

cans are a self-antigen, and so antibodies to such antigens are selected against). When broadly neutralizing antibodies to HIV-1 do arise, they display a number of highly unusual features, including high levels of somatic mutation. These features allow the antibodies to interact with the self-antigen glycan shield and to reach past it to contact foreign peptidic determinants on the HIV-1 spike protein (11). Thus, many of these antibodies recognize a particular combination of self- and foreign antigens, and their development may involve redemption of self-reactive B cells through mutation, as described by Burnett *et al.*

Consistent with a requirement for some level of self-reactivity, broadly neutralizing antibodies to HIV-1 frequently demonstrate cross-reactivity to self-antigens (12). More generally, poly-reactivity was found in 75% of a large collection of human monoclonal antibodies to HIV (13). The notion that some of these antibodies arise from self-reactive precursors is supported by antibody gene knock-in experiments that have demonstrated that B cells that express predicted germline versions of broadly neutralizing antibodies to HIV-1 frequently show precursor cell deletion or absence of allelic exclusion that is indicative of self-reactivity (14, 15). However, when present at high precursor frequencies and challenged with high-affinity multivalent antigen, the knock-in cells can participate in immune responses (14, 15).

The findings of Burnett *et al.* further our understanding of the biology of B cell anergy and provide a framework for thinking about why such cells might be allowed to persist in immune systems despite their self-reactivity. ■

REFERENCES AND NOTES

1. D. L. Burnett *et al.*, *Science* **360**, 223 (2018).
2. H. Wardemann *et al.*, *Science* **301**, 1374 (2003).
3. D. Gay, T. Saunders, S. Camper, M. Weigert, *J. Exp. Med.* **177**, 999 (1993).
4. D. A. Nemaee, K. Burki, *Nature* **337**, 562 (1989).
5. C. C. Goodnow *et al.*, *Nature* **334**, 676 (1988).
6. G. D. Victora, M. C. Nussenzweig, *Annu. Rev. Immunol.* **30**, 429 (2012).
7. C. T. Mayer *et al.*, *Science* **358**, eaao2602 (2017).
8. J. G. Cyster, S. B. Hartley, C. C. Goodnow, *Nature* **371**, 389 (1994).
9. T. Tiller *et al.*, *Immunity* **26**, 205 (2007).
10. T. D. Chan *et al.*, *Immunity* **37**, 893 (2012).
11. A. Escolano, P. Dosenovic, M. C. Nussenzweig, *J. Exp. Med.* **214**, 3 (2017).
12. B. F. Haynes *et al.*, *Science* **308**, 1906 (2005).
13. H. Mouquet *et al.*, *Nature* **467**, 591 (2010).
14. R. Zhang *et al.*, *Sci. Transl. Med.* **8**, 336ra362 (2016).
15. R. K. Abbott *et al.*, *Immunity* **48**, 133 (2018).

ACKNOWLEDGMENTS

We thank members of the Nussenzweig laboratory for comments. E.E.K. is supported by a National Health and Medical Research Council C. J. Martin Fellowship; M.C.N. is a HHMI investigator.

10.1126/science.aat5758



GENETICS

Crowdsourced genealogies and genomes

Genealogical study provides insight into history and life span and heralds crowdsourced genetic research

By Alexandre A. Lussier¹ and Alon Keinan^{1,2}

Genealogies are likely the first, centuries-old “big data,” with their construction as old as human civilization. Recent renewed interest led to the largest genealogical websites (Ancestry.com, MyHeritage, and Geni) amassing 130 million users who generated billions of online genealogical profiles, offering ample research opportunities that would otherwise require extensive recruitment. On page 171 of this issue, Kaplanis *et al.* (1) showcase the research potential of this type of crowdsourced data, studying genealogies based on processing 86 million public Geni profiles.

An important research tool throughout human history, genealogical studies reach from anthropology to modern genetics and medicine. Of note are centuries of Icelandic genealogical enthusiasts combining family information with early-age scriptures. The founding of deCODE genetics sped up the process to create an online genealogy of 864,000 Icelanders, the Íslendingabók, by 2003. It has proven an incredible research tool, for example, in studying the relative roles of genetic heritability and shared environment in many complex diseases and other traits (2). Another unique genealogy has been constructed from historical records by the Mormon Church since 1921. Continued more

recently by hundreds of thousands of volunteers, they created about half a million new profiles daily, with many connected to health care records (3)—a true crowdsourcing effort.

Entering the age of genomics, genealogy enthusiasts greeted a new tool. Direct-to-consumer (DTC) genetic testing companies all provide a service for finding relatives, while obtaining powerful, crowdsourced genome-wide data for millions of individuals. AncestryDNA and 23andMe applied their data to study migrations, structure, and admixture of U.S. populations (4, 5). However, most crowdsourced genetic research is medically driven, focusing on the genetic basis of complex traits. For example, a recent study describing how genetic risk factors are shared across many traits included analyses of 23andMe customers for 17 of the traits (6).

Kaplanis *et al.* demonstrate the potential of large-scale genealogies, although with no genetic data, but at the hands of statistical and population geneticists. Extensively processing and validating genealogical data, they compiled 5.3 million genealogies, including one with 13 million individuals that often depicts at least 20 generations.

They analyze relatedness and distance at birth between married couples. Distance for most was less than 10 km before the Industrial Revolution (1750), followed by a gradual increase, which then accelerated to over 100 km after the start of the Second Industrial Revolution (1870). Average relatedness remained the same (equivalent to fourth cousins) prior to the Second Industrial Revolution, when it began decreasing in line with increasing distance. Kaplanis *et al.* postulate that recently decreased relatedness is due to

¹Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. ²Cornell Center for Comparative and Population Genomics, Center for Vertebrate Genomics, and Center for Enervating Neuroimmune Disease, Cornell University, Ithaca, NY 14853, USA. Email: alon.keinan@cornell.edu

shifting cultural norms, rather than increased distance, because of the inconsistent relationship between relatedness and distance. This appears concurrent with popular writing from the time, which led 13 U.S. states to pass cousin marriage prohibitions by the 1880s (7) (although more distant relatedness is in question in Kaplanis *et al.*). A related study considered 160,000 couples in the Íslendingabók to show that increased couple relatedness (equivalent to third or fourth cousins) is associated with higher fertility that is not explained by socioeconomic influences on number of offspring, and hence is claimed to have a potential biological basis (8).

The main results of Kaplanis *et al.* involve life span. The resolution of the data set allows them to discern not only that average life span decreased during World War I and World War II, but also that the decrease was larger for individuals of military age. Despite these major events, life span appears to have increased at an almost constant rate of ~4 years per generation since ~1850. They conducted a meticulous study of factors affecting life span, attributing ~7% to gender, birth year, and geography combined. They estimated life span heritability at $16.1 \pm 0.4\%$, lower than most previous studies, although among them, the largest genealogy-based study until now provided a comparable estimate of $15 \pm 3\%$ in the Mormon genealogy (9). Kaplanis *et al.* estimate that an additional ~4% of life span is attributable to dominance (where having a single copy of a genetic variant constitutes the majority of the effect of having two) and none to interaction between different genetic variants.

Despite extensive analyses, Kaplanis *et al.* only scratch the surface of their resource, which is publicly available, stripped of personal information. It may be interesting to reanalyze life-span factors focused on very high longevity, and revisit other questions previously studied with smaller genealogies. The resource may benefit many disciplines, with unique promise in the combination with genetic data of the same individuals, an opportunity that led to large investments in DTC genetic services by the companies with the largest genealogical websites.

DTC genetic data are not publicly available, but Kaplanis *et al.* provide an academic version of their resource where individuals can consent to being identified. It can be used on websites to which participants upload their genetic data, as Kaplanis *et al.* implemented in DNA.Land, which, for example, collates family history of breast cancer and allows users to contribute their genomes to the National Breast Cancer Coalition (10). In a recent study, deCODE genetics highlighted yet again the power of large-scale genealogies with matched genetic data. They reconstructed

an ancestor's genome by mining descendants for inherited genetic fragments, which they tested via unique genealogical analyses (11).

One critical limitation of available crowdsourced data is that the "crowd" is mostly from 15% of the worldwide population that comprises Europe and North America. The overwhelming majority of DTC genetic testing customers are from these regions, as are 85% of the profiles in the Kaplanis *et al.* study. Partly due to local laws and consent, the potential unleashed by integrating worldwide diversity should provide an incentive to overcome these obstacles. Another shortcoming is the underutilization of the X chromosome by DTC genetic companies for both customer services and medical research (12). Its inclusion via newly developed analytical methods may improve these and, importantly, provide a key step toward closing the gender disparity in disease diagnosis and treatment (12).

The era of precision medicine heralds a greater potential for crowdsourcing, with distinct opportunities when familial, genetic, and medical data are integrated. Funding details of large-scale endeavors such as the U.S. National Institutes of Health *All of Us* program have put an effective price tag on the recruitment of each participant, their genetic data, and medical records. Recently founded companies, in turn, are attempting to resurrect the option for participants to lease their data to researchers. This may increase the potential for research based on crowdsourced, although not fully crowd-funded, data.

Beyond explosive growth of DTC genetic testing services (of ~16 million current customers, almost two-thirds joined since early 2017), whole-genome sequencing will likely become a cost-effective DTC choice within 2 to 3 years. This will enable tracing, and flagging of potentially harmful, *de novo* mutations in families and allow crowdsourced genetic research to more substantially advance disease risk prediction, diagnosis, and treatment.

Although many fields make use of crowdsourcing, none is better positioned, since all 7.5 billion of us have a genealogy, DNA, traits, and medical information to share. ■

REFERENCES

1. J. Kaplanis *et al.*, *Science* **360**, 171 (2018).
2. N. Zaitlen *et al.*, *PLoS Genet.* **9**, e1003520 (2013).
3. S. Knight *et al.*, *Hum. Hered.* **81**, 1 (2016).
4. E. Han *et al.*, *Nat. Commun.* **8**, 14238 (2017).
5. K. Bryc *et al.*, *Am. J. Hum. Genet.* **96**, 37 (2015).
6. J. K. Pickrell *et al.*, *Nat. Genet.* **48**, 709 (2016).
7. D. B. Paul, H. G. Spencer, *PLoS Biol.* **6**, e320 (2008).
8. A. Helgason *et al.*, *Science* **319**, 813 (2008).
9. R. A. Kerber *et al.*, *J. Gerontol. A Biol. Sci. Med. Sci.* **56**, B130 (2001).
10. J. Yuan *et al.*, *Nat. Genet.* **50**, 160 (2018).
11. A. Jagadeesan *et al.*, *Nat. Genet.* **50**, 199 (2018).
12. Editorial, *Nat. Med.* **23**, 1243 (2017).

10.1126/science.aat2634

NANOMATERIALS

A recipe for nanoporous graphene

Nanoporous graphene created from molecular precursors shows promise for electronic applications

By Alexander Sinitskii

Graphene is widely regarded as a promising material for electronic applications because the exceptionally high mobilities of its charge carriers enable extremely fast transistors (1). However, the lack of an energy band gap in graphene limits its use in logic applications; without a band gap, the devices remain highly conductive at any gate voltage and thus cannot be fully switched off. Researchers have therefore turned their attention to semiconducting forms of graphene that have the necessary band gap to enable transistors with high on-off ratios. On page 199 of this issue, Moreno *et al.* (2) report on the synthesis and device characterization of nanoporous graphene with semiconducting properties.

So far, researchers mostly have focused their attention on a different class of graphene-based materials that could have the desired energy band gap: one-dimensional graphene nanoribbons. Theoretical studies have shown that depending on their structural parameters, such as shape, width, and edge structure, graphene nanoribbons may possess not only a tunable electronic band gap (3), but also other intriguing physical properties, such as edge magnetism and highly localized electronic states. However, carving few-nanometer-wide strips is not the only nanostructuring approach that could create a band gap in graphene. The same can be achieved by patterning an array of closely spaced nanoscopic holes in graphene, thus forming a nanoporous graphene. Similar to nanoribbons, the properties of nanoporous graphene strongly depend on their structural parameters, such as the pore diameter and the periodicity of the structure (4).

Department of Chemistry and Nebraska Center for Materials and Nanoscience, University of Nebraska–Lincoln, Lincoln, NE 68588, USA. Email: sinitskii@unl.edu

Crowdsourced genealogies and genomes

Alexandre A. Lussier and Alon Keinan

Science **360** (6385), 153-154.
DOI: 10.1126/science.aat2634

ARTICLE TOOLS

<http://science.sciencemag.org/content/360/6385/153>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/360/6385/171.full>

REFERENCES

This article cites 12 articles, 2 of which you can access for free
<http://science.sciencemag.org/content/360/6385/153#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.