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 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...
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 1860s (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 ± 0.4%, lower than most previous studies, although not fully crowdfunded, 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

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

Published by AAAS

10.1126/science.aat2634
Crowdsourced genealogies and genomes
Alexandre A. Lussier and Alon Keinan

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