



Published in final edited form as:

Science. 2013 March 1; 339(6123): 1032–1033. doi:10.1126/science.339.6123.1032-c.

Data re-identification: societal safeguards

Russ B. Altman¹, Ellen Wright Clayton², Isaac S. Kohane³, Bradley A. Malin^{4,*}, and Dan M. Roden⁵

¹Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

²Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, TN 37232, USA

³Department of Pediatrics, Children's Hospital Boston, Boston, MA 02115, USA

⁴Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37232, USA

⁵Department of Medicine, Vanderbilt University, Nashville, TN 37232, USA

Head

The Report by M. Gymrek *et al.* ("Identifying personal genomes by surname inference," 18 January, p. 321) demonstrates that DNA samples can be combined with surname and other data to re-identify seemingly anonymous records. The study adds to the literature showing that large, publicly available data sets can be leveraged to infer personal information (1) and identify unique individuals (2). In 2005, a 15-year-old tracked down his sperm donor father using a similar approach (3). This demonstration may undermine individual research subjects' confidence that their DNA can be shared in a way that is not re-identifiable. To counteract this effect, we must augment imperfect technical safeguards with measures that make such re-identification socially, legally, and economically unacceptable.

Society can work in many ways to mitigate risks and maintain a climate of confidence that will continue to encourage research subject participation. Subjects must be informed of the risk that their DNA sequence will be identified with their names. Data stewards must make a good faith effort to protect DNA-based records, including continuing to de-identify the data (e.g., remove explicit identifiers). Systems for audit of access and use of the data should be routine, and data should be protected from anonymous access. Data use agreements should provide institutional and legal remedies when societal boundaries and robust research norms of respect for the privacy of individuals are violated.

The diversity of human genomes guarantees that each person will harbor markers for higher-than-average risk for some adverse outcomes. Thus, further risk mitigation should include regulatory approaches such as expansion of protections currently provided under the Genetic Information Nondiscrimination Act and other anti-discrimination laws, which currently protect against some forms of discrimination, but not others, such as long-term care or life insurance.

References

1. El Emam K, Jonker E, Arbuckle L, Malin B. PLoS One. 2011; 6:e28071. [PubMed: 22164229]
2. Jobling M. Trends Genet. 2001; 17:353. [PubMed: 11377798]
3. Sample, I. Teenager finds sperm donor dad on Internet. The Guardian. Nov 2. 2005
www.guardian.co.uk/science/2005/nov/03/genetics.news

*To whom correspondence should be addressed. b.malin@vanderbilt.edu.