

# Identification of individuals by trait prediction using whole-genome sequencing data

September 6 2017

---



Examples of real (Left) and predicted (Right) faces from the Human Longevity study predicting face and other physical traits from whole genome sequencing data. Credit: Human Longevity, Inc.

Researchers from Human Longevity, Inc. (HLI) have published a study in which individual faces and other physical traits were predicted using whole genome sequencing data and machine learning. This work, from lead author Christoph Lippert, Ph.D. and senior author J. Craig Venter, Ph.D., was published in the journal *Proceedings of the National Academy of Sciences (PNAS)*.

The authors believe that, while the study offers novel approaches for forensics, the work has serious implications for [data privacy](#), deidentification and adequately informed consent. The team concludes

that much more public deliberation is needed as more and more genomes are generated and placed in public databases.

For the IRB approved study, 1,061 ethnically diverse people ranging in age from 18 to 82 participated by having their genomes sequenced to an average depth of at least 30x. Researchers also collected phenotype data in the form of 3-D facial images, voice samples, eye and skin color, age, height, and weight.

The team predicted eye color, [skin color](#) and sex with high accuracy, but other more complex genetic traits proved more difficult. The team believes their [predictive models](#) are sound, but that large cohorts are needed to make prediction more robust. The team also developed a machine learning algorithm called a maximum entropy algorithm, which had novelty in that it found an optimal combination of all predictive models to match whole-genome sequencing data with phenotypic and demographic data and enabled the correct identification of, on average, 8 out of 10 participants of diverse ethnicity, and 5 out of 10 African American or European participants.

Venter, HLI's co-founder, executive chairman and head of scientific strategy, commented, "We set out to do this study to prove that your [genome](#) codes for everything that makes you, you. This is clearly a proof of concept with a limited cohort but we believe that as we increase the numbers of people in this study and in the HLI database to hundreds of thousands we will be able to accurately predict all that can be predicted from individuals' genomes."

He added, "We are also concerned that the public and the research community at large are not adequately focused on the need for better safeguards and policies for individual privacy in the genomics era and are urging more analysis, better technical solutions, and continued discussion."

Lippert, data scientist at HLI, added, "This study shows the potential of imaging technologies to screen the traits of large numbers of individuals. Machine learning enables fully automated data interpretation and plays a crucial role in scientific discovery."

**More information:** Christoph Lippert et al. Identification of individuals by trait prediction using whole-genome sequencing data, *Proceedings of the National Academy of Sciences* (2017). [DOI: 10.1073/pnas.1711125114](https://doi.org/10.1073/pnas.1711125114)

Provided by Human Longevity, Inc.

Citation: Identification of individuals by trait prediction using whole-genome sequencing data (2017, September 6) retrieved 24 April 2024 from <https://phys.org/news/2017-09-identification-individuals-trait-whole-genome-sequencing.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.