Published in final edited form as:

Am J Med. 2020 January ; 133(1): 143–146.e2. doi:10.1016/j.amjmed.2019.04.052.

Consistency of Direct to Consumer Genetic Testing Results Among Identical Twins

Anne M. Huml, MD^{1,2,3}, Catherine Sullivan, MS¹, Maria Figueroa¹, Karen Scott, MPA¹, Ashwini R. Sehgal, MD^{1,2,4}

¹Center for Reducing Health Disparities, Case Western Reserve University

²Division of Nephrology, Department of Medicine, MetroHealth Medical Center

³Glickman Urological and Kidney Institute, Cleveland Clinic

⁴Departments of Bioethics and Population and Quantitative Health Sciences, Case Western Reserve University Cleveland, Ohio

Abstract

Purpose: To evaluate the consistency of 3 commonly used direct to consumer genetic testing kits.

Background: Genetic testing kits are widely marketed by several companies but the consistency of their results is unclear. Since identical twins share the same DNA, their genetic testing results should provide insight into test consistency.

Methods: 42 identical twins (21 pairs) provided samples for three testing companies. Outcomes were concordance of ancestry results when i) twin pairs were tested by the same company and ii) the same participant was tested by different companies. Concordance of 8 self-reported traits with 23andMe genetic analyses were also examined.

Results: Concordance of ancestry results when twin pairs were tested by the same company was high, with mean percent agreement ranging from 94.5%–99.2%. Concordance of ancestry results when participants were tested by two different companies was lower, with mean percent agreement ranging from 52.7%–84.1%. Concordance of trait results was variable, ranging from 34.1% for deep sleep and detached earlobes to 90.2% for cleft chin.

Conclusion: The consistency of consumer genetic testing is high for ancestry results within companies but lower and more variable for ancestry results across companies and for specific traits. These results raise questions about the usefulness of such testing.

Keywords

genetic testing; consistency; identical twins

Introduction

An estimated 12 million Americans participated in direct to consumer genetic testing in 2017.¹ To get ancestry information, consumers submit a saliva or cheek swab sample. Testing companies also provide information on genetic traits and susceptibility to certain disease states. The outcomes of such testing may have implications for consumers and their health care providers, e.g. regarding measures to address specific health risks.² Genetic testing results across companies may not be comparable, and consistency of testing is not well understood. Since identical twins share the same DNA, their genetic testing results can provide insight into test consistency.

Methods

Participants

We recruited identical twins at Twins Days, an annual festival held in Twinsburg, Ohio. Eligible twin pairs were over 18 years of age and spoke English. We administered a validated zygosity questionnaire which confirmed that all participants were identical rather than fraternal twins.³

Genetic Testing Kits

We purchased kits from three popular direct to consumer genetic testing companies, including 23andMe, Ancestry, and MyHeritage. All three provide ancestry testing while 23andMe also offers testing for specific traits, e.g. deep sleep. After obtaining informed consent, study staff obtained samples and mailed the kits to each company. To protect confidentiality, each kit was registered and submitted using a fictitious name. The sample sent from one participant to 23andMe did not contain enough genetic material to be analyzed even when a second sample was submitted. The study was approved by the Institutional Review Board of MetroHealth Medical Center.

Questionnaires

Each participant completed a questionnaire on demographic characteristics and specific traits. We selected traits that would be obvious to participants regardless of their age.

Statistical Analysis

Concordance of ancestry was determined when i) twin pairs were tested by the same company and ii) the same participant was tested by different companies. When twin pairs were tested by the same company, concordance was calculated by comparing ancestry percentages for each reported ethnicity and using the lower of the two percentages as a measure of percent agreement. For example, if twin 1 was reported to have 20% Italian ancestry and twin 2 was reported to have 40% Italian ancestry, the percent agreement for Italian ancestry would be 20%. Similarly, if twin 1 was reported to have 80% Scandinavian ancestry and twin 2 was reported to have 60% Scandinavian ancestry, then the percent agreement for Scandinavian ancestry would be 60%. The total percent agreement was determined by adding across all ethnicities, in this example 20+60=80%. A similar approach was used to calculate concordance when the same participant was tested by different

companies. Because the three companies had somewhat different ways to categorize ethnicity, we created 13 common ethnicity categories in order to compare across companies (Appendix Table 3). We used each company's published tables and maps to re-categorize company reported ethnicity into these common ethnicity categories. Appendix Table 4 provides an example of ancestry concordance calculation using this re-categorization. Note that some of the 23andMe reported ethnicities were presented as broad regions, e.g. Broadly European. To be conservative in our concordance calculations, we distributed these results into common ethnicity categories to maximize concordance. Appendix Table 5 provides an example of this process. We anticipated a 60% concordance rate when participants were tested by different companies. To estimate this proportion with a precision of +15% and a 95% confidence level requires a sample size of 41.⁴ All analyses were performed using JMP Pro statistical software, version 14.0 (SAS Institute, Cary, NC).

Results

Participant Characteristics

42 participants (21 twin pairs) participated. Their mean age was 38.4 years (range 18–72 years). All were non-Hispanic white and 36 (85.6%) were female.

Ancestry Testing

The concordance of ancestry results when twin pairs were tested by the same company was high, with mean percent agreement ranging from 94.5%–99.2% (Table 1). The concordance of ancestry when participants were tested by different companies was lower, with mean percent agreement ranging from 52.7%–84.1%.

Trait Testing

For the health traits reported by 23andMe, the concordance between self-reported traits and genetic testing results varied greatly (Table 2). 21 participants reported deep sleep but only 6 genetic analyses reported this trait while 14 participants reported detached earlobes but all 41 genetic analyses reported this trait (concordance of 34.1%) Only 4 participants reported cleft chin and none of the genetic analyses reported this trait (concordance 90.2%).

Discussion

We found that the concordance of ancestry results when twin pairs were tested by the same company was high while the concordance when participants were tested by different companies was lower. The concordance between self- and company-reported traits were variable. Strengths of this study include the involvement of identical twins, performance of multiple tests on each participant, and use of three different widely available direct to consumer genetic testing kits.

Ancestry testing is based upon company-specific reference datasets. These datasets are comprised of the genotypes of previous consumers of each specific testing company as well as publically available genotypes from large scale sequencing projects. A new consumer's sample undergoes DNA sequencing and is compared to a reference dataset to determine

ancestry estimates.^{5–11} The differences among these reference datasets may be one explanation for the results of our study. Trait testing is based on genome-wide association sequencing that focuses on differences in small nucleotide polymorphisms.^{12,13}

Previous reports on the consistency of direct to consumer genetic ancestry testing among various companies have been limited to single cases. For example, a set of identical twins submitted samples simultaneously to five separate direct to consumer testing companies. The results were similar to our study findings, with good correlations when the twin pair was tested by the same company but not when they were tested by different companies.¹⁴

The results of our study have implications for patients, providers, policy makers, and researchers. A survey of genetic testing customers found that they expect information about their own genetics that are based on scientific evidence.¹⁵ Providers may be asked by patients to help explain results for a test that they did not order. Providers may also be faced with ordering diagnostics for asymptomatic patients based on genetic test results. There may be legal and insurance implications of test results.^{16,17} Policy makers may need to set performance standards before tests are marketed. Researchers should further examine the reliability and validity of genetic ancestry and trait testing.

Several limitations must be considered in interpreting our results. Our convenience sample is of modest size and consists primarily of white non-Hispanic women. We used a validated zygosity questionnaire instead of genetic testing to confirm that participants were identical twins. Our distribution of broad region 23andMe results into common ethnicity categories likely increased concordance (Appendix Table 5). Thus, the results that involve 23andMe may overestimate test consistency.

By providing detailed numerical results accompanied by color-coded maps showing where descendants came from, testing companies create the impression of rigor and precision. Our results raise questions about the consistency of both ancestry and trait genetic testing. Consumers should not be misled about the usefulness of these services.

Acknowledgements

National Institutes of Health grants U54MD002265 and K23DK101492 and Duncan Neuhauser Endowed Chair. The authors have no conflicts to disclose.

Appendix

Appendix Table 3.

Common ethnicity categories used to compare results across companies. Only ethnicities that appeared on at least one participant's analysis are listed. 23andMe also reported some results as broad categories, e.g. Broadly European, that are not listed below.

Common Ethnicity Category	Company Reported Ethnicity		
	Ancestry	MyHeritage	23andMe
• Britain and Ireland	• Ireland and Scotland	• Irish, Scottish, and Welsh	• British and Irish
	• England and Wales	• English	

Common Ethnicity Category	Company Reported Ethnicity		
	Ancestry	MyHeritage	23andMe
• Northwestern Europe	• Germanic Europe • France	• North and West European	• French and German
• Scandinavian and Finnish	• Sweden	• Scandinavian	• Scandinavian
• European/Ashkenazi Jewish	• European Jewish	• Ashkenazi Jewish	• Ashkenazi Jewish
• Italian	• Italy	• Italian	• Italian
• Greece, Balkans, Baltic, and East European	• Greece and the Balkans • Baltic states • East European and Russia	• Greek • Balkan • Baltic • East European	• Balkan • Eastern European
• Iberian		• Iberian	• Iberian
• Sardinian		• Sardinian	• Sardinian
• West Asia	• Turkey and the Caucasus	• West Asian	• Western Asian
• North Africa and Middle East	• Middle East	• North African • Middle East	• North African and Arabian • Northern East Africa
• West Africa		• Nigerian	• Nigerian • Coastal West African • Senegambian and Guinean
• East Asian and American		• Central American	• Native American • East Asian and Native American
• South Asian			• South Asian

Appendix Table 4.

Example of ancestry concordance calculation when participant tested by Ancestry and MyHeritage.

A. Company Reported Ethnicity (Ancestry)		B. Common Ethnicity Category		C. Company Reported Ethnicity (MyHeritage)		D. Common Ethnicity Category		E. Agreement (B vs. D)
• England and Wales	79.0%	• Britain and Ireland	97.0%	English	27.6%	• Britain and Ireland	52.4%	52.4%
• Ireland and Scotland	18.0%			• Irish, Scottish, and Welsh	24.8%			

A. Company Reported Ethnicity (Ancestry)		B. Common Ethnicity Category		C. Company Reported Ethnicity (MyHeritage)		D. Common Ethnicity Category		E. Agreement (B vs. D)
• Sweden	1.0%	• Scandinavian and Finnish	1.0%	• Scandinavian	19.4%	• Scandinavian and Finnish	19.4%	1.0%
				• East European	19.8%	• Greece, Balkans, Baltic, East European	19.8%	0.0%
				• Italian	8.4%	• Italian	8.4%	0.0%
• Germanic Europe	2.0%	• Northwestern Europe	2.0%					0.0%
Total agreement 53.4%								

Appendix Table 5.

Example of ancestry concordance calculation when participant tested by MyHeritage and 23andMe. To be conservative in our concordance calculations, 23andMe results reported as broad categories were redistributed to maximize concordance according to maps and tables provided by 23andMe as follows: Broadly Southern European (Greece, Balkans, Baltic, East European; Iberian; Italian; Sardinian), Broadly Northwestern Europe (Britain and Ireland, Northwestern Europe, Scandinavian and Finnish), Broadly European (all European regions), Broadly West Asian and North African (West Asia, North Africa and Middle East), and Broadly Unassigned (all regions). In the example below, the reported result of 17.1% Broadly European was redistributed as 5.9% Italian and 11.2% Northwestern Europe before calculating concordance.

A. Company Reported Ethnicity (MyHeritage)		B. Common Ethnicity Category		C. Company Reported Ethnicity (23andMe)		D. Common Ethnicity Category		step 2	E. Agreement (B vs. D step 2)
• East European	23.8%	• Greece, Balkans, Baltic, East European	35.9%	• Eastern European	21.9%	• Greece, Balkans, Baltic, East European	26.3%	$26.3 + 6.3 = 32.6\%$	32.6%
• Balkan	12.1%			• Balkan	4.4%				
• Italian	24.1%	• Italian	24.1%	• Italian	18.2%	• Italian	18.2%	$18.2 + 5.9 = 24.1\%$	24.1%
• North and West European	40.0%	• Northwestern Europe	40.0%	• French and German	11.5%	• Northwestern Europe	11.5%	$11.5 + 14.0 + 11.2 = 36.7\%$	36.7%
				• British and Irish	6.6%	• Britain and Ireland	6.6%	6.6%	0.0%
				• Broadly Northwestern European	14.0%	• Northwestern Europe	14.0%		
				• Broadly Southern European	6.3%	• Greece, Balkans, Baltic, East European	6.3%		
				• Broadly European	17.1%	• Italian • Northwestern Europe	5.9% 11.2%		
Total agreement 93.4%									

References

- 1). Regalado A 2017 was the year consumer DNA testing blew up. MIT Technology Review <https://www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up/> 2 12, 2018.
- 2). Bloss CS, Darst BF, Topol EJ, Schork NJ. Direct-to-consumer personalized genomic testing. Hum Mol Genet 2011 15;20(R2):R132–41. [PubMed: 21828075]
- 3). Ooki S, Yamada K, Asaka A, Hayakawa K. Zygosity diagnosis of twins by questionnaire. Acta Genet Med Gemellol 1990;39:109–15. [PubMed: 2392887]
- 4). Browner WS, Newman TB, Hulley SB. Estimating Sample Size and Power: Applications and Examples in Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB Designing Clinical Research Philadelphia: Lippincott Williams and Wilkins; 2007, 65–96.
- 5). Shriver MD and Kittles RA. Genetic ancestry and the search for personalized genetic histories. Nat Rev Genet 2004;5:611–18. [PubMed: 15266343]
- 6). Novembre J, Johnson T, Bryc K, et al. Genes mirror geography within Europe. Nature 2008;456(7218):98–101. [PubMed: 18758442]
- 7). Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science 2008 22;319(5866):1100–4. [PubMed: 18292342]
- 8). Auton A, Brooks LD, Durbin RM, et al.; 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526(7571):68–74. [PubMed: 26432245]
- 9). Birney E and Soranzo N. Human genomics: The end of the start for population sequencing. Nature 2015 1;526(7571):52–3. [PubMed: 26432243]
- 10). Durand EY, Do CB, Mountain JL, Macpherson JM. Ancestry Composition: A novel, efficient pipeline for ancestry deconvolution. 23andMe White Paper <https://permalinks.23andme.com/pdf/23-16ancestrycomposition.pdf> 7 1, 2018.
- 11). Noto K, Wang Y, Song S, et al. Ethnicity estimate 2018 white paper. Ancestry DNA https://www.ancestrycdn.com/dna/static/pdf/whitepapers/WhitePaper_2018_1130update.pdf 2 15, 2019.
- 12). Spencer EG and Topol EJ. Direct to consumer fitness DNA testing. Clin Chem 2019;65(1):45–47. [PubMed: 30274978]
- 13). Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 2011;187(2):367–83 [PubMed: 21115973]
- 14). Argo C and Denne L. Twins get some ‘mystifying’ results when they buy 5 DNA ancestry kits to the test <https://www.cbc.ca/news/technology/dna-ancestry-kits-twins-marketplace-1.4980976> 2 15, 2019.
- 15). Bollinger JM, Green RC, Kaufman D. Attitudes about regulation among direct-to-consumer genetic testing customers. Genet Test Mol Biomarkers 2013;17(5):424–8. [PubMed: 23560882]
- 16). Marietta C and McGuire AL. Currents in contemporary ethics. Direct-to-consumer genetic testing: is it the practice of medicine? J Law Med Ethics 2009 37(2):369–74. [PubMed: 19493081]
- 17). Brothers KB and Knapp EE. How should primary care physicians respond to direct-to-consumer genetic testing results? AMA J Ethics 2018;20(9):E812–18. [PubMed: 30242811]

Table 1.

Concordance of ancestry among identical twins tested by three companies.

Comparison	Company	Subjects	Percent agreement (95%CI) [range]
<i>Twins tested by same company</i>			
Two twins	23andMe	20 twin pairs	94.5% (93.2–95.9%) [89.0–98.9%]
Two twins	MyHeritage	21 twin pairs	98.7% (98.0–99.4%) [93.0–100.0%]
Two twins	Ancestry	21 twin pairs	99.2% (98.6–99.4%) [96.0–100.0%]
<i>Participants tested by different companies</i>			
Two companies	23andMe, MyHeritage	41 participants	66.9% (62.8–71.0%) [40.9–90.5%]
Two companies	23andMe, Ancestry	41 participants	84.1% (81.0–87.3%) [63.7–98.9%]
Two companies	MyHeritage, Ancestry	42 participants	52.7% (46.6–58.8%) [7.0–100.0%]

Table 2.

Concordance of self (“Self”)- and 23andMe-reported genetic analyses (“Genetic”) among 41 participants.

Trait	Self yes Genetic yes	Self yes Genetic no	Self no Genetic yes	Self no Genetic no	Concordance n/N (%) [95% CI]
Deep sleep	0	21	6	14	14/41 (34.1%) [21.6–49.5%]
Detached earlobes	14	0	27	0	14/41 (34.1%) [21.6–49.5%]
Sweet taste	6	22	4	9	15/41 (36.6%) [23.6–51.9%]
Cheek dimples	0	15	0	26	26/41 (63.4%) [48.1–76.4%]
Freckles	15	8	4	14	29/41 (70.7%) [55.5–82.4%]
Brown or hazel eyes	8	2	5	25	33/41 (80.4%) [66.0–89.8%]
Lactose intolerance	0	6	0	35	35/41 (85.3%) [71.6–93.1%]
Cleft chin	0	4	0	37	37/41 (90.2%) [77.5–96.1%]