien qui ne la fasse entrer dans un univers unique de préoccupations morales, juridiques et politiques.

- Murray, T. H. (2019) *Hastings Center Report*

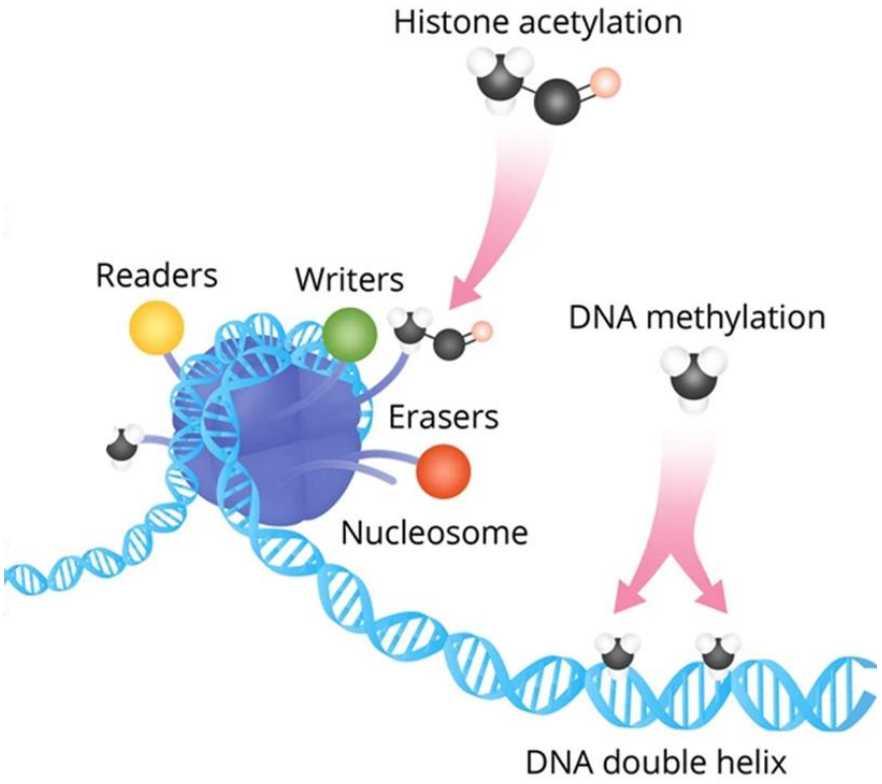
Bien que la plupart des commentateurs aient critiqué l'exceptionnalisme génétique, la quasi-totalité de la législation récente ... a été spécifique à la génétique.

- Wright Clayton et al. (2019) Journal of Law and the Biosciences

Épigénétique

"l'étude des changements dans la fonction des gènes, qui sont … héritables, mais qui n'impliquent pas de changement de la sequence d'ADN."

- Wu and Morris (2001)



Smigielski et al. (2020) *Molecular Psychiatry*

Environnement physique/chimique (toxiques, pollution)
Environnement social (traumatisme, stress, culture)
Comportement et style de vie (alimentation, consommation)
Diversité génétique (entre individus et groupes)

PRÉVENTION

Variants épigénétiques

(entre cellules/tissus, individus et groupes)

TRAITEMENT

Risques épigénétiques — PRÉDICTION
Signatures épigénétiques — DIAGNOSTIC

Héritabilité transgénérationnelle
(de variants acquis)

Les horloges épigénétiques

Âge chronologique

Les horloges épigénétiques sont *plus précises* que les horloges à télomères et peuvent prédire l'âge 'calendrier' d'une personne, lorsqu'il est incertain ou non divulgué.

- **Hannum clock:** 71 sites CpG; sang
- **Horvath clock :** 353 sites CpG; pan-tissulaire

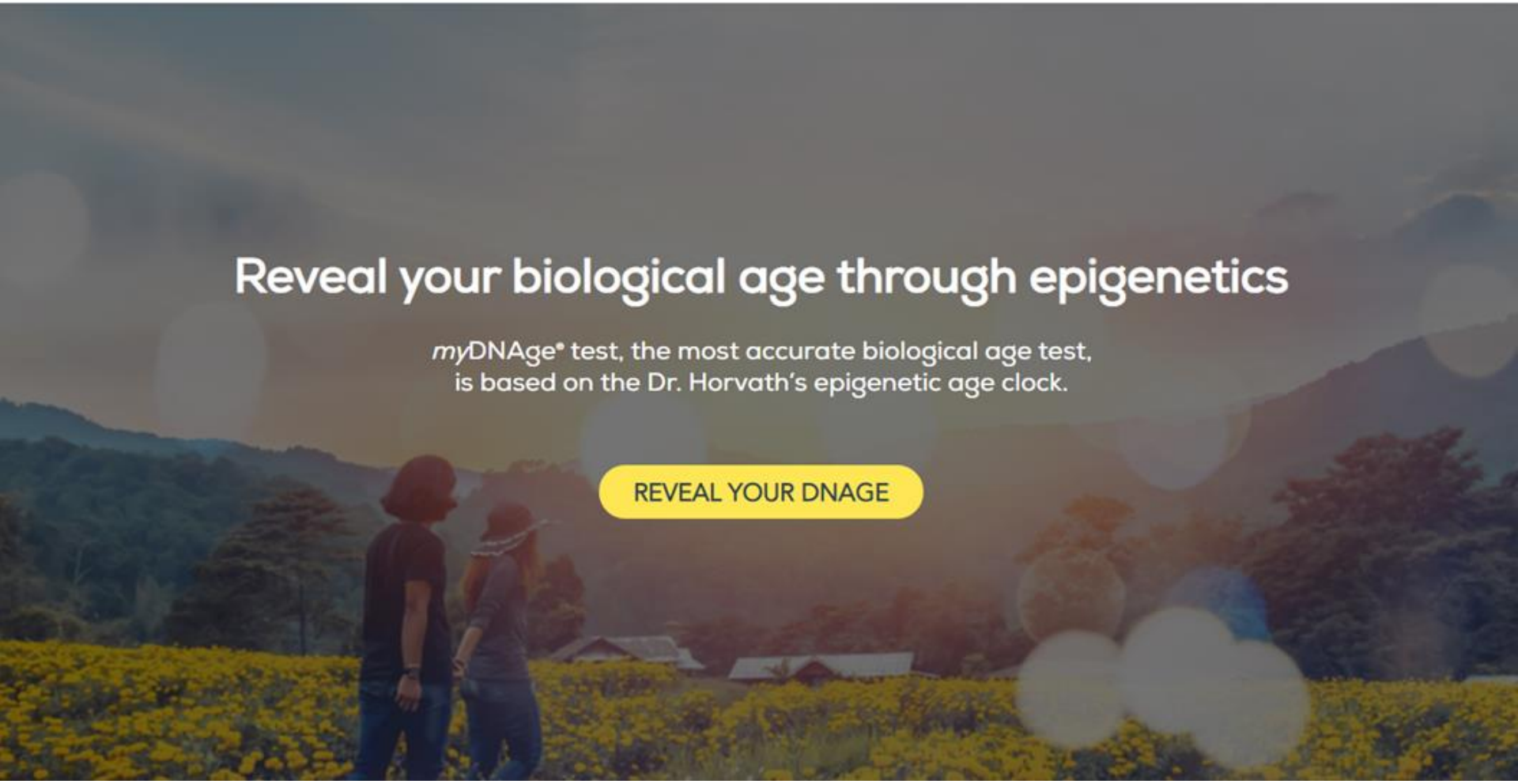
Âge biologique

Les horloges épigénétiques pourraient également être utiles pour évaluer la variation de l'espérance de vie à la suite de l'accélération ou de la décélération de l'âge biologique d'une personne.

- **PhenoAge clock:** 513 sites CpG (liés 9 phénotypes), sang
- **GrimAge clock:** 172 sites CpG (liés exposition), sang

myDNAge

ACTIVATE KIT CLINIC LEARN BLOG CONTACT US LOG IN ORDER



START YOUR JOURNEY

From start to finish

Order a kit

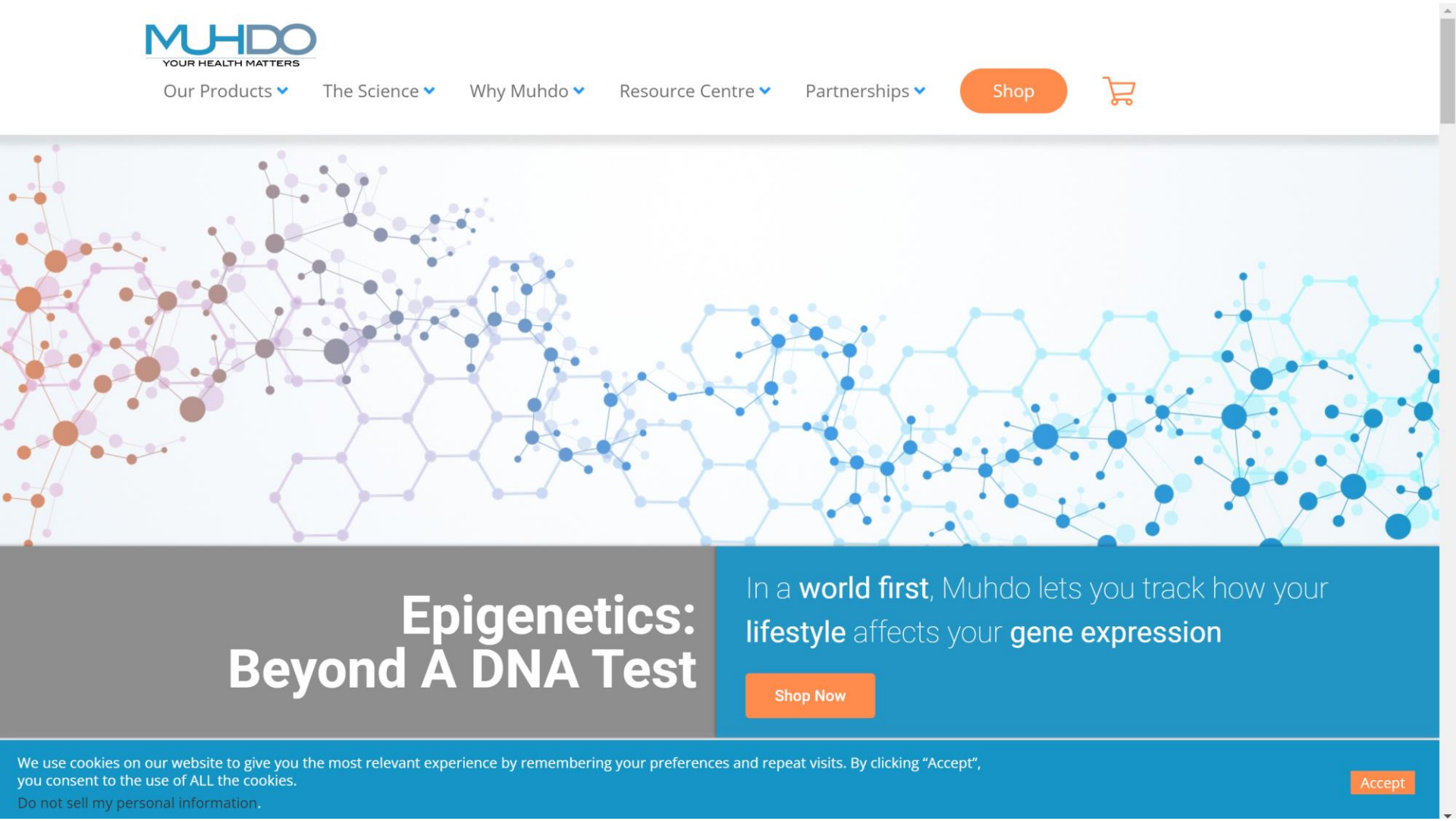
Choose your preferred sample collection kit, blood or urine. The collection device will arrive within a week.

Send us your sample

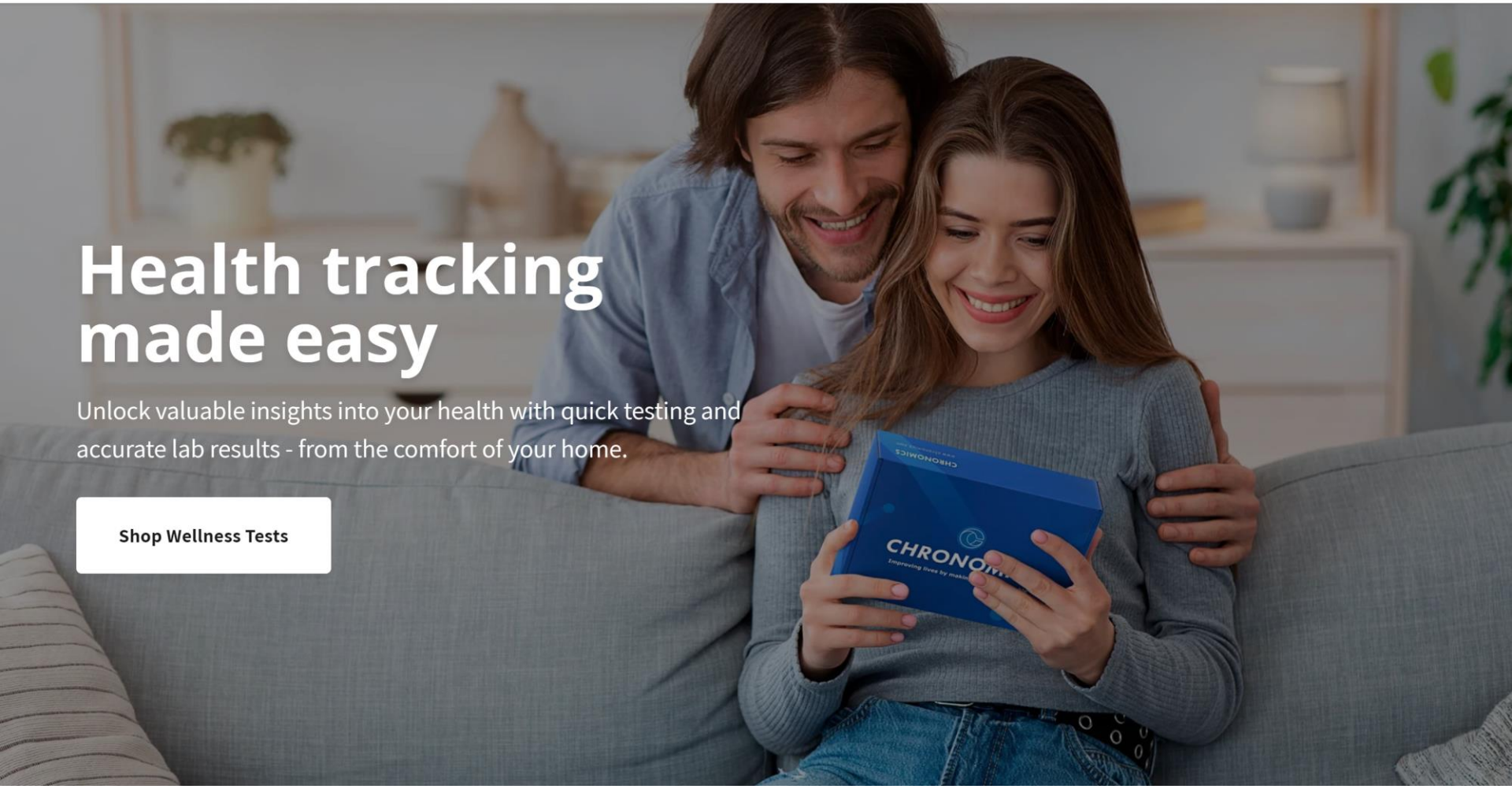
Follow the kit's instructions to collect your blood or urine sample. Mail it back using the provided prepaid return label.

Reveal your DNAge

Your result will be ready in 4-6 weeks. The report will be delivered to you by email. You can track your DNAge® changes if you take multiple tests.


MUHDO
YOUR HEALTH MATTERS
Our Products
The Science
Why Muhdo
Resource Centre
Partnerships
Shop
Epigenetics:
Beyond A DNA Test
In a world first, Muhdo lets you track how your lifestyle affects your gene expression
Shop Now
We use cookies on our website to give you the most relevant experience by remembering your preferences and repeat visits. By clicking "Accept", you consent to the use of ALL the cookies.
Do not sell my personal information.
Accept


EPIGENETIC MODULATION
DIET
MICROBIOME
EXERCISE
ALTERNATIVE MEDICINE
DISEASE EXPOSURE
THERAPEUTIC DRUGS
TOXIC CHEMICALS
PSYCHOLOGICAL STATE
DRUG ABUSE
FINANCIAL STATUS



Health tracking made easy

Unlock valuable insights into your health with quick testing and accurate lab results - from the comfort of your home.

Shop Wellness Tests

	Novembre 2018	Juillet 2019	Mars 2020
Épigénétique	Chronomics epigenCare	Chronomics epigenCare Muhdo myDNAge TruMe	Chronomics epigenCare Muhdo myDNAge TruMe Inside Tracker Elysium Health Epigenetics Experts TruDiagnostics
Microbiomique	uBiome	Atlas Biomed Carbiotix Ixcela Thryve uBioDiscovery uBiome Viome	Atlas Biomed Carbiotix Ixcela Thryve uBioDiscovery Viome Join Zoe Biohm DayTwo BiomeSight Microba

Knoppers et al. (2021) *New Genetics & Society*

consumer epigenetics

- medical relevance
- information sensitivity
- data protection
- secondary use

Dupras, Beauchamp & Joly (2020) *Nature Reviews Genetics*

the « omics of our lives »

- diversifying industry
- lifestyle biomonitoring
- terms and conditions readibility
- no standard of practice

Knoppers et al. (2021) *New Genetics & Society*

Confidentialité et utilisation secondaire

“les informations non personnellement identifiables ou agrégées peuvent être utilisées par nous à toutes les fins autorisées par la loi”

- Ixcela

“nous pouvons divulguer toute information vous concernant au gouvernement, aux autorités chargées de l'application de la loi ou à des parties privées”

- Thryve

Knoppers et al. (2021) The omics of our lives: practices and policies of direct-to-consumer epigenetic and microbiomic testing companies. *New Genetics and Society*

consumer epigenetics

- medical relevance
- information sensitivity
- data protection
- secondary use

Dupras, Beauchamp & Joly (2020) *Nature Reviews Genetics*

the « omics of our lives »

- diversifying industry
- lifestyle biomonitoring
- terms and conditions readibility
- no standard of practice

Knoppers et al. (2021) *New Genetics & Society*

Future of Privacy Forum (2018)

Privacy Best Practices for Consumer Genetic Testing Services

23andMe, Ancestry, Helix, MyHeritage, Habit, African Ancestry, and Living DNA

- Les politiques de protection de la vie privée devraient être 'complètes, accessibles et faciles à lire'

- Le consentement spécifique est requis pour:

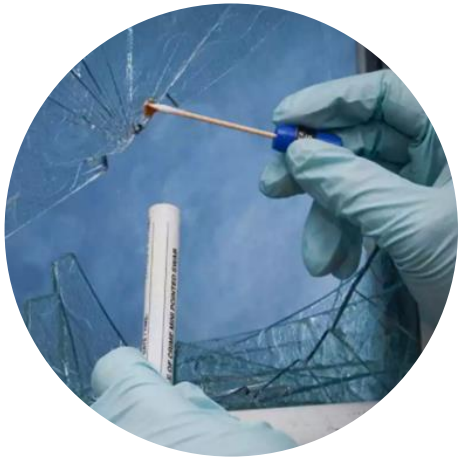
- l'utilisation des données d'une manière incompatible avec les conditions de la politique initialement prévue,
- le transfert ultérieur des données d'un seul individu,
- les utilisations en dehors de l'objectif premier du service et les utilisations contextuelles inhérentes,
- la soumission par procuration d'un échantillon à des fins d'analyse,
- le transfert à des tiers à des fins de recherche, et
- l'utilisation à des fins de recherche interne en l'absence d'approbation dans le cadre d'une procédure d'évaluation éthique ;

- Le partage des données avec les employeurs, les compagnies d'assurance, les établissements d'enseignement et les agences gouvernementales est expressément interdit, sauf si la loi l'exige ou si l'on dispose d'un consentement explicite distinct.

**** Les informations dépersonnalisées sont exclues de ces protections.***

Majumder *et al.* (2020) *Annual Review of Medicine*

Utilisations non médicales



Criminalistique

Immigration

Assurance

Dupras et al. (2019) Potential (mis)use of epigenetic age estimators (...) *Environmental Epigenetics*

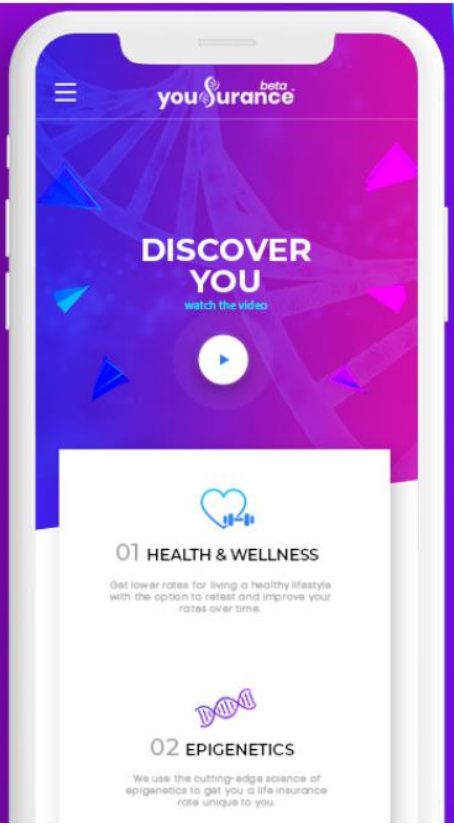
youSurance beta™

Insurance Science Agents

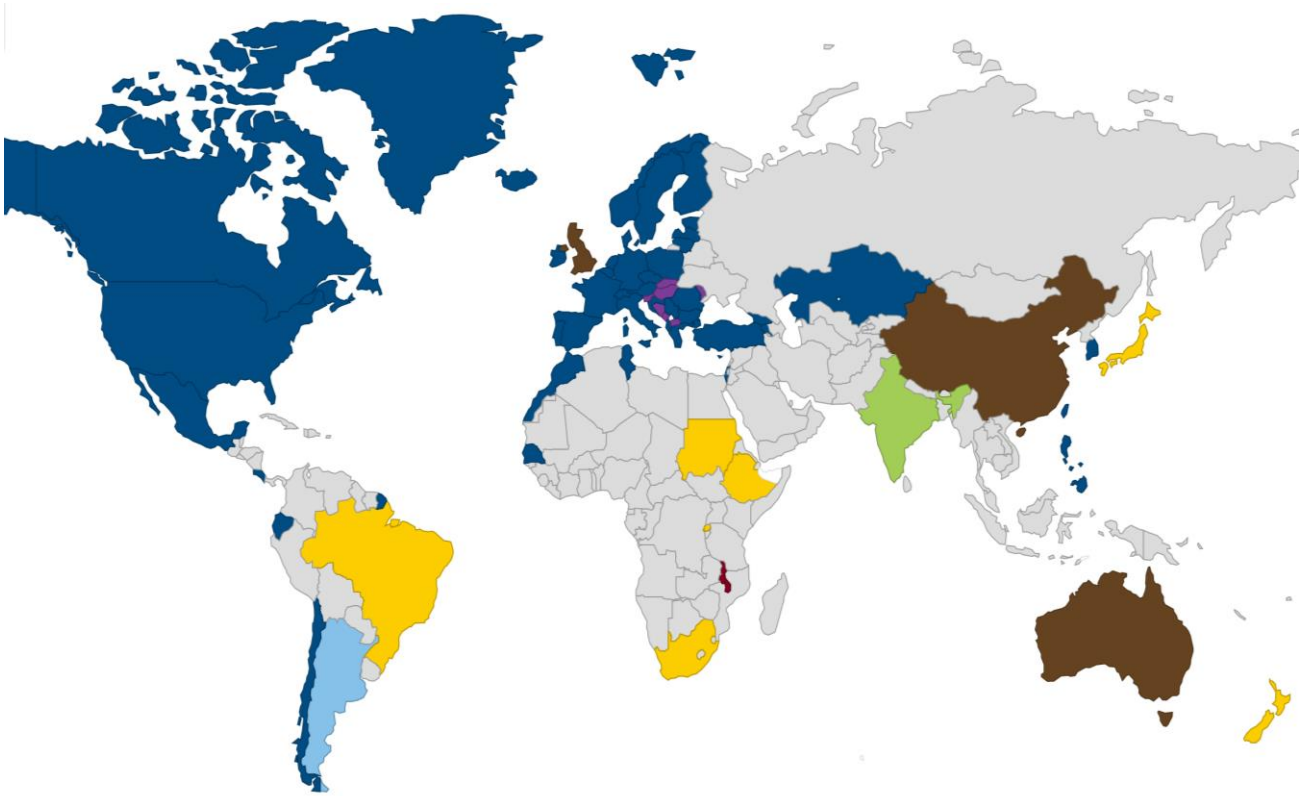
THE FUTURE OF
LIFE INSURANCE IS YOU

We use the cutting-edge science of epigenetics to get you a life insurance rate unique to you.

LEARN MORE



A GEOGRAPHICAL OVERVIEW OF APPROACHES ADOPTED AROUND THE WORLD TO PREVENT GENETIC DISCRIMINATION



frontiers in Genetics

PERSPECTIVE
published: 08 June 2018
doi: 10.3389/fgene.2018.00202

Epigenetic Discrimination: Emerging Applications of Epigenetics Pointing to the Limitations of Policies Against Genetic Discrimination

Charles Dupras, Lingqiao Song, Katie M. Saulnier and Yann Joly*

Centre of Genomics and Policy, McGill University and Genome Quebec Innovation Centre, Montreal, QC, Canada

Over more than two decades, various policies have been adopted worldwide to restrict the use of individual genetic information for non-medical reasons by third parties and prevent 'genetic discrimination'. In this paper, we bring attention to the growing interest for individual *epigenetic information* by insurers and forensic scientists. We question whether such interest could lead to 'epigenetic discrimination' – the differential adverse treatment or abusive profiling of individuals or groups based on their actual or presumed epigenetic characteristics – and argue that we might already be facing the limitations of recently adopted normative approaches against genetic discrimination. First, we highlight some similarities and differences between genetic and epigenetic modifications, and stress potential challenges to regulating epigenetic discrimination. Second, we argue that most existing normative approaches against genetic discrimination fall short in providing oversight into the field of epigenetics. We conclude with a call for discussion on the issue, and the development of comprehensive and forward-looking preventive strategies against epigenetic discrimination.

Keywords: epigenetics, DNA methylation, discrimination, insurance, forensic science, ethics, justice, policy

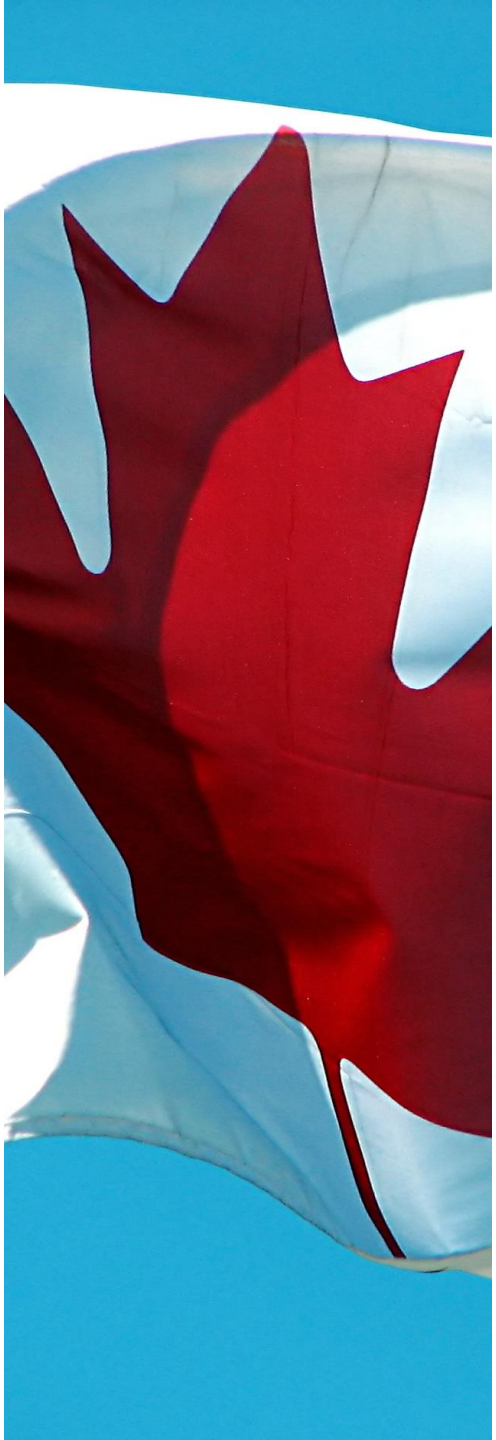
OPEN ACCESS

Edited by:
Christopher Thomas Scott,
Stanford University, United States

Reviewed by:
Maurizio Meloni,
Deakin University, Australia
Laura Specker Sullivan,
Harvard Medical School,
United States

****Correspondence:***
Charles Dupras
charles.dupras2@mcgill.ca

Specialty section:
This article was submitted to ELSI in Science and Genetics,



Loi sur la non-discrimination génétique (S.C. 2017, c. 3)

Sommaire

1. Droit de **refuser de se soumettre** à un *test génétique*

2. Droit de **refuser de divulguer les résultats** d'un *test génétique*

3. Droit de **ne pas faire l'objet d'une discrimination** dans la fourniture de biens et de services sur la base d'un refus de tester (1) ou de divulguer les résultats (2).

4. Il est interdit à toute personne engagée dans la fourniture de biens ou de services contractuels* de **collecter, d'utiliser ou de divulguer** les résultats d'un test génétique **sans le consentement écrit** de l'autre personne.

*des exceptions peuvent s'appliquer aux médecins et aux chercheurs

LIFE | Europe EXTENSION®
The Science of a Healthier Life®

Search products, blog entries, authors

Login

Shop | Learn | Brands | Experts and Reviews | Loyalty benefits

Contact us | Partners | English (€)

ANTI-AGING | NEW | HOME TESTS

EpiAge epigenetic age test EU/UK

Discover your real age based on the latest scientific and technological standards

Item #90005 | 1 Tests | In Stock

For you, if you want a precise personalized biological age check. English version for all customers outside Germany

The epiAge epigenetic age test EU can offer:

- The next-generation epigenetic age test for home use
- A user-friendly DNA analysis kit to calculate your biological age
- Analysis of sample in the laboratory of one of the leading epigeneticists
- Age-monitoring with scientific accuracy
- Effectiveness test of your personal anti-aging regimen

Age is much more than a number of chronological years. Biological years is the age of the cells that make up your body based on various properties and biomarkers that have been researched to correlate with aging. epiAge epigenetic age test EU can provide state-of-the-art DNA research embedded in a user-friendly product.

This is how it works:

1. You take a saliva test and send it to HKG, a specialized laboratory
2. The laboratory analyses your test triple times for the utmost scientific accuracy
3. After a few weeks you have the result of your biological age

Buy 1 | Buy 2 (5% off) | Buy 4 (10% off)

€199.00

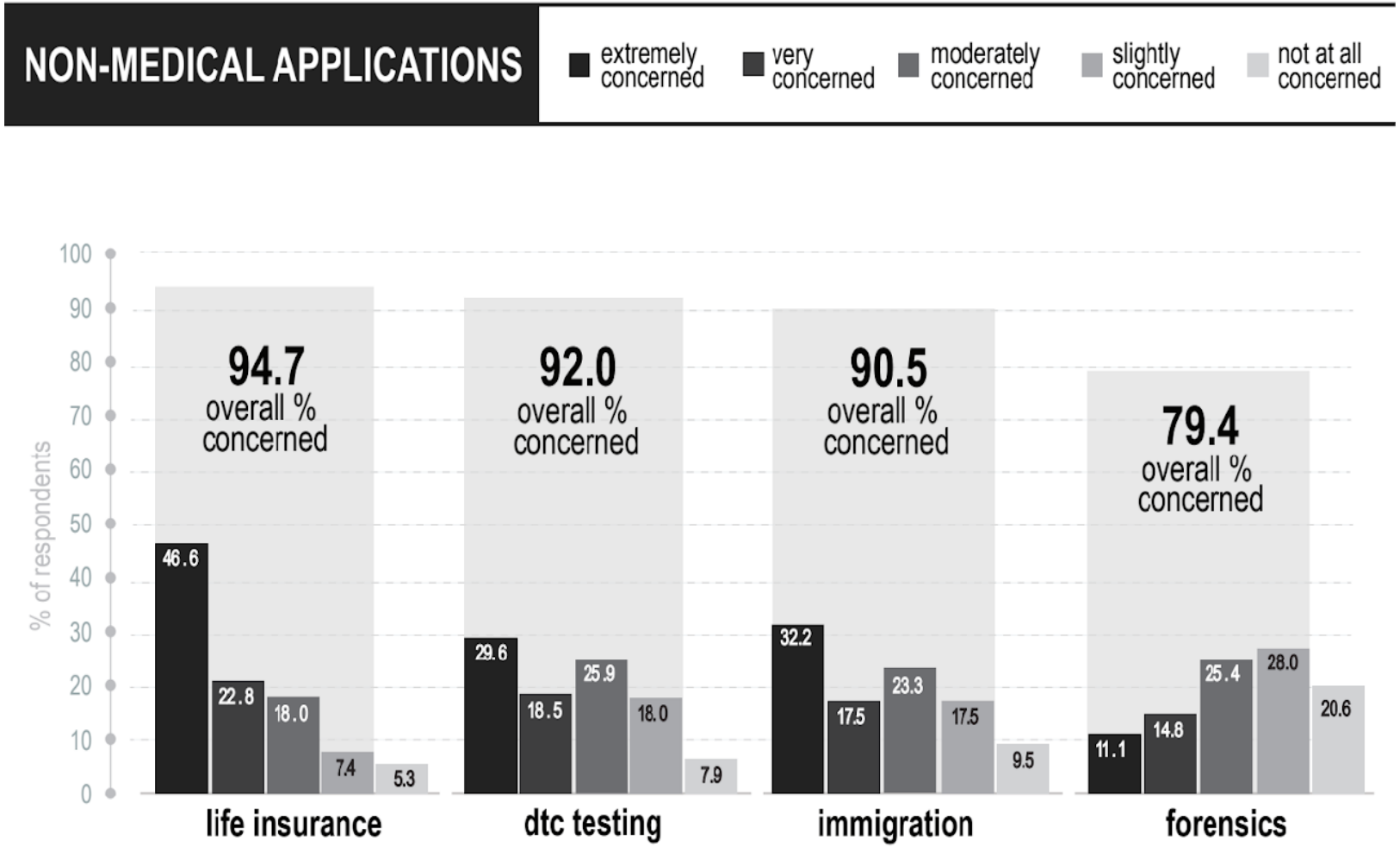
− 1 + Add to shopping cart

Free Shipping
on orders over € 69

Money Back Guarantee
We have a 100 days return policy, not the standard 30 days!

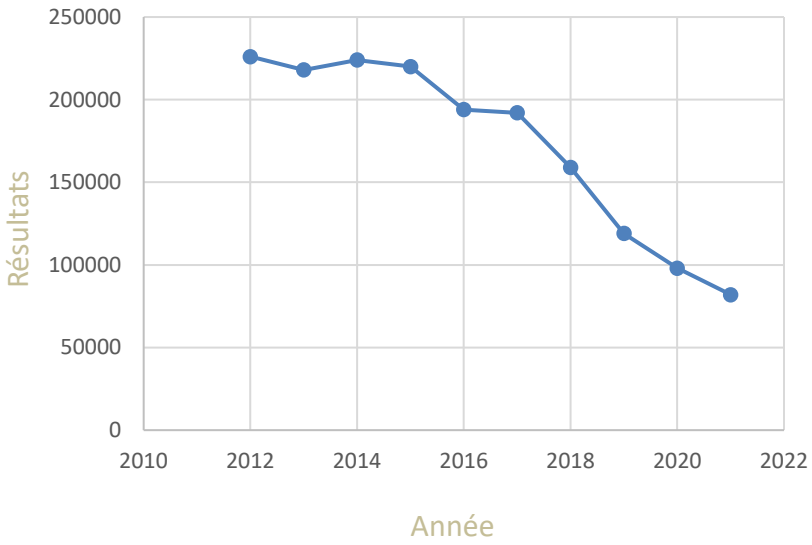
Daily Shipping
Order before 11:00 CET, for same day shipping (Mon-Fri)

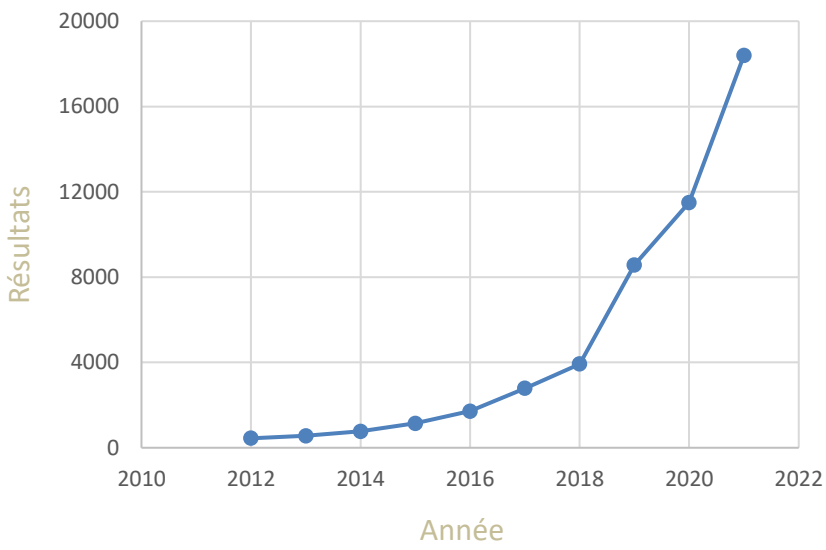
Customer Care Advisory
We can help you in 6 languages



Dupras et al. (2022) Researcher perspectives on ethics considerations in epigenetics. *Clinical Epigenetics*

À l’ère de la multi-omique



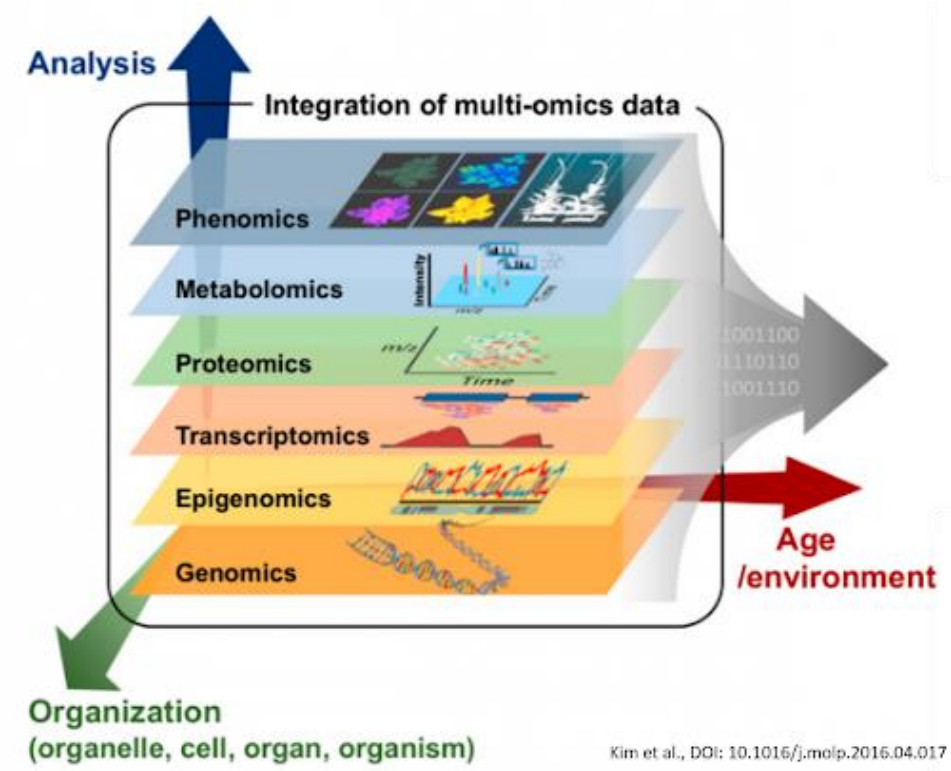


« Genomics »

« Multi-omics »

Google Scholar 2022-04-06

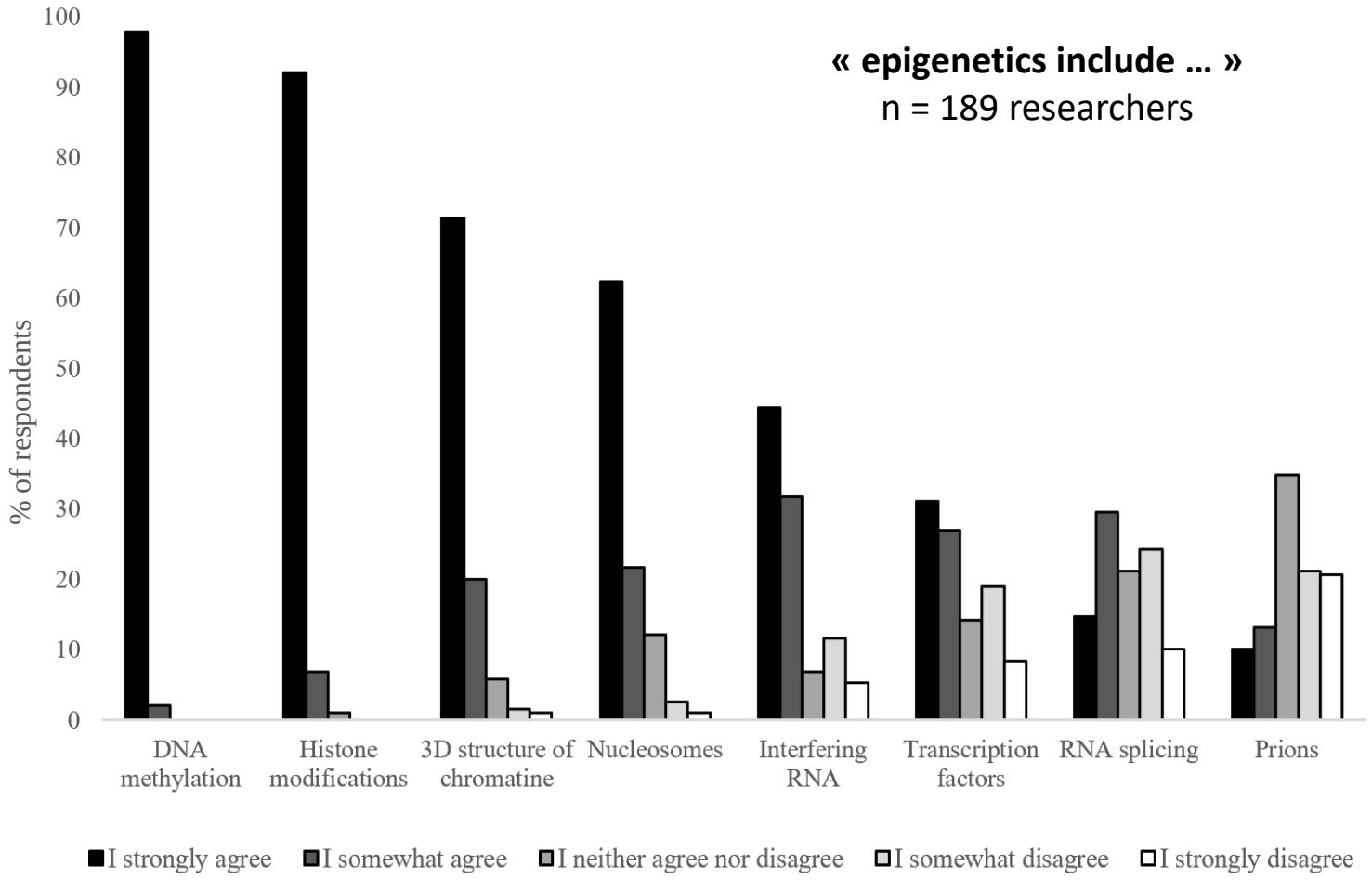
À l'ère de la multi-omique



À l’ère de la multi-omique

1. **Multijuridictionnel:** circulation internationale facilitée des données omiques
2. **Multisectoriel:** circulation entre bases de données privées et publiques
3. **Multiplication:** accumulation dans les grandes bases de données
4. **Multiplicité:** diversification des données omiques (et digitales)

Elle élargit le champ d'attention - et les devoirs - des chercheurs, des comités d'évaluation éthique de la recherche et des administrateurs de données pour y inclure la protection des données autres que génomiques.



Dupras et al. (2022) Researcher perspectives on ethics considerations in epigenetics. *Clinical Epigenetics*

Méthode d'évaluation des risques pour la vie privée

THE AMERICAN JOURNAL OF BIOETHICS
2021, VOL. 21, NO. 12, 46–64
https://doi.org/10.1080/15265161.2020.1863516

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

Check for updates

Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases

Charles Dupras[a] and Eline M. Bunnik[b]

[a]McGill University; [b]Erasmus MC

ABSTRACT

While the accumulation and increased circulation of genomic data have captured much attention over the past decade, privacy risks raised by the *diversification* and *integration* of omics have been largely overlooked. In this paper, we propose the outline of a framework for assessing privacy risks in multi-omic research and databases. Following a comparison of privacy risks associated with genomic and epigenomic data, we dissect ten privacy risk-impacting omic data properties that affect either the risk of re-identification of research participants, or the sensitivity of the information potentially conveyed by biological data. We then propose a three-step approach for the assessment of privacy risks in the multi-omic era. Thus, we lay grounds for a data property-based, 'pan-omic' approach that moves away from genetic exceptionalism. We conclude by inviting our peers to refine these theoretical foundations, put them to the test in their respective fields, and translate our approach into practical guidance.

KEYWORDS

Confidentiality & privacy; genetic research; human subjects research; IRB (Institutional Review Board); research ethics

INTRODUCTION

Over the past twenty years, the number and size of biobanks and databases set up for the purposes of biological and health research have increased exponentially. Their content has not only exploded quantitatively; it has also diversified qualitatively. Today, both public and private repositories contain massive amounts of genomic data about individuals from various countries and ranging from single nucleotide polymorphism to whole-genome sequencing data (Sudlow et al. 2015; Stoeklé et al. 2016; Canela-Xandri, Rawlik, and Tenesa 2018; Brieger et al. 2019). But repositories may also contain data related to other types of biological systems –often referred to as 'omics'– such as epigenomic, transcriptomic, proteomic, lipidomic, metabolomic, phenomic and microbiomic data (Komaki et al. 2018; Maiella et al. 2018; Cazaly et al. 2019; Vangay, Hillmann, and Knights 2019). The collection, sharing and use of these complementary data types provides an opportunity for researchers to better understand, among other things, the multiple intra- and extra-cellular variables that influence gene regulation, expression and function (Conesa and Beck 2019; Perez-Riverol et al. 2019).

To study associations and causal relationships between different omics, researchers not only collect these complementary data types, they also routinely merge them into multi-omic databases and computation systems, allowing them to perform increasingly sophisticated integrative analyses (Creanza et al. 2015; Blekhman et al. 2015; Hasin, Seldin and Lusi 2017; Yang et al. 2018; Karimi et al. 2018). Integrative single-cell analysis, for instance, aims to study different omics systems simultaneously to provide more accurate and greater information on their specific biological functions and responsiveness (Stuart and Satija 2019). To perform these complex multi-omic analyses, researchers have notably started implementing machine learning technologies (Lin and Lane 2017; Hamamoto et al. 2019).

While the accumulation (multiplication) and increased transdisciplinary, trans-sectoral, and transnational circulation of genomic data has captured much attention by ethicists, legal scholars, social scientists and policymakers over the past decades (Greely 2007; Hoeyer 2012; Kamm et al. 2013; Wang et al. 2017; Dankar, Ptitsyn, and Dankar 2018), the potential privacy issues raised specifically by the *diversification* of omics data (multiplicity), and their intended or

CONTACT Charles Dupras charles.dupras2@mcgill.ca Centre of Genomics and Policy, McGill University, 740 avenue du, Docteur-Penfield, Montreal, Quebec, H3A 0G1, Canada.

Postulat: **10 propriétés** des données augmentent les risques pour la vie privée.

Étape 1: présence des propriétés des données ayant un impact sur le **risque de réidentification**.

Étape 2: présence des propriétés des données ayant un impact sur leur **niveau de sensibilité**.

Étape 3: présence d'**effets d'interrelations** entre les types de données

- Effet de corrélation
- Effet de synergie

Dupras & Bunnik (2021) Toward a framework for assessing privacy risks in multi-omic research and databases. *American Journal of Bioethics*

Méthode d'évaluation des risques pour la vie privée

THE AMERICAN JOURNAL OF BIOETHICS
2021, VOL. 21, NO. 12, 46–64
https://doi.org/10.1080/15265161.2020.1863516

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

Check for updates

Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases

Charles Dupras[a] and Eline M. Bunnik[b]

[a]McGill University; [b]Erasmus MC

ABSTRACT

While the accumulation and increased circulation of genomic data have captured much attention over the past decade, privacy risks raised by the *diversification* and *integration* of omics have been largely overlooked. In this paper, we propose the outline of a framework for assessing privacy risks in multi-omic research and databases. Following a comparison of privacy risks associated with genomic and epigenomic data, we dissect ten privacy risk-impacting omic data properties that affect either the risk of re-identification of research participants, or the sensitivity of the information potentially conveyed by biological data. We then propose a three-step approach for the assessment of privacy risks in the multi-omic era. Thus, we lay grounds for a data property-based, 'pan-omic' approach that moves away from genetic exceptionalism. We conclude by inviting our peers to refine these theoretical foundations, put them to the test in their respective fields, and translate our approach into practical guidance.

KEYWORDS

Confidentiality & privacy; genetic research; human subjects research; IRB (Institutional Review Board); research ethics

INTRODUCTION

Over the past twenty years, the number and size of biobanks and databases set up for the purposes of biological and health research have increased exponentially. Their content has not only exploded quantitatively; it has also diversified qualitatively. Today, both public and private repositories contain massive amounts of genomic data about individuals from various countries and ranging from single nucleotide polymorphism to whole-genome sequencing data (Sudlow et al. 2015; Stoeklé et al. 2016; Canela-Xandri, Rawlik, and Tenesa 2018; Brieger et al. 2019). But repositories may also contain data related to other types of biological systems –often referred to as 'omics'– such as epigenomic, transcriptomic, proteomic, lipidomic, metabolomic, phenomic and microbiomic data (Komaki et al. 2018; Maiella et al. 2018; Cazaly et al. 2019; Vangay, Hillmann, and Knights 2019). The collection, sharing and use of these complementary data types provides an opportunity for researchers to better understand, among other things, the multiple intra- and extra-cellular variables that influence gene regulation, expression and function (Conesa and Beck 2019; Perez-Riverol et al. 2019).

To study associations and causal relationships between different omics, researchers not only collect these complementary data types, they also routinely merge them into multi-omic databases and computation systems, allowing them to perform increasingly sophisticated integrative analyses (Creanza et al. 2015; Blekhman et al. 2015; Hasin, Seldin and Lusi 2017; Yang et al. 2018; Karimi et al. 2018). Integrative single-cell analysis, for instance, aims to study different omics systems simultaneously to provide more accurate and greater information on their specific biological functions and responsiveness (Stuart and Satija 2019). To perform these complex multi-omic analyses, researchers have notably started implementing machine learning technologies (Lin and Lane 2017; Hamamoto et al. 2019).

While the accumulation (multiplication) and increased transdisciplinary, trans-sectoral, and trans-national circulation of genomic data has captured much attention by ethicists, legal scholars, social scientists and policymakers over the past decades (Greely 2007; Hoeyer 2012; Kamm et al. 2013; Wang et al. 2017; Dankar, Ptitsyn, and Dankar 2018), the potential privacy issues raised specifically by the *diversification* of omics data (multiplicity), and their intended or

CONTACT Charles Dupras charles.dupras2@mcgill.ca Centre of Genomics and Policy, McGill University, 740 avenue du, Docteur-Penfield, Montreal, Quebec, H3A 0G1, Canada

Postulat: 10 propriétés des données augmentent les risques pour la vie privée.

Étape 1: présence des propriétés des données ayant un impact sur le **risque de réidentification**.

Étape 2: présence des propriétés des données ayant un impact sur leur **niveau de sensibilité**.

Étape 3: présence d'**effets d'interrelations** entre les types de données

- Effet de corrélation
- Effet de synergie

Dupras & Bunnik (2021) Toward a framework for assessing privacy risks in multi-omic research and databases. *American Journal of Bioethics*

Table 1. Summary of omic data properties which may impact privacy risks.

Privacy-relevant omic data properties	Yes/No	Step 1: impact on identifying power	Step 2: impact on the level of data sensitivity
i. Does it convey **observable** phenotypic information?	Yes	Identifying power increases (e.g., if facial *traits* are revealed)	
ii. Is it **acquired** through the life course?	Yes	Identifying power increases (e.g., if either active or passive *exposures* are revealed)	May convey information about a person's *life*[1] (e.g., history of exposures)
iii. Is it **stable** over time and the life course?	Yes	Identifying power increases by *persisting* over time	Risk of *deterministic* interpretations
iv. Is it a **rare** (combination of) variant(s)?	Yes	Identifying power increases with *distinctiveness* (e.g., unique to an individual)	Risk of *stigmatization* and/or *discrimination*
v. Is it (partly) **shared**[2] by members of some specific groups?	Yes	Identifying power increases with potential to link an individual's data to the data of another group member (e.g., a family member)	The information conveyed may be sensitive for many persons, have implication *for third parties,* and lead to stigmatization/ discrimination of groups
vi. Is it **ubiquitous** among cell types and tissues	Yes	High ubiquity (e.g. genotype ubiquity) can make it possible to link data from a single person but obtained using different sample sources (i.e. different types of cells or tissues)	High ubiquity (e.g. genotype ubiquity) increases the likelihood that sensitive information (e.g. high risk to breast cancer) be revealed by cell types or tissues which are neither functionally related to the particular information (e.g., saliva sample), nor the primary research object.
vii. Is it conceived as **abnormal**[3]?	Yes		Can convey *medical* information, or other *disreputable* information
viii. Does it convey some **predictive** power?	Yes		Psychological/socioeconomic impact, risk of unfair treatment of asymptomatic persons (i.e., biological *profiling*)
ix. Is it determined (in part) by acts or behaviors conceived as **willful**[4]?	Yes		Persons or institutions may be *blamed* for harming or *pressured* not to harm themselves and/or others.
x. Is the information provided by the data **actionable**[5]?	Yes		*Pressure* may be imposed on persons or institutions to act upon the at-risk variant; potential for attribution of both prospective and retrospective responsibility

Méthode d'évaluation des risques pour la vie privée

THE AMERICAN JOURNAL OF BIOETHICS
2021, VOL. 21, NO. 12, 46–64
https://doi.org/10.1080/15265161.2020.1863516

Taylor & Francis
Taylor & Francis Group

TARGET ARTICLE

Check for updates

Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases

Charles Dupras[a] and Eline M. Bunnik[b]

[a]McGill University; [b]Erasmus MC

ABSTRACT

While the accumulation and increased circulation of genomic data have captured much attention over the past decade, privacy risks raised by the *diversification* and *integration* of omics have been largely overlooked. In this paper, we propose the outline of a framework for assessing privacy risks in multi-omic research and databases. Following a comparison of privacy risks associated with genomic and epigenomic data, we dissect ten privacy risk-impacting omic data properties that affect either the risk of re-identification of research participants, or the sensitivity of the information potentially conveyed by biological data. We then propose a three-step approach for the assessment of privacy risks in the multi-omic era. Thus, we lay grounds for a data property-based, 'pan-omic' approach that moves away from genetic exceptionalism. We conclude by inviting our peers to refine these theoretical foundations, put them to the test in their respective fields, and translate our approach into practical guidance.

KEYWORDS

Confidentiality & privacy; genetic research; human subjects research; IRB (Institutional Review Board); research ethics

INTRODUCTION

Over the past twenty years, the number and size of biobanks and databases set up for the purposes of biological and health research have increased exponentially. Their content has not only exploded quantitatively; it has also diversified qualitatively. Today, both public and private repositories contain massive amounts of genomic data about individuals from various countries and ranging from single nucleotide polymorphism to whole-genome sequencing data (Sudlow et al. 2015; Stoeklé et al. 2016; Canela-Xandri, Rawlik, and Tenesa 2018; Brieger et al. 2019). But repositories may also contain data related to other types of biological systems –often referred to as 'omics'– such as epigenomic, transcriptomic, proteomic, lipidomic, metabolomic, phenomic and microbiomic data (Komaki et al. 2018; Maiella et al. 2018; Cazaly et al. 2019; Vangay, Hillmann, and Knights 2019). The collection, sharing and use of these complementary data types provides an opportunity for researchers to better understand, among other things, the multiple intra- and extra-cellular variables that influence gene regulation, expression and function (Conesa and Beck 2019; Perez-Riverol et al. 2019).

To study associations and causal relationships between different omics, researchers not only collect these complementary data types, they also routinely merge them into multi-omic databases and computation systems, allowing them to perform increasingly sophisticated integrative analyses (Creanza et al. 2015; Blekhman et al. 2015; Hasin, Seldin and Lusi 2017; Yang et al. 2018; Karimi et al. 2018). Integrative single-cell analysis, for instance, aims to study different omics systems simultaneously to provide more accurate and greater information on their specific biological functions and responsiveness (Stuart and Satija 2019). To perform these complex multi-omic analyses, researchers have notably started implementing machine learning technologies (Lin and Lane 2017; Hamamoto et al. 2019).

While the accumulation (multiplication) and increased transdisciplinary, trans-sectoral, and transnational circulation of genomic data has captured much attention by ethicists, legal scholars, social scientists and policymakers over the past decades (Greely 2007; Hoeyer 2012; Kamm et al. 2013; Wang et al. 2017; Dankar, Ptitsyn, and Dankar 2018), the potential privacy issues raised specifically by the *diversification* of omics data (multiplicity), and their intended or

CONTACT Charles Dupras charles.dupras2@mcgill.ca Centre of Genomics and Policy, McGill University, 740 avenue du, Docteur-Penfield, Montreal, Quebec, H3A 0G1, Canada.

Postulat: **10 propriétés** des données augmentent les risques pour la vie privée.

Étape 1: présence des propriétés des données ayant un impact sur le **risque de réidentification**.

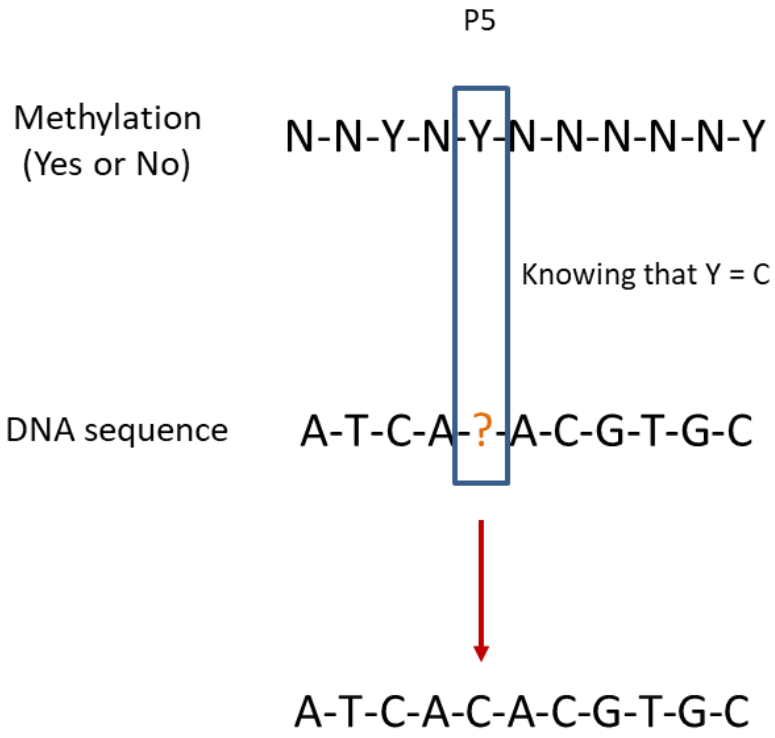
Étape 2: présence des propriétés des données ayant un impact sur leur **niveau de sensibilité**.

Étape 3: présence d'**effets d'interrelations** entre les types de données

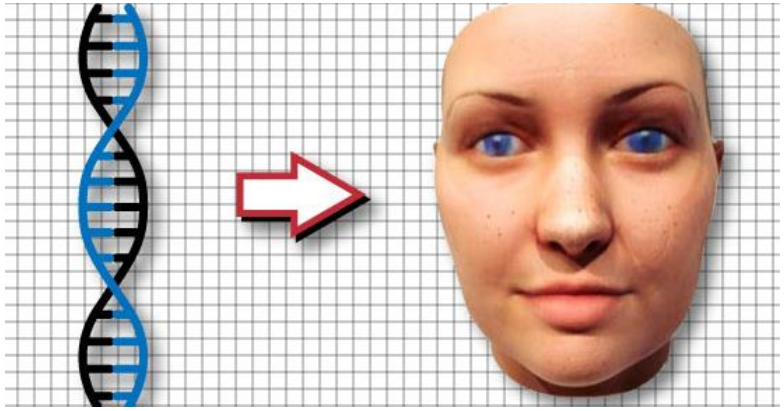
- Effet de corrélation
- Effet de synergie

Dupras & Bunnik (2021) Toward a framework for assessing privacy risks in multi-omic research and databases. *American Journal of Bioethics*

Effet de corrélation



Effet de synergie



Phénotypage ADN

Phénotypage multi-omique

THE AMERICAN JOURNAL OF BIOETHICS
2022, VOL. 22, NO. 9, W4–W6
https://doi.org/10.1080/15265161.2022.2105436

Taylor & Francis
Taylor & Francis Group

CORRESPONDENCE

Response to Open Peer Commentaries on Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases

Charles Dupras[a] and Eline M. Bunnik[b]

[a]Université de Montréal; [b]Erasmus MC

In 'Toward a Framework for Assessing Privacy Risks in Multi-Omic Research and Databases' (Dupras and Bunnik 2021), we argued against the assessment of privacy risks and protection requirements based on broad biological data types. More specifically, we questioned the assumption that genomic data generally deserves greater caution than other omic data types. Rather, we argued, it is the presence or absence of privacy-relevant *data properties*—and their specific combination—that affect the level of risk and call for more or less elaborate privacy protection strategies. Privacy-relevant properties are not unique to genomic data; many are shared across various data types (cf. epigenomics, microbiomics, transcriptomics, proteomics, lipidomics, metabolomics, neuromics, phenomics, exposomics).

Following an analysis of the similarities and dissimilarities between genomic and epigenomic data, we identified ten properties that may increase risk of re-identification using the data and/or the level of sensi-

DIGNITY, PRIVACY AND CONFIDENTIALITY RISKS

Safarlou et al. and Alex and Winkler (Alex and Winkler 2021) observe that we did not make explicit the conception of privacy inherent in our framework. They remind us that privacy concerns are not limited to information concealment and data protection by third parties, but also include crucial questions about data ownership and control by the persons to whom the information pertains. Similarly, Dorrington et al. propose moving away from an approach to privacy that focuses on concealment of personal data toward an approach that situates ownership or control with individuals who provide omic data in research settings. They write: "In the context of multi-omics research and databases, we believe that upholding the dignity of potential participants requires more than privacy protection, regardless of how thorough the assessment of those privacy risks." They argue that individuals should themselves control "when, for how long, with whom, and for what purpose any portion

Financement


National Defence


Contact

Charles Dupras, PhD
Professeur adjoint et responsable, Programmes de bioéthique
Département de médecine sociale et préventive (DMSP)
École de santé publique de l'Université de Montréal (ESPUM)
charles.dupras@umontreal.ca