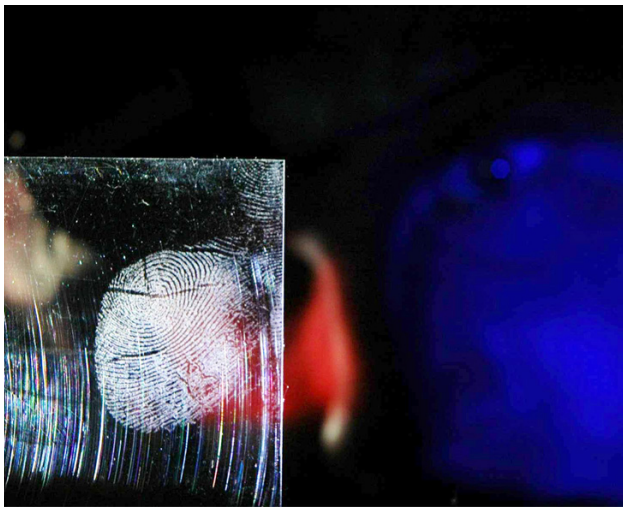
Securing the Next Biometric

Vulnerabilities of DNA-based Identity Verification Systems

Ranjani | Parthasarathy



Brief No. 10.7

P | I | P | S

The Project on International Peace and Security
Institute for the Theory and Practice of International Relations
The College of William and Mary

Please direct inquiries to:
The Project on International Peace and Security (PIPS)
Institute for the Theory and Practice of International Relations
The College of William and Mary
427 Scotland Street
Williamsburg, Virginia 23185

tele. 757.221.1441
fax. 757.221.4650
pips@wm.edu

Electronic copies of this report are available at: www.wm.edu/pips

Securing the Next Biometric
Vulnerabilities of DNA-based Identity Verification Systems

APRIL 2018

Ranjani Parthasarathy

P|I|P|S *The Project on International Peace and Security*
Institute for the Theory and Practice of International Relations
The College of William and Mary

Securing the Next Biometric
Vulnerabilities of DNA-based Identity Verification Systems

DNA is a biometric identifier that may be used for instantaneous identification within the next decade. This technology has three major vulnerabilities: sale of private data from commercial genetic analysis, spoofing based on DNA synthesis capabilities, and introduction of DNA-based malware. Current legislation inadequately safeguards US citizens' genetic information. Inexpensive DNA synthesis will enable spoofing by non-state actors. Further, open-access publications will accelerate the proliferation of malware design and production capabilities. To maintain the viability of DNA-based biometric security, policymakers will need to better protect US persons' genetic information and review publication of sensitive biotechnology research.

Introduction

DNA will be the next biometric identifier,[1] replacing fingerprints and retina scans, with applications ranging from immigration and border security to facility security and rapid personnel identification.[2] In the next decade, improving DNA sequencing technologies will enable timely DNA-based identity verification at sites critical to national security. Perceived to be "impossible to spoof,"[3] on-site DNA sequencing may verify identities in near real time—in minutes or even seconds.

While current US government research focuses on the feasibility and implementation of a *DNA-based identity verification system* (DIVS), three significant vulnerabilities threaten the future role of DNA as a real-time identifier.[4] Sequencing of US persons' DNA by private corporations and sale of this genetic data to unknown end users has disseminated personally identifying information. Exploitation of these sequences may allow spoofing attacks—that is, false presentation of one individual's genetic sequence by another individual in order to trick a DIVS. Lastly, any network connected to a DIVS may also become susceptible to attack by malware encoded by synthetic DNA.[5]

Current US policy inadequately protects citizens' genetic security. US policy allows unrestricted sale of private genetic data, endangering future DIVS security, while the scientific community openly publishes information allowing the design and synthesis of DNA-encoded malware. In order to protect the opportunity to adopt DNA-based authentication, the United States must better secure US persons' genetic information and review publication of DNA-based network exploits, while monitoring the accessibility of DNA synthesis technologies.

DNA as the Next Biometric Authenticator

Global interest in biometric security is rising.[6] Advancing beyond badges and PINs, biometric identifiers such as fingerprints, iris, or retina scans have become nearly ubiquitous even for everyday applications such as unlocking a smartphone or laptop.[7] At the same time, biometric identifiers are growing more popular for border security and industrial applications.[8] Fingerprints are linked to visas in the European Union and biometrics secure sensitive laboratory and nuclear facilities.[9]

Current real-time biometrics, such as fingerprints and iris scans, are not perfect identifiers. Some biometrics are shared among twins, family members, or even unrelated individuals.[10] Reports of twins being able to represent each other based on fingerprints and iris scans, as well as unrelated coworkers being able to open a single phone based on facial recognition technology, have eroded confidence in existing biometric security features. Growing concerns about the security of existing biometrics will likely accelerate a move towards DNA-based identity verification systems once feasible in near real time.

From Fingerprints to DNA

DNA may replace traditional biometric identifiers, allowing differentiation even between "identical" twins based on minute differences in genetic sequence.[11] DNA will likely become a feasible authenticator within the next decade, as scientists achieve accurate, near-instantaneous sequencing.[12] Decreasing cost and increasing speed of DNA sequencing, in tandem, enable a probable future where DNA will be used as an effective biometric authenticator.

- *Cheaper sequencing will increase the appeal of DNA as a biometric identifier.* In 1990, the Human Genome Project was started with a projected budget of roughly $3 billion. Today, sequencing any specific individual's genome costs less than $1,000, and is projected to fall below $100 in the next few years.[13]

- *Faster sequencing will allow DNA to be a feasible near-real-time biometric identifier.* Not only was the Human Genome Project costly, it was also lengthy: the project required DNA samples from more than eleven individuals, involved more than 250 global researchers and countless lab technicians, and produced a draft of the first consensus genome sequence after over a decade of dedicated effort.[14] Sequencing a complete human genome in January 2017 required only one hour on a high-end sequencer, and that time will continue to fall.[15]

Improvement in sequencing technologies is largely independent of interest in security applications, and is pursued by scientists who seek to maximize the use of genetic information for medical applications. These improvements will, however, also render DNA-based biometric identification feasible in near real time. Notably, one field-ready tool has already been developed to take advantage of DNA as a perfect biometric identifier. DNA sequencing kits used by US Special Forces are able to reach an ID within 90 minutes, demonstrating the potential for DNA to serve as a first-line authenticator in the near future.[16].

The Road to DNA-based Identification

A state-level reference DNA database is a prerequisite for implementing DIVS; however, DNA collection has led to privacy challenges. Additionally, state-level collection on legal grounds may enable targeting of ethnic or political dissidents. Finally, state-level DNA databases not only offer an indicator of state potential to implement DIVS, but also reveal the state's technological and scientific capacity to exploit the vulnerabilities of US DNA-based biometric security.

- *DNA collection for non-instantaneous identification.* DNA databases allow identification of individuals, with a time delay of at least 90 minutes. Some states, such as Saudi Arabia, have linked DNA samples to national biometric ID databases, allowing the government to better track and verify the identities of immigrants or residents.[17] Additionally, many states have DNA databases that cover only a portion of the population for the purpose of post-mortem identification (as with the US military[18]) or state-wide databases for the same purpose (e.g. Kuwait).[19]

  These databases may face legal challenges on privacy grounds. Kuwait's national DNA database was established following an attack on a mosque by ISIS-affiliated extremists in 2015.[20] In 2017, challenges on privacy grounds were upheld by Kuwait's Supreme Court and DNA sample collection will not continue, although whether existing samples will be preserved and analyzed in the wake of this ruling is not yet known.[21]

- *DNA information from legal proceedings.* Forensic databases with genetic collection on suspected or arrested criminals exist in 64 countries, including the United States,[22] the UK,[23] Russia,[24] and China.[25] The criteria for DNA collection vary, with some countries requiring that the accused citizen be found guilty and serve some part of their sentence while others allow DNA collection of suspects in any investigation. [26]

  In particular, the compilation of DNA samples from political and ethnic dissidents may warrant heightened scrutiny. In China, DNA collection restricted to minority ethnic groups has raised concerns.[27] Potential implications may include ethnic targeting, as wide-ranging DNA collection combined with computational genome processing and medical research into ethnic DNA markers allow fine-tuned identification of ethnic make-up.[28] Speculatively, bioweapons may be tuned to an individual's DNA or to more general ethnic DNA markers.[29]

- *DNA sequencing for medical applications.* Precision Medicine Initiatives (PMIs) are efforts to collect high-quality, comprehensive genetic information on more than one million individuals. Such initiatives require extensive investment into genetic sequencing technologies, many of which are designed by US-based companies. Recently, US-China competition has expanded into biotechnology, as both countries are pursing PMIs that have accelerated the development of their genetic analysis infrastructures.[30]

  Thus far, the preponderance of genetic analysis has directly or indirectly relied on US companies for sequencing technologies or data analysis.[31] Investment into entirely independent sequencing technologies by other countries challenges the United States'

technological advantage in this field, posing a dual-use threat.[32] Capacity-building will not only allow countries to implement DNA-based identity verification systems, but will also position them to exploit its vulnerabilities.

State-level DNA collection serves as an indicator of an emerging capacity to exploit the vulnerabilities of DNA-based identity verification systems (DIVS). While sequencing and other genetic analysis technologies have legitimate medical and legal applications, their dual-use nature may warrant additional scrutiny when in the hands of potential US adversaries.[33]

Vulnerabilities of DNA-based Identity Verification Systems

> *"A Chinese company, Beijing Genomics Institute (BGI), is the world's largest genetic research center… and [it] has sequenced the genomes of millions of Americans."*
>
> – Paul Scharre, in testimony to the House Armed Services Committee, 2018[34]

Despite confidence in DNA as an ideal authenticator, the capacity to develop workarounds or otherwise exploit future DNA-based identity verification systems (DIVS) poses a threat to US security. The United States faces three key vulnerabilities that will jeopardize the security of DIVS: distributed genetic information may be exploited by potential adversaries, synthetic DNA may allow DIVS spoofing, and DNA-based malware may threaten DIVS-associated networks.

Commercialization of Genetic Information

Private companies offering genetic analysis services batch, de-identify, and ultimately sell genetic data to third parties. This largely unregulated distribution of uniquely-identifying genetic information threatens future opportunities for the United States to pursue DNA-based biometric security.

Commercial kits to collect and analyze DNA samples for ancestry information and medical outcomes have been widely popularized in the United States. The breadth of the consumer base, the diverse interests of users, and the relative accessibility of these kits have allowed private companies to capture genetic data on a significant portion of the US population.[35]

While the consumer retains a copy of their genetic information, the genetic data collected through these commercial offerings is even more valuable to the company.[36] Businesses compile hundreds or thousands of genomes into one dataset, strip the consumers' names from the data, and sell these datasets to unknown end users who may include health insurance companies or even foreign actors.[37]

Although nominally de-identified, the genetic sequences in these datasets are still personal identifiers. The consumer's name is not attached to the DNA sequence, but the sequence is still unique to a single individual. In research scenarios, de-identified genomes have successfully been

traced back to the individuals to whom they belong.[38] Forensic analysis also demonstrates the ease of re-identification given an anonymized DNA sample.[39] Varying levels of sequencing coverage may allow for DNA-based identification:

- *Forensics methods allow DNA identification but cannot guarantee perfectly confident identification*. Forensics databases use short tandem repeats (STR) to identify individuals. These repeated sequences are copied with low fidelity by DNA replication machinery *in vivo*, resulting in near-unique patterns of STRs across several different sites of any individual's genome.[40]

- *Sequencing the full genome guarantees individualized identification.* The STR pattern of an individual, though typically unique, may be shared among close relatives or twins. Among identical twins, only whole-genome sequencing (WGS), which allows comparison of the entire genome between the two individuals, allows unique identity verification.[41] STR sequencing alongside computational analysis may allow stronger identification, however, potentially accelerating the implementation of DIVS.[42]

  The only rapid DNA-based biometric ID system in use relies on the STR model and has a sequencing time of 90 minutes. It is likely that the decision to use STR identification was driven both by having an existing forensic database of targets and by the relatively accelerated sequencing timeline. In the future, DIVS relying on sequencing the entire genome may be preferred for a heightened security guarantee.

Commercial sale of genetic data disproportionately threatens US implementation of DIVS, as the majority of individuals seeking commercial genetic analysis services are US citizens.[43] Additionally, Chinese sequencing of US citizens' genomes has widened the set of individuals with access to US genetic information beyond US businesses to foreign consumers.[44] US persons' genetic information is being disseminated before DIVS becomes feasible and may then be used to spoof DNA-based biometric security.

Synthesis and Spoofing

Current trends in DNA synthesis technology will enable spoofing of a DNA-based identity verification system (DIVS). Spoofing is defined as the presentation of a false identity to any authentication system. In the case of a DIVS, spoofing would take the form of one individual using another's DNA sequence. DNA synthesis technology may allow actors with knowledge of the DNA sequence required for authentication to spoof the system using synthetic DNA.

Made-to-order, biologically active synthetic DNA constructs are presently limited in practice to approximately 6,000 base pairs: a fraction of the human genome, which measures 3.2 billion base pairs.[45] DNA synthesis, however, is sharply decreasing in cost and increasing in accessibility.[46] Two main factors will determine whether the capability for DIVS spoofing will be restricted to state actors or extend to non-state actors:

- *Length of authenticating DNA sequence.* STR-based identity verification will rely on the presence of a few dozen short sequences of DNA that are more easily synthesized and obtained commercially. STR sequences may be spoofed using existing technologies, as spoofing genome STRs would require only a mix of shorter synthetic stretches that may be within the price range of a non-state actor. STRs are among the most difficult regions of the human genome to synthesize accurately, however, which may increase the resilience of STR-based DIVS to this vulnerability.

- *Limitations on DNA synthesis.* While synthesis length is restricted by technological limits, WGS-based DIVS will be less easily spoofed by synthetic DNA compared to an STR-based DIVS. The required scientific knowledge and instrumentation to assemble a genome ordered in multiple pieces will be higher.

  Current synthesis technologies barely enable whole-genome synthesis of small genomes: less than 10,000 base pairs.[47] Assembling a complete human genome will be technologically demanding even for state actors, and most likely impossible for a non-state actor.[48]

Fundamentally, the capacity to create synthetic DNA from a digital sequence creates a vulnerability for DIVS, although exploiting this vulnerability will require significant scientific expertise and technological advancement. Technological limits on synthesizing full human genomes will persist for some time after DIVS is enabled by sequencing advances, allowing policymakers to better assess the threat from non-state actors.

DNA Malware

DNA malware, a network attack encoded by synthetic DNA, has been explored to a limited degree by a single lab at the University of Washington. A proof-of-concept paper published in August 2017 establishes that synthetic DNA sequences, in conjunction with poor network hygiene, can allow a remote actor access to a secure network.[49]

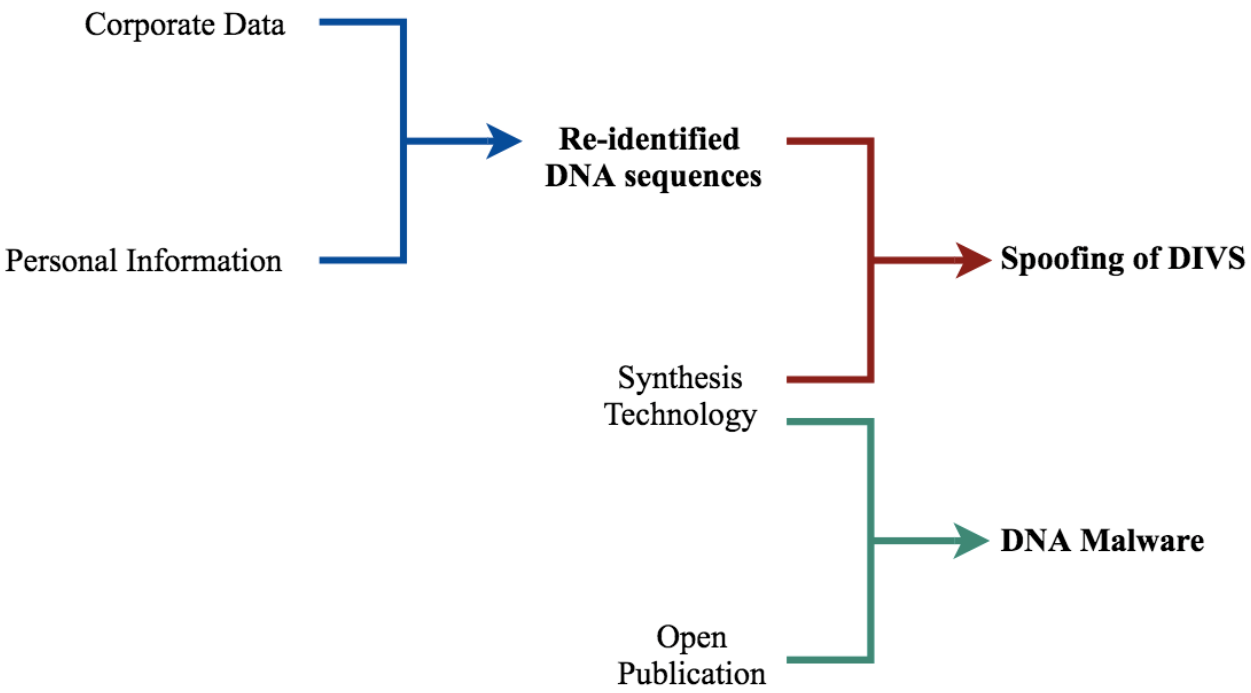
DIVS on portable systems used in the field may be particularly vulnerable to such exploits if separated from the individual to whom the system was issued, as DNA malware would allow access to the system's secure network while bypassing user authentication. This vulnerability may be exploited by states with resources in both cyberspace and biotechnology. Non-state actors may also be able to take advantage of this vulnerability if current publication and policy practices are maintained.

DIVS Vulnerabilities in Summary

The United States has three significant vulnerabilities that undermine the security of future DIVS. First, commercial sale of US persons' genetic information leads to the dissemination of US persons' DNA sequences that may, in conjunction with other public data such as genealogies, be

re-identified by potential adversaries. This re-identified sequence, paired with DNA synthesis technologies, can be used to spoof a DIVS by presenting synthetic DNA. Finally, open publication of malware design alongside commercial access to DNA synthesis options will allow DIVS to be a point of entry for DNA malware to attack a secure network. The diagram below summarizes these vulnerabilities:

Figure 1. Summary of DIVS vulnerabilities



Current US Policy and Research Practices

Existing US federal policy regarding genetic data focuses on the medical implications of sequencing, but does not consider future security applications—though some US state laws offer stronger protections from a personal privacy standpoint. Current US policy and research practices may accelerate the failure of DIVS for the following reasons:

- *US policies overlook security implications of genetic collection for DIVS implementation.* Policies at the federal and state levels address privacy concerns surrounding the collection and retention of genetic information. The Genetic Information Nondiscrimination Act, passed in 2008, prohibits discrimination in insurance decisions on the basis of predispositions to disease discovered through genetic analysis for medical applications.

Similarly, the most recent National Defense Authorization Act considers genomics only in a medical context.[50]

At the state level, Texas, Illinois, and Washington have passed laws prohibiting the retention of biometric information on private individuals, including information derived from commercial DNA analysis. These state laws—although driven by privacy concerns rather than security threats—may better protect genetic information and serve as a model for future federal policy.

- *Previous US inaction following major national security leaks magnifies existing DIVS vulnerabilities.* Compiling de-identified genetic data with previously publicized information about US national security employees could rapidly accelerate the failure of DIVS by allowing more effective re-identification of DNA sequences. Information regarding US national security personnel was extensively compromised in a 2015 hack of the federal Office of Personnel Management.[51] More broadly, US citizens' social security and credit information was leaked in 2017 when Equifax was breached.[52] Both hacks have been tentatively attributed to Chinese or Chinese-sponsored actors, demonstrating that potential adversaries may collect sensitive information on US persons.[53]

  Knowledge of an individual's health history will also accelerate the process of matching DNA sequences to their originators, based on DNA markers that indicate specific medical outcomes. Wide-scale breaches of health information on US persons have occurred with great regularity as the use of digitized health records has become the norm.[54] One projection estimates that every patient in the United States will have had their health records leaked by the year 2024.[55] The relative insecurity of health records thus intensifies genetic insecurity in the United States.

- *Current US research into DNA-based malware is published with no review for national security applications, publicizing a DIVS vulnerability.* While research into DNA malware is presently undertaken by US institutions and close allies, open publication accelerates the pace with which other actors may exploit US vulnerabilities.[56] Safeguards cover research into other dual-use biotechnology, but DNA malware is not yet designated as a dual-use concern.

Commercial collection of genetic information, inaction following information leaks, and open publication of sensitive research allow the creation of vulnerabilities even prior to the successful implementation of DNA-based identity verification systems (DIVS). Changes to US policy and research practices are required to preserve the opportunity for future US use of DIVS.

Protecting Future Implementation of DIVS

To protect future implementation of DIVS, US policy and research practices must address key vulnerabilities now. Unrestricted sale of US persons' genetic information, commercial access to

synthesis technologies, and DNA malware publications must all be monitored or curtailed if DIVS will be used in sensitive facility or network security.

In order to pursue a DNA-based identity verification system in the future, US policy must more stringently monitor current data sales. Restrictive US state laws that prohibit the retention or sale of privately collected genetic information—such as those in Texas, Illinois, and Washington—may serve as a model for federal policy, keeping US citizens' genetic data more secure. Specifically, policy designed to reduce US citizens' vulnerability to foreign information gathering via the sale of genetic data may strengthen the future potential of DIVS.

Expansion of the existing protocols to identify pathogenic sequences in synthetic DNA requests may allow detection of commercial synthetic DNA intended to spoof a DIVS. While such expansion will not rule out state capacity to synthesize either synthetic genomic DNA or DNA malware, it may reduce the ability of non-state actors to acquire these technologies.[57]

Additional review of sensitive biotechnological topics, such as the development of DNA malware, will better secure networks against DIVS as a weak point of entry. While a proof-of-concept paper on DNA malware has already been published by an academic institution, the vulnerabilities it identifies could be patched before DIVS is implemented. Future research into DNA malware must be reviewed before publication in order to avoid advertising vulnerabilities to potential adversaries. Some dual-use biological research is already reviewed by the National Science Advisory Board for Biosecurity, whose purview could be expanded to include publications on DNA malware design and synthesis.[58]

Finally, the United States must monitor the accessibility and cost of synthesis technologies in order to assess whether non-state actors have spoofing and malware capabilities. While DNA malware design and DIVS spoofing presently require the resources of a state actor, given current trends, cheap synthesis and open-source publication will eventually allow non-state actors to exploit these vulnerabilities. At that point, the United States may undertake a strategic reevaluation of future DIVS usage. The United States must also monitor states researching medical genomics to assess whether potential adversaries have a homegrown technological or scientific capacity to exploit US DIVS vulnerabilities.

Conclusion

DNA-based identity verification systems (DIVS) have the potential to be the biometric security systems of the future, offering perfect, one-to-one rapid identification of individuals to authenticators. Current US policies and research practices, however, are creating vulnerabilities for the United States before DIVS is ready for implementation.

Commercial sale of US persons' genetic information is currently unrestricted, and pairing genetic information available on the open market with personal information from data breaches or hacks will allow re-identification of DNA sequences. Subsequently, artificial synthesis of DNA sequences will enable both spoofing of a DIVS and the injection of DNA malware into a secure

network. Without adequate protection of US citizens' genetic and personal information, and with continued open publication on sensitive topics including DNA malware design and synthesis, US implementation of DIVS will fail.

---

[1] Goodman, Marc. *Future Crimes: Everything Is Connected, Everyone Is Vulnerable and What We Can Do About It.* London: Gorgi Books, 2016.

[2] Vallone, Peter. "DNA Biometrics." *NIST*. July 13, 2017. https://www.nist.gov/programs-projects/dna-biometrics.

[3] Le, Chien. "A Survey of Biometrics Security Systems." November 28, 2011. https://www.cse.wustl.edu/~jain/cse571-11/ftp/biomet/#sec3.6.

[4] Vallone, Peter. "DNA Biometrics." *NIST*. July 13, 2017. https://www.nist.gov/programs-projects/dna-biometrics.

[5] For more malware news sources, see: Greenberg, Andy. "Biohackers Encoded Malware in a Strand of DNA." *Wired*. August 10, 2017. https://www.wired.com/story/malware-dna-hack/; Yong, Ed. "These Scientists Took Over a Computer by Encoding Malware in DNA." *The Atlantic.* August 10, 2017. https://www.theatlantic.com/science/archive/2017/08/these-scientists-took-over-a-computer-by-encoding-malware-in-dna/536361/; "Researchers: Hackers Could Encode Human DNA With Malicious Software." *NPR*. August 12, 2017. https://www.npr.org/2017/08/12/542998566/researchers-hackers-could-encode-human-dna-with-malicious-software.

[6] For a discussion of the outlook for biometrics in markets including Latin America, the Asia-Pacific, and Africa, see respectively: Mayhew, Stephen. "Latin Americas Growing Biometric Technology Market." *Biometric Update.* November 25, 2014. http://www.biometricupdate.com/201411/latin-americas-growing-biometric-technology-market; "Rising Asia-Pacific Biometrics Market Looks Set to Overtake North America." *Biometric Technology Today*. February 23, 2015. http://www.sciencedirect.com/science/article/pii/S0969476515300084; and "Key Drivers for Biometrics in Africa in the Next 5 Years." *Iritech, Inc. – A World-Leading Biometrics Technology Provider.* July 21, 2015. http://www.iritech.com/blog/biometrics-africa/.

[7] Seifert, Dan. "Fingerprint, Face Scan, or Password: What's the Best Way to Unlock your Galaxy S8?" *The Verge.* April 21, 2017. https://www.theverge.com/2017/4/21/15360584/samsung-galaxy-s8-unlock-face-iris-fingerprint-scanner-most-secure.

[8] For biometrics in border security in the EU, *see*: "Biopass II: Automated Biometric Border Crossing Systems based on Electronic Passports and Facial Recognition" European Agency for the Management of Operation Cooperation at the External Borders of the Member States of the European Union. 2010. http://frontex.europa.eu/assets/Publications/Research/Biopass_Study_II.pdf. In the United States, *see*: "Biometrics." U.S. Customs and Border Protection. October 30, 2017. https://www.cbp.gov/travel/biometrics. For biometrics in industry, *see*: "Moving Forward with Cybersecurity and Privacy." PwC. 2017. https://www.pwc.com/gx/en/information-security-survey/assets/gsiss-report-cybersecurity-privacy-safeguards.pdf.

[9] For examples in laboratory applications, *see*: "High Security Applications for Biometrics." Allegion. 2014. https://us.allegion.com/content/dam/allegionus2/webdocuments2/Article/High_Security_Applications_for_Biometrics_110129.pdf. For an example application to nuclear facilities, *see* plans detailed in: "Biometric System to Control Access to Nuclear Facilities." The International Science and Technology Center. http://www.istc.int/en/project/0CD7D638EA152E22C3256E00004124B8.

[10] For instance, *see*: Glaser, April. "Biometrics are Coming, Along with Serious Security Concerns." *Wired*. March 9, 2016. https://www.wired.com/2016/03/biometrics-coming-along-serious-security-concerns/ and Brandom, Russell. "Your Phone's Biggest Vulnerability is your Fingerprint." *The Verge*. May 2, 2016. https://www.theverge.com/2016/5/2/11540962/iphone-samsung-fingerprint-duplicate-hack-security.

[11] For a discussion of epigenetic changes, *see*: Hamzelou, Jessica. "Police Can Now Tell Identity Twins Apart – Just Melt Their DNA." *New Scientist.* April 24, 2015. https://www.newscientist.com/article/dn27411-police-can-now-tell-identical-twins-apart-just-melt-their-dna/. For a discussion of deep-sequencing that can differentiate between twins, *see*: Weber-Lehmann et al. "Finding the Needle in the Haystack: Differentiating "Identical" Twins in Paternity Testing and Forensics by Ultra-Deep Next Generation Sequencing." *Forensic Science International* 9 (March 2014): 42-46. https://doi.org/10.1016/j.fsigen.2013.10.015. http://www.sciencedirect.com/science/article/pii/S1872497313002275.

[12] *See*: Betancourt, Dario. "Why is DNA Biometric Technology Is The Most Commonly Used." *Biometric News Portal*. October 5, 2017. https://www.biometricnewsportal.com/dna_biometrics.asp. For more about the potential for real-time sequencing, especially nanopore sequencing, *see*: Gottschling, Irimia R. "A Decade's Perspective on DNA Sequencing Technology." *Nature* 470 (February 9, 2011): 198-203. doi: 10.3897/bdj.4.e7720.figure2f.

[13] For a discussion of how this cost has changed over time, *see*: "The Cost of Sequencing a Human Genome." *National Human Genome Research Institute*. July 6, 2016. https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/.
[14] For the original draft of the human genome, *see*: International Human Genome Sequencing Consortium. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409 (February 15, 2001): 860-921. https://www.nature.com/articles/35057062.
[15] For the figure of sequencing in one hour, *see*: Fikes, Bradley. "New Machines Can Sequence Human Genome in One Hour, Illumina Announces." *The San Diego Union-Tribune*. January 9, 2017. http://www.sandiegouniontribune.com/business/biotech/sd-me-illumina-novaseq-20170109-story.html.
[16] For how U.S. Special Operations Command is planning to use DNA readers for biometric ID purposes, *see*: Tucker, Patrick. "Special Operators are Using Rapid DNA Readers." *Defense One.* May 20, 2015. http://www.defenseone.com/technology/2015/05/special-operators-are-using-rapid-dna-readers/113383/.
[17] For a brief discussion of Saudi Arabia's use of DNA, *see*: "DNA Testing in Saudi Arabia." *DNA Diagnostics Center*. 2018. https://dnacenter.com/immigration-dna-testing/dna-testing-in-saudi-arabia/.
[18] Reference: "The Collection of DNA from Military Personnel." *Council for Responsible Genetics.* http://www.councilforresponsiblegenetics.org/geneticprivacy/DNA_mil.html.
[19] Kuwait's collection has since been shut down, although the future of the DNA that has already been logged is unclear: "Kuwait: Court Strikes Down Draconian DNA Law." *Forensic Genetics Policy Initiative*. October 17, 2017. http://dnapolicyinitiative.org/kuwait-court-strikes-down-draconian-dna-law/.
[20] When not all of the bodies could be identified, the Kuwaiti government passed legislation for the creation of a DNA database on all Kuwaiti citizens and implemented it, starting collection that year. *See*: "Kuwait: Court Strikes Down Draconian DNA Law." *Forensic Genetics Policy Initiative*. October 17, 2017. http://dnapolicyinitiative.org/kuwait-court-strikes-down-draconian-dna-law/.
[21] Ibid.
[22] For some explanation of US judicial databases, *see*: "Frequently Asked Questions on CODIS and NDIS." Federal Bureau of Investigation. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/codis-and-ndis-fact-sheet, "Biometric Analysis." Federal Bureau of Investigation. https://www.fbi.gov/services/laboratory/biometric-analysis, "Combined DNA Index System (CODIS)." Federal Bureau of Investigation. https://www.fbi.gov/services/laboratory/biometric-analysis/codis.
[23] Reference: Dr. Wallace, Helen. "International Standards for DNA Databases." GeneWatch UK. http://www.ncl.ac.uk/media/wwwnclacuk/policyethicsandlifesciences/files/Helen_Wallace_Forensic_DNA_databases.pdf.
[24] Reference: "Russia Steps up Crime Fighting with Creation of DNA Database." *RT*. October 23, 20122. https://www.rt.com/news/russia-dna-criminal-database-161/.
[25] This database, although legal in basis, may also be used to target specific populations, "focus personnel," or "migrants." *See*: "Privacy Concerns as China Expands DNA Database." *BBC News*. May 17, 2017. http://www.bbc.com/news/world-asia-china-39945220.
[26] *See*: "Global Summary." *DNA Policy Initiative*. August 31, 2017. http://dnapolicyinitiative.org/wiki/index.php?title=Global_summary. Typically, limited genetic analysis is conducted and the data may or may not be retained indefinitely, with the duration of sample retention and difficulty of data removal varying widely among countries. For more on different standards regarding legal collection, *see*: http://www.councilforresponsiblegenetics.org/dnadata/fullreport.pdf.
[27] *See*: "Rights Group Criticizes China for Mass DNA Collection in Xinjiang." *Reuters*. December 13, 2017. https://www.reuters.com/article/us-china-xinjiang/rights-group-criticizes-china-for-mass-dna-collection-in-xinjiang-idUSKBN1E71DX.
[28] For more on ethnic markers in DNA, *see*: "How Does a Patient's Ethnic Background Affect Matching?" *Be the Match.* https://bethematch.org/transplant-basics/matching-patients-with-donors/how-does-a-patients-ethnic-background-affect-matching/.
[29] For more on this concept, *see* Marc Goodman's book *Future Crimes.* Goodman, Marc. *Future Crimes: Everything Is Connected, Everyone Is Vulnerable and What We Can Do About It*. London: Gorgi Books, 2016. Russia has also accused the US of pursuing such goals. *See*: "Russians' DNA Taken by Foreign Agents, Kremlin Says." *BBC News*. October 31, 2017. http://www.bbc.com/news/world-europe-41816857.
[30] For more on the US Precision Medicine Initiative, established in 2015, *see*: "The Precision Medicine Initiative." *National Archives and Records Administration*. https://obamawhitehouse.archives.gov/node/333101. For more information on the Chinese equivalent, and its impact on Chinese genetic analysis capacity *see*: Perez, Bien. "China's 'precision medicine' Initiative Gets Lift from Latest Genomics Company Funding." *South China Morning*

Post. May 2, 2017. http://www.scmp.com/tech/china-tech/article/2092362/chinas-precision-medicine-initiative-gets-lift-latest-genomics.

[31] The bulk of sequencing has been carried out using technology from Illumina, a major biotechnology firm; *see*: Yong, Ed. "A DNA Sequencer in Every Pocket." *The Atlantic.* April 28, 2016. https://www.theatlantic.com/science/archive/2016/04/this-technology-will-allow-anyone-to-sequence-dna-anywhere/479625/, and Fikes, Bradley. "New Machines Can Sequence Human Genome in One Hour, Illumina Announces." *The San Diego Union-Tribune*. January 9, 2017. http://www.sandiegouniontribune.com/business/biotech/sd-me-illumina-novaseq-20170109-story.html.

[32] For information on China's investment into sequencing technologies, *see*: Baker, Monya. "China Buys US Sequencing Firm." *Nature*. September 25, 2012. https://www.nature.com/news/china-buys-us-sequencing-firm-1.11472, and Molteni, Megan. "A Chinese Genome Giant Sets its Sights on the Ultimate Sequencer." *Wired*. May 18, 2017. https://www.wired.com/2017/05/chinese-genome-giant-sets-sights-uitimate-sequencer/.

[33] China's investment into a homegrown capacity for DNA sequencing has been a particular cause for US concern, *see*: Scharre, Paul. "Testimony Before the House Armed Services Subcommittee on Emerging Threats and Capabilities: China's Pursuit of Emerging and Exponential Technologies." *Center for a New American Security*. January 9, 2018. http://docs.house.gov/meetings/AS/AS26/20180109/106756/HHRG-115-AS26-Wstate-ScharreP-20180109.pdf.

[34] Scharre, Paul. "Testimony Before the House Armed Services Subcommittee on Emerging Threats and Capabilities." *CNAS*. January 9, 2018. https://www.cnas.org/publications/congressional-testimony/paul-scharre-testimony-before-hasc.

[35] A public promotion was initially scheduled in order to encourage fans to seek out Origin, a company offering ancestry breakdowns based on DNA markers; however, it was ultimately delayed and may be cancelled due to privacy concerns. *See*: Pryce, Meghan. "Promotion Offering DNA Test Kits to Ravens Fans to be Rescheduled." *The Baltimore Sun*. September 17, 2017. http://www.baltimoresun.com/business/bs-bz-ravens-dna-rescheduled-20170917-story.html. Despite the cancellation, such a promotion highlights the degree to which the US public may be willing to yield DNA samples without a detailed understanding of the consequences of their actions. Fans were disappointed that the promotion was cancelled, while privacy activists – academics and lobbyists – were thrilled. Such businesses include 23andMe (https://www.23andme.com/), Ancestry (ancestry.com), National Geographic's Geno 2.0 (https://genographic.nationalgeographic.com/), and Nutrigene (https://www.mynutrigene.com/)

[36] Such data commercialization has been largely accepted by the public; however, recent journalism has started to question the premise of selling DNA information. For example, *see*: Schulson, Michael. "Spit and Take." Slate. http://www.slate.com/articles/technology/future_tense/2017/12/direct_to_consumer_genetic_testing_has_tons_of_privacy_issues_why_is_the.html.

[37] Ibid.

[38] Several paths have been identified that allow re-identification of DNA sequences. For more, *see*: "Are Guarantees of Genome Anonymity Realistic?" Personal Genome Project – Open Humans Foundation. December 2003. http://arep.med.harvard.edu/PGP/Anon.htm.

[39] Forensic analysis uses DNA phenotyping to better enable re-identification. Phenotyping is the use of genetic sequences to project physical characteristics. For more on the use of DNA phenotyping in forensics, *see*: Augustine, Seth. "DNA Phenotyping Recreates the Face of an Alleged Serial Killer." *Forensic Magazine*. August 30, 2016. https://www.forensicmag.com/article/2016/08/dna-phenotyping-recreates-face-alleged-serial-killer.

[40] For a plethora of resources on STR-based identification, *see*: "STR Training Materials." NIST. https://strbase.nist.gov/training.htm.

[41] To clarify, whole-genome sequencing still relies on differences at a single base pair level, known as SNPs. *See*: "Telling the Difference Between Identical Twins" Illumina. August 3, 2015. https://www.illumina.com/company/news-center/feature-articles/telling-the-difference-between-identical-twins.html. It is these SNPs at known locations in the human genome that are used to ID individuals, as the computing power required to process base-by-base comparisons of the full human genome would be staggering, *see*: Vallone, Peter. "Underlying Sequence Variation within STRs: Considerations for Nomenclature, Storage, Searching, and Reporting." NIST. https://strbase.nist.gov/pub_pres/ISHI_NGS_Workshop_2015_Vallone.pdf.

[42] See: Afolabi, A., and O. Akintaro. "Design of DNA Based Biometric Security System for Examination Conduct." *Journal of Advances in Mathematics and Computer Science* 23, no. 6 (2017): 1-7. doi:10.9734/jamcs/2017/27251; "RapidHIT ID" IntegenX, Inc. 2016. https://integenx.com/wp-content/uploads/RapidHIT-ID-Brochure-Desktop-DNA-is-here.pdf.

[43] For more on the scope of the direct-to-consumer genomics, see: Hogarth, Stuart and Saukko, Paula. "A Market in the Making: The Past, the Present, and the Future of Direct-to- Consumer Genomics" *New Genetics and Society* 36, no. 3 (2017): 197-208. https://doi.org/10.1080/14636778.2017.1354692.
[44] Scharre, Paul. "Testimony Before the House Armed Services Subcommittee on Emerging Threats and Capabilities." *CNAS*. January 9, 2018. https://www.cnas.org/publications/congressional-testimony/paul-scharre-testimony-before-hasc.
[45] Cost of synthesizing a human genome currently: with synthesis prices around $0.10/bp, $320 million dollars. Even with the discounts offered by some services, e.g. "Gene Synthesis Service." GenScript. https://www.genscript.com/gene_synthesis.html, this is clearly beyond the reach of an interested garage scientist.
[46] For a description of trends in synthesis, *see*: Goldberg, Martin. "Applying Moore's Law to DNA Synthesis." *Genetic Engineering and Biotechnology News*. February 20, 2013. https://www.genengnews.com/gen-articles/applying-moore-39-s-law-to-dna-synthesis/4739/?kwrd=monsanto. For a discussion of synthesis futures, *see*: Carlson, Robert et al. "Genome Synthesis and Design Futures: Implications for the US Economy." *Bio-era*. http://www.bio-era.net/reports/genome.html.
[47] Smith, Hamilton et al. "Generating a Synthetic Genome by Whole Genome Assembly: φX174 Bacteriophage from Synthetic Oligonucleotides." *Proc Natl Acad Sci USA* (December 2003): 15440–15445. doi: 10.1073/pnas.2237126100, and Mueller, Steffen, J. Robert Coleman, and Eckard Wimmer. "Putting Synthesis into Biology: A Viral View of Genetic Engineering through De Novo Gene and Genome Synthesis." *Chemistry & Biology*16, no. 3 (2009): 337-47. doi:10.1016/j.chembiol.2009.03.002.
[48] For more on the challenges associated with attempting to assemble a large genome from fragments, *see*: Henson, Joseph et al. "Next-Generation Sequencing and Large Genome Assemblies" Pharmacogenomics, no. 8 (2012): 901-215. doi: 10.2217/pgs.12.72
[49] Ney, Peter et al. "Computer Security, Privacy, and DNA Sequencing: Compromising Computers with Synthesized DNA, Privacy Leaks, and More." 2017 USENIX Security Symposium. August 16, 2017. https://dnasec.cs.washington.edu/dnasec.pdf.
[50] "National Defense Authorization Act for Fiscal Year 2018." Committee on Armed Services, House of Representatives. 76-77. July 6, 2017. https://www.congress.gov/115/crpt/hrpt200/CRPT-115hrpt200.pdf.
[51] *See*: Moon, Mariella. "FBI Nabs Chinese National Linked to massive OPM Hack. *Engadget*. August 25, 2017. https://www.engadget.com/2017/08/25/fbi-nabs-chinese-national-opm-hack/.
[52] *See*: Clements, Nick. "Equifax's Enormous Data Breach Just Got Even Bigger." *Forbes*. March 5, 2018. https://www.forbes.com/sites/nickclements/2018/03/05/equifaxs-enormous-data-breach-just-got-even-bigger/#367955de53bc.
[53] Equifax hack attribution is unclear at best; *see*: Riley, Michael. "The Equifax Hack Has the Hallmarks of State-Sponsored Pros." *Chicago Tribune*. October 2, 2017. http://www.chicagotribune.com/business/ct-equifax-hack-state-sponsored-pros-20171002-story.html.
[54] Landi, Heather. "2017 Breach Report: 477 Breaches, 5.6M Patient Records Affected" *Healthcare Informatics*. January 23, 2018. https://www.healthcare-informatics.com/news-item/cybersecurity/2017-breach-report-477-breaches-56m-patient-records-affected.
[55] *See*: Mearian, Lucas. "Hackers are Coming for Your Healthcare Records – Here's Why." *Computer World*. June 30, 2016. https://www.computerworld.com/article/3090566/healthcare-it/hackers-are-coming-for-your-healthcare-records-heres-why.html, Sweeney, Brigid. "The Frightening New Frontier for Hackers: Medical Records." *Modern Healthcare.* April 10, 2017. http://www.modernhealthcare.com/article/20170410/NEWS/170419987, and Yaraghi, Niam. "Hackers, Phishers, and Disappearing Thumb Drives: Lessons Learned from Major Health Care Data Breaches. *Brookings*. May 5, 2016. https://www.brookings.edu/research/hackers-phishers-and-disappearing-thumb-drives-lessons-learned-from-major-health-care-data-breaches/.
[56] In fact, the University of Washington's genetic research is conducted in direct partnership with China. *See*: Molteni, Megan. "A Chinese Genome Giant Sets its Sights on the Ultimate Sequencer. *Wired*. May 18, 2017. https://www.wired.com/2017/05/chinese-genome-giant-sets-sights-uitimate-sequencer/.
[57] For more on current screening procedures to identify potential pathogenic elements in requested synthetic DNA sequences, *see*: "Where Gene Synthesis and Biosecurity Align." International Gene Synthesis Consortium. https://genesynthesisconsortium.org/; "Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA." Health and Human Services Department. October 13, 2010. https://www.federalregister.gov/documents/2010/10/13/2010-25728/screening-framework-guidance-for-providers-of-synthetic-double-stranded-dna.

[58] NSABB information is available online, *see: "*National Science Advisory Board for Biosecurity (NSABB)" National Institutes of Health: Office of Science Policy. https://osp.od.nih.gov/biotechnology/national-science-advisory-board-for-biosecurity-nsabb/.

The Project on International Peace and Security (PIPS)
Institute for the Theory and Practice of International Relations
The College of William and Mary