# How I analyzed my DNA with Claude Code and Claude Science

*Two experiments with the same 23andMe raw-data file - first with Claude Code, then with Claude Science*

**Pavel Gurov** · ORCID 0009-0001-0152-8056
Written 4 July 2026 · 940 words · gurovdigital.com/research/dna-analysis-claude-code-claude-science

Privacy · AI agents · Language models · Anthropic · Claude

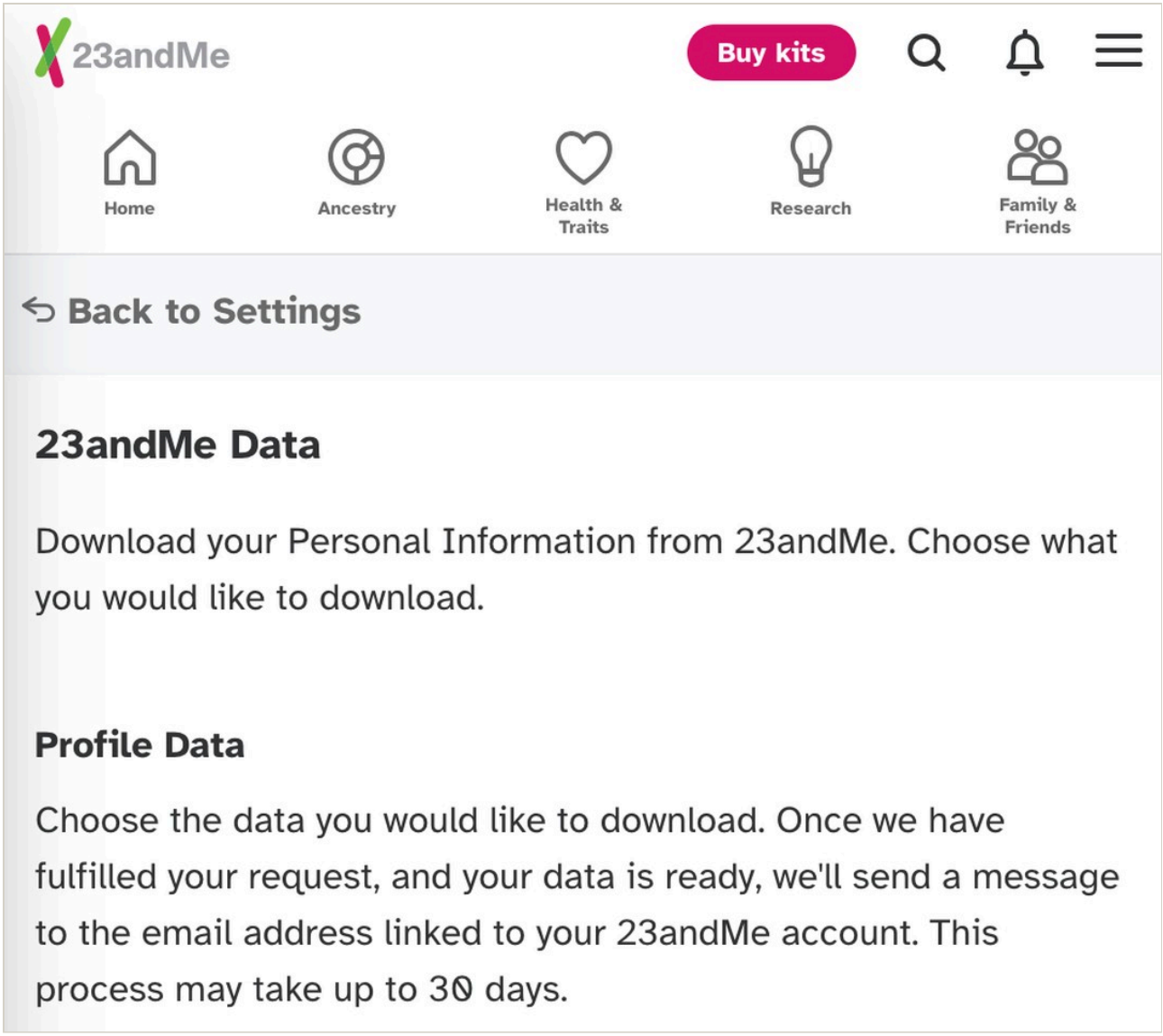
## Claude Code

How I analyzed my DNA using Claude Code.

If you try to replicate it, you do so at your own risk.

1. First, your genes need to be digitized. I used 23andMe for that. Spit in a tube, sent it off by DHL courier. After analysis, the service mostly gave me entertaining lines like "You're likely to prefer vanilla over chocolate ice cream" - nothing meaningful about health (my guess is their lawyers won't allow it).

2. I went into my 23andMe account settings and downloaded the Raw Data.




> The download screen in the 23andMe account settings, where the raw genotype file comes from 23andMe. 23andme.com. Screenshot, quoted under §51 UrhG.

Technically, it was a .txt file weighing 16.6 MB. This isn't my full DNA. Full genome sequencing involves about 3 billion nucleotide pairs - that file would be around 100 GB.

But all humans are 99.9% identical. Companies like 23andMe don't read the entire genome - they record only specific variations, SNPs (single nucleotide polymorphisms). So those 16 MB aren't my full DNA, but rather an "encyclopedia of Pavel Gurov's mutations" - the variations that define my uniqueness.

DEFINITION

**Single nucleotide polymorphism (SNP)** — a single position in the genome where the letter differs between people. A consumer genotyping array reads several hundred thousand of these chosen positions rather than the whole sequence, which is why the download is measured in megabytes and not in gigabytes.

---

The term dates to the late 1990s mapping effort - The International SNP Map Working Group, *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms*, Nature 409, 928-933, 2001[1]

3. The next issue: the text file was about 650,000 lines of data, or roughly 4.5 million tokens. A standard ChatGPT will choke on that (or worse, hallucinate results). What do you do instead? You hand the task to an AI agent that can break it into smaller chunks and actually process this huge dataset.

I used the Claude Code agent, which is currently one of the strongest options on the market in terms of security. And yes, Claude Code can work not just with software code, but with genetic code as well.

4. I created an isolated folder on my computer, uploaded the DNA file there, granted the agent access, and gave it the big fat prompt:

EXACT PROMPT

---

```
Act as an expert bioinformatics AI agent. Using Python and the pandas library, write and
execute a local script to parse my 23andMe raw DNA file (txt/tsv format) located in this
current directory. Extract SNPs related to key, well-documented health markers: pharma-
cogenomics (e.g., warfarin sensitivity), methylation/folate metabolism (MTHFR),
Alzheimer's risk (APOE), and severe monogenic traits (e.g., celiac disease, hemochromato-
sis). Cross-reference the extracted rsIDs with established clinical databases like ClinVar
or SNPedia. CRITICAL: Account for strand orientation (plus/minus strand flip) to avoid
misreporting alleles. 23andMe sometimes reports on the negative strand. Do NOT output the
raw Python code to me unless I explicitly ask. Do NOT output the entire file. Output a
clean, structured Markdown table with: the medical condition, the specific gene and rsID,
my exact genotype, and the clinical significance based on peer-reviewed scientific
literature.
```

Instead of uploading 650,000 lines to the cloud, the AI agent autonomously wrote a Python script and executed it locally, using the Apple Silicon M2 chip.

In other words, my DNA file never left my hard drive (security!)

What did the AI actually find in my biological data? It produced a massive dossier with insights no commercial service would have given me:

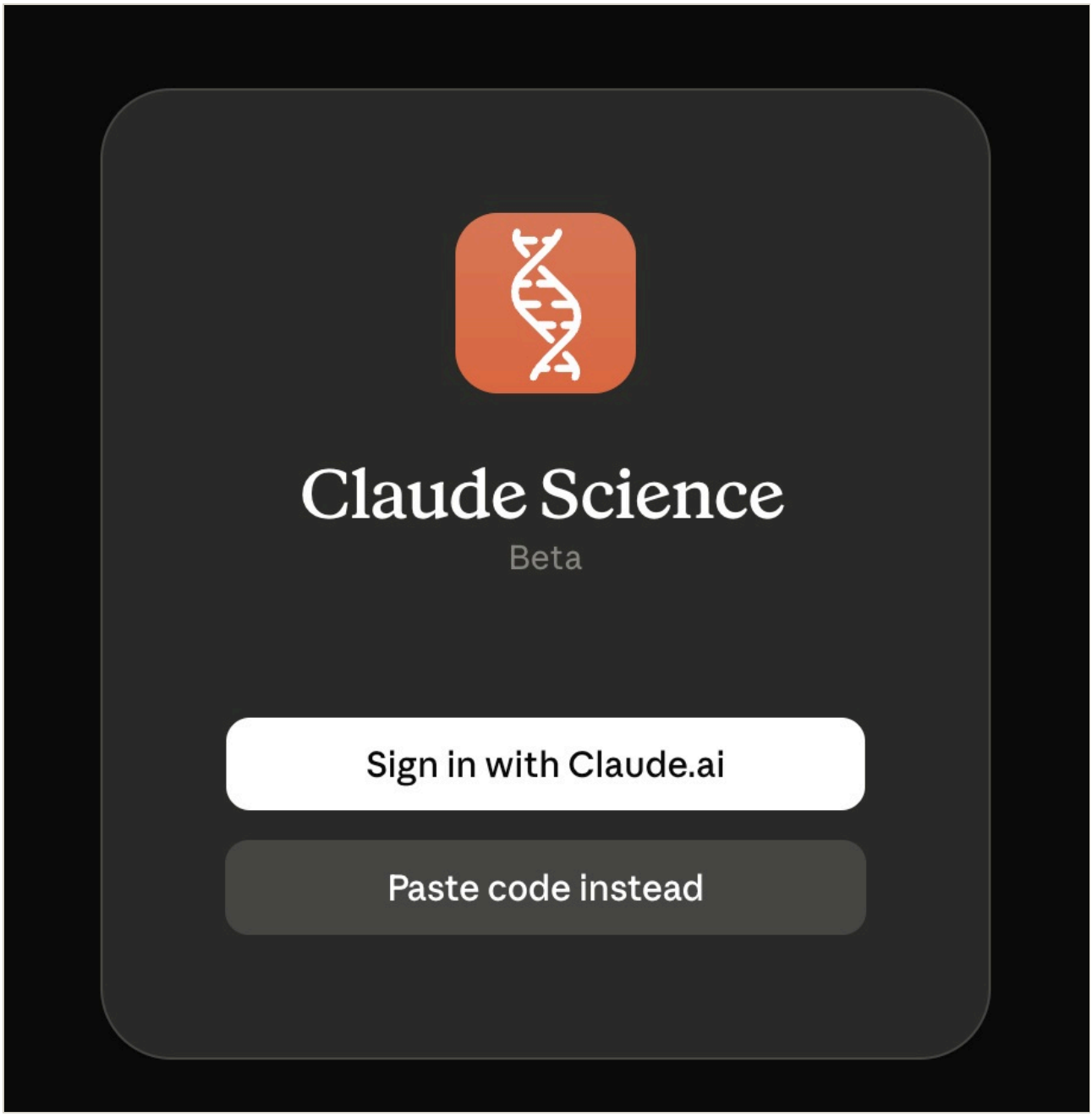
- my sensitivity to anticoagulant drugs
- my susceptibility to some viruses

- likelihood of age-related degeneration of eyesight
- folate metabolism (relevant to cancer risk)
- Alzheimer's risk, and much more.

This approach minimizes hallucinations because the model isn't guessing - it's grounded in deterministic script execution and concrete databases (Retrieval-Augmented Generation combined with Agentic Code Execution).

Give it a try if you happen to have your genome data lying around! You won't get answers to everything, but it's a genuinely powerful capability - especially at roughly $20, which is what Claude Pro costs to access the Claude Code agent.

## Claude Science




> The Claude Science sign-in screen, the desktop app in its beta release Anthropic. anthropic.com. Screenshot, quoted under §51 UrhG.

Anthropic just shipped a new desktop app, and I've been using it non-stop: Claude Science. It's not a new model. It's a full research workbench that runs on AI agents. You ask a question in plain English, and the agents go query the actual scientific sources for you - papers, gene and protein databases, molecular structures - then run the analysis and draw the result as a chart or a 3D model.

More than 60 databases are wired in, so you don't have to jump between PubMed, a terminal, and five separate viewers.

What I did with it: I fed it my own raw DNA file. Claude Science read 631,454 nucleotide polymorphisms - the single-letter spots in DNA that vary from person to person - and checked each one against a large public database of genetic studies (the GWAS Catalog).

A few minutes later it handed me a clean chart showing where I carry more or fewer risk-linked variants than the population average, across things like blood pressure, cholesterol, type 2 diabetes, coronary artery disease, and body mass index.


> The chart Claude Science drew - effect alleles carried per trait, against the population-average tally Anthropic. anthropic.com. Screenshot, quoted under §51 UrhG.

To be clear, this is genetic predisposition, not a diagnosis. But seeing my own genome mapped against decades of published research, in one plain-language session, is not something I could easily do myself a year ago.

## Three limitations

- The 23andMe download is a genotype file covering selected markers, not a full genome or a clinical test.[2]

- Claude Code ran the Python script locally, but Claude itself was not offline: prompts and context were processed through Anthropic's application programming interface.[3]
- The Claude Science chart was exploratory. Genome-wide association matches are not a validated polygenic risk score or a diagnosis.[4][5][6]

## REFERENCES


1The International SNP Map Working Group, *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms*, Nature 409, 928-933, 2001 - https://doi.org/10.1038/35057149↩

223andMe, *Accessing Your Raw Genetic Data* - https://customercare.23andme.com/hc/en-us/articles/212196868-Accessing-Your-Raw-Genetic-Data↩

3Anthropic, *Claude Code data usage* - https://code.claude.com/docs/en/data-usage↩

4Anthropic, *Claude Science, an AI workbench for scientists, is now available*, 30 June 2026 - https://www.anthropic.com/news/claude-science-ai-workbench↩

5National Human Genome Research Institute and European Bioinformatics Institute, *Genome-Wide Association Studies Catalog* - https://www.ebi.ac.uk/gwas/↩

6National Human Genome Research Institute, *Polygenic risk scores* - https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores↩

7Gurov, P., *How I analyzed my DNA using Claude Code*, LinkedIn, 10 April 2026 - https://www.linkedin.com/posts/hellogurov_how-i-analyzed-my-dna-using-claude-code-activity-7448393556651692032-EVxS

8Gurov, P., *Anthropic just shipped a new desktop app: Claude Science*, LinkedIn, 4 July 2026 - https://www.linkedin.com/posts/hellogurov_anthropic-just-shipped-a-new-desktop-app-activity-7479047601934360577-ptd5



CITE THIS ARTICLE

Gurov, P. (2026). *How I analyzed my DNA with Claude Code and Claude Science*. Gurov Digital Research. https://gurovdigital.com/research/dna-analysis-claude-code-claude-science

Published at https://gurovdigital.com/research/dna-analysis-claude-code-claude-science · Author ORCID 0009-0001-0152-8056