

<http://www.nature.com/news/reproducibility-1.17552>; <https://fi1000research.com/gateways/PRR>). The newly launched International MAQC Society will strive to work with various scientific communities to develop consensus on best practices for enhanced reproducibility in generation, analysis, and interpretation of massive data from increasingly innovative biomedical fields. More information about the MAQC Society can be found at <http://www.maqcsociety.org>.

ACKNOWLEDGMENTS

The first MAQC Society meeting was supported by the SAS Institute, Burroughs Wellcome Fund, and the National 863 Program of China (2015AA020104, to L.S.).

DISCLAIMER

The views presented in this article do not necessarily reflect current or future opinion or policy of the US Food and Drug Administration. Any mention of commercial products is for clarification and not intended as endorsement.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Leming Shi^{1,2}, Rebecca Kusko³,
Russell D Wolfinger⁴,
Benjamin Haibe-Kains⁵⁻⁷, Matthias Fischer^{8,9},
Susanna-Assunta Sansone¹⁰,
Christopher E Mason¹¹, Cesare Furlanello¹²,
Wendell D Jones¹³, Baitang Ning¹⁴ &
Weida Tong¹⁴

¹State Key Laboratory of Genetic Engineering, School of Life Sciences and Shanghai Cancer Center, Fudan University, Shanghai, China.

²Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China. ³Immuneering Corporation, Cambridge, Massachusetts, USA. ⁴SAS Institute Inc., Cary, North Carolina, USA. ⁵Princess Margaret Cancer Centre, Toronto, Ontario, Canada. ⁶Departments of Medical Biophysics and Computer Science, University of Toronto, Toronto, Ontario, Canada. ⁷Ontario Cancer Research Institute, Toronto, Ontario, Canada. ⁸Department of Experimental Pediatric Oncology, University Children's Hospital of Cologne, Medical Faculty, Cologne, Germany. ⁹Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. ¹⁰Oxford e-Research Centre, Engineering Science Department, University of Oxford, Oxford, UK. ¹¹Weill Cornell Medicine, New York, New York, USA. ¹²Fondazione Bruno Kessler, Trento, Italy. ¹³Q2 Solutions, Morrisville, North Carolina, USA. ¹⁴National Center for Toxicological Research, US Food and Drug Administration, Jefferson, Arkansas, USA. e-mail: weida.tong@fda.hhs.gov.

- Shi, L. *et al. Nat. Biotechnol.* **24**, 1151–1161 (2006).
- Shi, L. *et al. Nat. Biotechnol.* **28**, 827–838 (2010).
- Su, Z. *et al. Nat. Biotechnol.* **32**, 903–914 (2014).
- Begley, C.G. & Ellis, L.M. *Nature* **483**, 531–533 (2012).

- Anonymous. *Nat. Biotechnol.* **30**, 806 (2012).
- Collins, F.S. & Tabak, L.A. *Nature* **505**, 612–613 (2014).
- McNutt, M. *Science* **343**, 229 (2014).
- Anonymous. *Nat. Biotechnol.* **33**, 319 (2015).
- Alberts, B. *et al. Science* **348**, 1420–1422 (2015).
- Nosek, B.A. *et al. Science* **348**, 1422–1425 (2015).
- Baker, M. *Nature* **533**, 452–454 (2016).
- Beaulieu-Jones, B.K. & Greene, C.S. *Nat. Biotechnol.* **35**, 342–346 (2017).
- Di Tommaso, P. *et al. Nat. Biotechnol.* **35**, 316–319 (2017).
- Ioannidis, J.P.A. *et al. PLoS Med.* **2**, 0696–0701 (2005).
- Micheel, C.M. *et al. Evolution of Translational Omics; Lessons Learned and the Path Forward.* (National Academies Press, Washington, DC, 2012).
- Ioannidis, J.P.A. *et al. Nat. Genet.* **41**, 149–155 (2009).
- Wilkinson, M.D. *et al. Sci. Data.* **3**, 160018 (2016).

Challenges and recommendations for epigenomics in precision health

To the Editor: In March 2017, US life insurance company, GWG Life (Minneapolis) started to require policy owners to submit saliva samples. The company was not interested in the genes that their customers inherited, but in the epigenetic state of genes in the form of DNA methylation, for which it had licensed an epigenomic technology to predict an individual's health and life span^{1,2}. This raises the issue of individual epigenomic profiles being used to charge more or less for insurance coverage—or to deny life insurance altogether. The silver lining is that the epigenetic profile is not fixed in stone; you may improve your epigenome by changes in diet, exercise, or other modifications.

Seven years after epigenetics was featured on the cover of *Time* magazine, the development of novel epigenomic techniques has led to a better understanding of how the epigenome changes across individuals and health states³. For example, researchers have shown that the T cells of the immune system change at the epigenomic level during aging or functional exhaustion^{4,5}. The response of certain cancer patients to a drug treatment can be predicted from DNA accessibility of target loci during treatment⁶. Whether a person smokes cigarettes, and thus may be at risk for myriad cancers, can be inferred from DNA methylation patterns⁷. The origin of cell-free DNA, released into the blood by damaged tissues can be gleaned from nucleosome positioning in those fragments⁸. Thus, in the coming years, we can expect epigenomic findings to be used increasingly in determining diagnosis, treatment course, and even the cost of insurance. Here, we describe five recommendations for continued development in this field, formulated with an interdisciplinary group of experts toward realizing the full potential of epigenomic medicine.

The authors' perspectives for this piece came together via the Centers of Excellence in Genomic Science (CEGS), sponsored by the US National Human Genome Research

Institute (NHGRI). CEGS aim to develop and disseminate novel genomic and epigenomic technologies. As technology developers, we recognize that the field of epigenomics is rapidly maturing in both the technological and biological sciences and many further applications of epigenomic technology have been proposed for both clinical and commercial purposes. For the technology to be optimally designed for discovery or research purposes as well as robust application, priorities for use across these fields should be considered together at the outset. Thus we sought to bring together an interdisciplinary group consisting of both developers and end users of epigenomic technology to propose priorities for the field. The Center for Personal Dynamic Regulomes at Stanford, California, joined with the CEGS investigators from Harvard Medical School (Boston), Massachusetts Institute of Technology (Cambridge, MA, USA), Dana-Farber Cancer Institute (Boston), Massachusetts General Hospital (Boston), the Salk Institute (La Jolla, CA, USA), and the University of Chicago, as well as thought leaders in academic medicine, executives from companies specializing in diagnosis, lifestyle, and data analysis, and biosecurity and bioethics experts. Our recommendations reflect discussions that began with a consideration of clinical and consumer needs and moved to technical feasibility, commercial opportunities, and regulatory and ethical considerations.

The promise and challenge of epigenomics

Precision medicine promises to greatly improve individualized medical care, and this promise hinges not only on genetic tests and therapies, but also epigenetic insights. Although the massive power of DNA sequencing has largely been applied to genome and exome sequencing as a means to trace sequence variants in myriad diseases, such applications do not capture

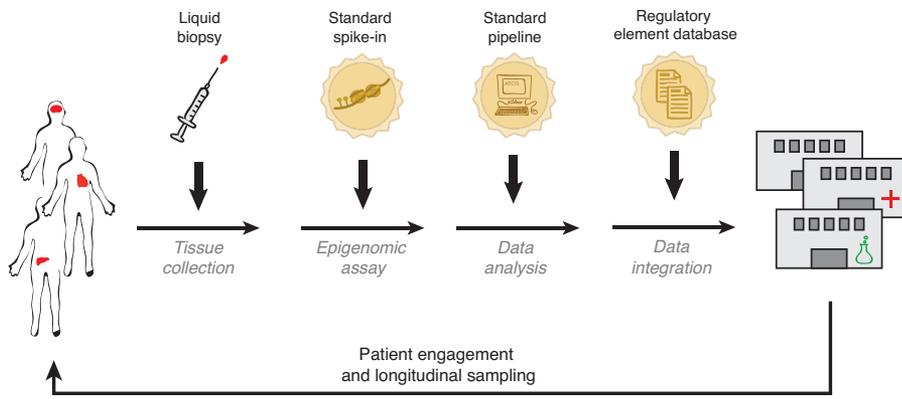


Figure 1 Summary of challenges and recommendations to advance epigenomics in precision medicine.

the tissue, context, or temporal dynamism to the epigenome that defines cell state and drives phenotypes⁹. Take the case of a woman who brings her young son to the pediatrician, explaining a set of symptoms. After some testing, the doctor confirms that the child has a fairly common autoimmune disease. The doctor is intrigued, however, when the woman explains that her son's identical twin does not show any symptoms. He suggests that although the two boys share the same genetic information, their discordance for the disease could perhaps be explained by differences in epigenomic states, leading to changes in gene expression. The woman consents to giving blood samples from her children for a research study, which finds a handful of changes in DNA methylation and chromatin accessibility in the blood of the two boys that could be related to the disease. This study is powerful in its ability to ascertain how epigenomic changes lead to disease or other phenotypes because of the controlled genetic makeup of the twins^{10,11}.

Although much evidence for epigenomic biomarkers and drivers of disease exist in cases like the above, their application to broader clinical medicine has yet to be fully realized. Before they can be applied in the context of precision medicine, the following questions need to be answered. How does one know the measurements are accurate? Which changes are transient versus true biomarkers of disease? How can results be compared across studies and individuals? What are the best ways to compare different emerging technologies for measurement, and how do we ensure that these new insights are used for maximal benefit toward precision health?

Many of the challenges associated with making personalized genomic medicine a reality hold true for epigenomic medicine as well. Beyond the technical challenge of

resolving the complex molecular basis of clinical phenotypes, these include issues of patient privacy and identification, ethical considerations of editing the genome or epigenome, difficulties in gaining access to diverse patient cohorts, and the cost of individualized care. Though critical, these considerations have been addressed repeatedly in the literature and in the realm of policy and thus are not discussed here¹². Instead, we describe the opportunities unique to epigenomic medicine and the hurdles that must be overcome for this vision to become reality.

Epigenomic science

The epigenome encompasses a broad range of molecular signals associated with DNA, from modifications to DNA itself, to nucleosomes and their modifications, to the folding of chromatin and the accessibility of regulatory regions to DNA binding proteins. The epigenome integrates genetic information, such as inherited sequence variants, together with environmental information—propagated via signaling cascades and pathways—to affect functional outcomes for the cell. Reading the epigenome, which, unlike the genome is highly dynamic, can reveal disease-, condition-, and tissue-associated changes in gene regulation over time that do not involve changes at the sequence level¹³. Whether these epigenomic modifications are the cause or simply the consequence of changes in cellular state, they are potentially valuable as diagnostic biomarkers and/or as targets for therapeutic intervention.

Several methods have been developed in the past few years to make epigenetic measurements on small, clinical samples on a reasonable time scale; for example, ATAC-seq (assay for transposase-accessible chromatin using sequencing) reveals information about protein–DNA binding and regulatory activity

of a majority of non-coding functional elements in the genome on a timescale of one day. It requires only a small number of cells as input, making it an attractive technology for clinical application^{14–16}. Other techniques that involve sequencing of epigenomic modifications, such as cytosine and histone methylation, provide opportunities for both invasive (i.e., biopsy) and non-invasive (i.e., cell-free DNA/liquid biopsy) diagnosis and prognosis¹⁷. The rapid development of these techniques points toward further integration of epigenomic metrics, alongside genomic metrics, in clinical medicine and lifestyle applications.

The nature of the epigenome also raises unique challenges related to two key properties¹⁸: dynamism and tissue specificity, and continuous scale.

The dynamism and tissue-specificity of epigenomic modification require that for appropriate interpretation, assays be performed longitudinally from biopsies or blood samples from a single person over the course of disease progression, aging, or lifestyle intervention. Furthermore, epigenomic measurements are most often measured on a continuous scale without a 'ground truth' value or standard threshold for presence or absence of a feature. These facts bring challenges in assay and data analysis standardization and/or storage and in long-term patient engagement and feedback. For those in academia and industry seeking to tackle these unique challenges, we describe below five recommended areas of focus: first, develop a standard chromatinized DNA sample to benchmark assays; second, create a standardized analysis pipeline for quality control of epigenomic data, such as ATAC-seq data; third, name and catalog regulatory elements in the genome in a manner that can evolve as knowledge increases around regulatory element structure and function; fourth, develop experimental and computational biomarker assays to acquire epigenomic information from all tissues using non-invasive collection methods; and fifth, cultivate a large and diverse patient pool for participation in longitudinal studies over long time frames. The first three of these recommendations relate to standardization; the final two to clinical application (Fig. 1).

Recommendation 1: develop a commercially available and affordable standard spike-in chromatin sample

We recommend the development of a commercially available and affordable standard chromatinized DNA sample that can be added from the start of an assay to

control for differences among laboratories and over time. Imagine that a researcher in another country wants to test whether some children that have developed certain symptoms in his clinic are in the early stages of the same autoimmune disease that presented in the single twin example described above. These patients likely would not have healthy identical twins for comparison, and thus this researcher would need to compare against other healthy individuals whose data are available, like the original healthy twins. To compare the DNA methylation and chromatin accessibility in these new patients, this researcher must be able to benchmark his assay against those done at the previous institution. This comparison across large cohorts of longitudinal samples to discover mechanisms of condition-related gene regulation requires standardization of sample preparation and analysis. This type of standardization would also expedite the process for regulators to sign off on such discoveries to be turned into common clinical practices. Though written protocols for sample preparation are currently available, we recommend that a standard 'lynchpin' sample be used in all epigenomic experiments as a benchmark to ensure assay quality, instrument performance, and normalization.

The use of standard 'ground truth' samples is common in RNA-seq experiments, which employ ERCC (External RNA Control Consortium) spike-in RNA samples for comparison¹⁹. These spike-ins, combined with unique molecular indexing (UMI) allow an absolute measurement of molecules of a given type in a sample. Similar standard samples are being developed for genomic DNA sequencing by the Genome In a Bottle Consortium²⁰. Developing this type of standard is especially important for epigenomic markers that are measured as continuous variables and in reference to an internal control. In the case of the twins, imagine that the brother with autoimmune symptoms showed a fourfold increase in chromatin accessibility at an important immune-modulating gene over his brother. In another laboratory across the country, technical variables may yield an increase in a patient of threefold over an unaffected individual. To determine whether this difference is significant, a reference sample is required. Overall enrichment of signal in assays like chromatin immunoprecipitation sequencing (ChIP-seq) and ATAC-seq, and conversion rates for bisulfite sequencing vary between experiments, even within the

same laboratory. Furthermore, these can change based on the library preparation and the sequencing instruments used, which can affect biases in DNA fragment lengths and sequence content. An internal standard is absolutely imperative for defining 'normal' versus 'pathogenic' in these cases.

A standard chromatin sample could come from renewable mouse and human sources, like common ENCODE (Encyclopedia of DNA Elements) cell lines, and should have high confidence data available on DNA accessibility, histone modifications, nucleosome positions, methylation, and sequence polymorphisms. It is important that this type of standard be prepared from a single batch, or as few batches as possible, of cells in an industry-standard setting to eliminate the current problems arising because of various confounding factors. These include different passage numbers, karyotypes, culture conditions, and lab-to-lab variation. In addition, so that this program can continue unbroken into the future, we will need a strategy whereby new, or renewed, standard cell lines are periodically brought into the program and characterized well before established cell lines are exhausted or too far altered (genetically or epigenetically). We recognize the value of normal karyotypes, isogenic samples, and ancestral diversity. Finally, coordination of the choice of cell lines with ongoing efforts to describe the long-range structure of genomes through chromatin capture techniques (e.g., Hi-C) and imaging technologies will greatly facilitate the integration of three- and four-dimensional (4D; including time) genome information.

Recommendation 2: create a consensus analysis pipeline and data repository for epigenomic assays

There is a need for a standardized analysis pipeline for quality control of epigenomic data. At present, to analyze data published by multiple laboratories, researchers must start from raw sequencing reads owing to differences in downstream pipelines, making comparison inefficient. In addition, unlike base calls in genomic data, epigenomic measurements are continuous, and thus thresholds are arbitrarily chosen. To confidently assign biological meaning to epigenomic changes, researchers must be capable of integrating very large sample cohorts from multiple institutions analyzed in the same way. This standardized analysis would apply to the most upstream steps of sequence analysis, specifically pertaining to quality control. As new techniques

become available and multiple institutions apply them to multiple tissue types, quality metrics related to signal-to-noise ratios and bias in the data must be recorded in a standardized way. This type of pipeline would still allow novel downstream analyses in individual laboratories, while improving reproducibility and comparison across studies.

Although we recognize that the development of novel analyses is just as important to driving research forward as the development of new assays, analyses for clinical samples need to be standardized to facilitate comparisons and meet regulatory standards. Public pipelines generated by consortiums like ENCODE are a good start, but there is currently no requirement or incentive for individual laboratories to make use of them. Furthermore, it is hard to overemphasize the importance of data visualization methods. Methods for simplification and visualization of complex epigenomic data should be developed and standardized for ease of dissemination and human interpretability. Finally, a common data repository with standardized metrics would allow laboratories to deposit data analyzed in the same way and compare new data against high-quality data from others in the field. Simplifying the process of comparing data sets via a common data deposit space and user-friendly interface would speed this process.

Recommendation 3: catalog and register regulatory elements in a standardized index for easy searchability and comparison across laboratories, genome builds, and species

Concurrent to standardized analytic methods, there is an urgent need for a way to name and catalog regulatory elements in the genome as knowledge increases concerning their structure and function. Unlike genes, regulatory elements are not currently indexed in a standardized manner, which makes cross-referencing among studies both manual and laborious. A gene (or more precisely a transcript) is identified not only by its sequence, strand, structure, and coordinates in the reference genome, but also by a standardized unique identifier (its gene name) that links it across genome builds, species, and assays as well as to its product. At the gene level, confusion often arises because there are multiple databases with unique transcript names and identifiers^{21–23}. This problem is dramatically expanded in the regulatory space. Regulatory elements, which are estimated to comprise ~10–20% of the

human genome (aggregated across all cellular contexts), have no universally accepted identifiers at present.

Creating a dictionary of regulatory elements is an essential foundation for the construction of a body of evidence linking a given regulatory region to a phenotype or disease. Take for example, the set of loci whose accessibility or methylation are changed between the healthy and diseased twin. Another study at another institution finds that in a set of patients, those affected with a related autoimmune disease, show changes in the epigenomic state at 20% of these same loci. Without a common naming system, these two studies may never link these regions together, and the latter group of researchers likely would not be able to find that these loci had previously been linked to autoimmunity.

The task of cataloguing regulatory elements is complicated by several factors. First, regulatory elements do not have well-defined starts and stops (unlike, for example, an open reading frame of a gene). Thus, setting a convention for choosing the starts and ends of enhancers, promoters, and other regulatory elements is necessary. Second, regulatory elements may expand or contract in size across different cellular states or form higher order coordinated units. Thus, a higher order structure may need to be imposed on a dictionary of elements that is flexible enough to capture such complex relationships. Third, regulatory elements are not as well conserved as coding genes. Thus, comparisons across species may require the use of operational and spatial identification instead of sequence-based identification. Fourth, not all accessible elements in the genome are functional, and not all nucleotides within a defined element are functional, necessitating high-throughput methods for assaying necessity and sufficiency of elements in different contexts²⁴.

Several collaborative functional genomics projects, such as the ENCODE Project, The Roadmap Epigenomics Project, the International Human Epigenome Consortium (IHEC), and Functional Annotation of the Mammalian Genome (FANTOM) Consortium have made substantial progress toward identifying and cataloguing candidate regulatory elements by integrating multiple types of functional genomic data (e.g., chromatin accessibility, RNA-seq, ChIP-seq, cap analysis gene expression (CAGE), and DNA methylation) to predict chromatin state for a given region across multiple common cell

types^{25–27}. Using the ENCODE portal, for example, regulatory regions can be queried by coordinate or by nearby gene name, yielding an output of expression level or accessibility in a host of tissues. Despite these major advances, at present this approach has limitations. We support a convention for naming and numbering be extended more broadly in the field such that publications from all groups report regulatory elements in a standard way for all samples, not only those included in a specific consortium. Furthermore, an ontology, similar to that employed by ENCODE, would simplify making functional connections between samples, regions, cell types, and species²⁸.

A combination of a standardized chromatin spike-in and analysis framework and a cataloguing system for the non-coding genome will expedite understanding of the relationship between epigenetic changes and biological meaning and thus its translation into applications in clinical practice.

Recommendation 4: develop assays for inferring tissue-of-interest epigenomic modifications from accessible tissues such as blood, skin, saliva, and urine

Unlike genome sequencing, which requires often just one sample from an accessible tissue at a single time point, epigenomic assays require multiple samples over time from tissues of interest. This necessitates two advances: first, the development of biomarker assays that can infer the state of the epigenome in an inaccessible tissue, such as the brain, from blood, skin, saliva, or urine; and second, long-term engagement in epigenomic investigations from enrolled patients.

For longitudinal studies on healthy participants to be feasible, as well as for epigenomic medicine to reach its full potential, non-invasive collection methods must be sufficient to acquire epigenomic information from all tissues. The ultimate goal is to be able to infer the cell-type specific state of an epigenomic feature in the forebrain or in a pancreatic tumor through a surrogate tissue source that can be collected rapidly and repeatedly. This is especially important for measurements in healthy people, who ideally would not have to visit a hospital for a biopsy to participate. These kinds of surrogate tissue assays have proven useful at the DNA sequence level in both pre-natal diagnosis and cancer. In these cases, DNA found in the blood, called cell-free DNA (cfDNA), can be sequenced and assigned to its tumor or fetal origin by a mutational signature or sex chromosomes and genetic polymorphisms,

respectively^{29,30}. Studies in cancer have shown that the amount of circulating DNA from a known tumor with a common mutational signature correlates with tumor size and thus the number of cells dying and releasing DNA into the bloodstream.

For conditions that do not involve changes to the DNA sequence itself, epigenomic biomarker assays in surrogate tissues would be highly valuable. These assays would measure epigenomic features correlated with health outcomes, or identify the tissue source of DNA being released as a result of tissue damage. As the twin with autoimmune disease progresses through his life, complications may arise as the immune system attacks his own tissues. Imagine that doctors could identify exactly which tissues of the gut, for example, are being damaged and releasing their DNA into the blood based on epigenomic features in cfDNA. This would allow the doctor to specifically treat symptoms in that organ.

As described above, DNA methylation and nucleosome positioning measured in the blood have been shown to correlate with the tissue of origin and a number of health traits, including life expectancy and smoking^{2,7,8}. Liquid biopsy-based epigenomic assays have already shown their utility for clinical application. Researchers in academia and industry should strive to continue developing a broader range of biochemical and analytical methods for inferring or deconvoluting the epigenetic state of various cell types from blood, skin, saliva, and urine.

Recommendation 5: improve long-term participant involvement via enhanced result reporting and patient-brokered data sharing

One of the greatest challenges in biomedical research is cultivation of a large and diverse patient pool for participation in longitudinal studies over long time periods. This is especially important to capture temporal dynamics in epigenomic measurements. Such a large-scale longitudinal study requires an elevated level of patient participation. One reason for challenges in this area has been the fact that patients are not always notified of and encouraged to engage in the findings of research studies. One way to improve participant engagement is via increased feedback to participants regarding study results. We also encourage experiments with a data clearinghouse. Such an online clearinghouse could allow individuals to submit their own data (e.g., ATAC-seq data from blood) to multiple ongoing studies

in order to gain more information about their health from a single donation. This may also facilitate data sharing among researchers, but with patients controlling the use of their data by acting as their own brokers.

Personalized epigenomic measurements hold enormous and transformative potential for clinical medicine and even for lifestyle applications. The epigenome has the power to reveal why the two identical twins present differently for diseases, which medication they may respond to, and the influence of environmental factors on the progression of the disease throughout their lifetime. But the complexity and dynamic nature of the epigenome over time and health state bring unique difficulties surrounding standardization, data analysis and sharing, sample acquisition, and patient participation.

Editor's note: This article has been peer-reviewed.

ACKNOWLEDGMENTS

We thank J.A. Stamatoyannopoulos (U. Washington), C.-T. Wu (Harvard Medical School) and J. Schloss (NHGRI) for input. We thank the National Human Genome Research Institute for support.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Ava C Carter¹, Howard Y Chang¹, George Church², Ashley Dombkowski³, Joseph R Ecker⁴, Elad Gil⁵, Paul G Giresi⁶, Henry Greely⁷, William J Greenleaf^{1,8}, Nir Hacohen⁹, Chuan He¹⁰, David Hill¹¹, Justin Ko¹², Isaac Kohane¹³, Anshul Kundaje¹⁴,

Megan Palmer¹⁵, Michael P Snyder^{1,8}, Joyce Tung¹⁶, Alexander Urban^{1,17}, Marc Vidal¹¹ & Wing Wong^{1,18}

¹Center for Personal Dynamic Regulomes, Stanford University, Stanford, California, USA.

²Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA, and Wyss Institute, Boston, Massachusetts, USA.

³BEFORE Brands, Menlo Park, California, USA. ⁴The Salk Institute for Biological Studies, La Jolla, California, USA, and Howard Hughes Medical Institute. ⁵Color Genomics, Burlingame, California, USA. ⁶Epinomics, Menlo Park, California, USA. ⁷Center for Law and the Biosciences, Stanford University, Stanford, California, USA. ⁸Department of Genetics, Stanford University, Stanford, California, USA. ⁹Massachusetts General Hospital, Boston, Massachusetts, USA. ¹⁰University of Chicago, Chicago, Illinois, USA, & Howard Hughes Medical Institute ¹¹Center for Cancer Systems Biology, Dana Farber Cancer Institute, Boston, Massachusetts, USA. ¹²Department of Dermatology, Stanford University, Stanford, California, USA. ¹³Department of Medical Informatics, Harvard Medical School, Boston, Massachusetts, USA. ¹⁴Departments of Genetics and Computer Science, Stanford University, Stanford, California, USA. ¹⁵Center for International Security and Cooperation, Stanford University, Stanford, California, USA. ¹⁶23andMe, Mountain View, California, USA. ¹⁷Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA. ¹⁸Department of Statistics, Stanford University, Stanford, California, USA. e-mail: howchang@stanford.edu.

¹Anonymous. GWG Life becomes first insurtech firm to collect epigenetic samples to analyze biomarkers of life insurance policy owners. <https://globenews-wire.com/news-release/2017/03/02/930557/0/en/>

2. Chen, B.H. *et al.* *Aging* **8**, 1844–1865 (2016).

3. Cloud, J. Why your DNA isn't your destiny. *Time* <http://content.time.com/time/subscriber/article/0,33009,1952313-2,00.html> (2010).

4. Moskowitz, X. *et al.* *Sci. Immunol.* **2**, 1–2 (2016).

5. Sen, D.R. *et al.* *Science* **354**, 1165–1169 (2016).

6. Qu, K. *et al.* *Cancer Cell* **32**, 27–41.e4 (2017).

7. Joehanes, R. *et al.* *Circ. Cardiovasc. Genet.* **9**, 436–447 (2016).

8. Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M. & Shendure, J. *Cell* **164**, 57–68 (2016).

9. Soon, W.W., Hariharan, M. & Snyder, M.P. *Mol. Syst. Biol.* **9**, 640 (2013).

10. Fraga, M.F. *et al.* *Proc. Natl. Acad. Sci. USA* **102**, 10604–10609 (2005).

11. Castillo-Fernandez, J.E., Spector, T.D. & Bell, J.T. *Genome Med.* **6**, 60 (2014).

12. McGuire, A.L., Caulfield, T. & Cho, M.K. *Nat. Rev. Genet.* **9**, 152–156 (2008).

13. Allis, C.D. & Jenuwein, T. *Nat. Rev. Genet.* **17**, 487–500 (2016).

14. Buenostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. *Nat. Methods* **10**, 1213–1218 (2013).

15. Adey, A. & Shendure, J. *Genome Res.* **22**, 1139–1143 (2012).

16. Schmidl, C., Rendeiro, A.F., Sheffield, N.C. & Bock, C. *Nat. Methods* **12**, 963–965 (2015).

17. Fraser, M. *et al.* *Nature* **541**, 359–364 (2017).

18. Soshnev, A.A., Josefowicz, S.Z. & Allis, C.D. *Mol. Cell* **62**, 681–694 (2016).

19. Jiang, L. *et al.* *Genome Res.* **21**, 1543–1551 (2011).

20. Zook, J.M. *et al.* *Sci. Data* **3**, 160025 (2016).

21. Gray, K.A., Seal, R.L., Tweedie, S., Wright, M.W. & Bruford, E.A. *Hum. Genomics* **10**, 6 (2016).

22. Mockus, S.M., Patterson, S.E., Stutz, C., Bult, C.J. & Tsongalis, G.J. *Clin. Chem.* **62**, 442–448 (2016).

23. Wright, M.W. & Bruford, E.A. *Hum. Genomics* **5**, 90–98 (2011).

24. Kellis, M. *et al.* *Proc. Natl. Acad. Sci. USA* **111**, 6131–6138 (2014).

25. The ENCODE Project Consortium. *Nature* **489**, 57–74 (2012).

26. Roadmap Epigenomics Consortium. *Nature* **518**, 317–330 (2015).

27. The FANTOM Consortium. *Nature* **507**, 455–461 (2014).

28. Malladi, V.S. *et al.* *Database* **2015**, bav010 (2015).

29. Newman, A.M. *et al.* *Nat. Med.* **20**, 548–554 (2014).

30. Fan, H.C. *et al.* *Nature* **487**, 320–324 (2012).

31. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

32. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

33. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

34. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

35. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

36. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

37. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

38. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

39. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

40. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

41. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

42. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

43. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

44. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

45. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

46. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

47. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

48. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

49. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

50. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

51. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

52. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

53. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

54. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

55. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

56. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

57. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

58. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

59. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).

60. Anonymous. GWG-Life-Becomes-First-Insurtech-Firm-to-Collect-Epigenetic-Samples-to-Analyze-Biomarkers-of-Life-Insurance-Policy-Owners.html (2017).