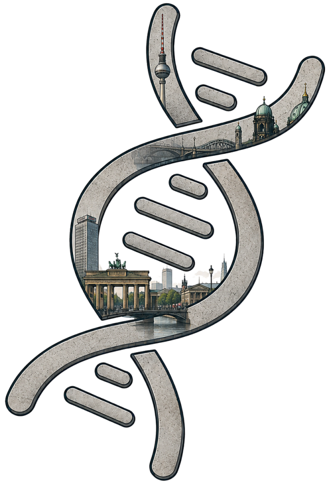
LLMs for personal genomics?

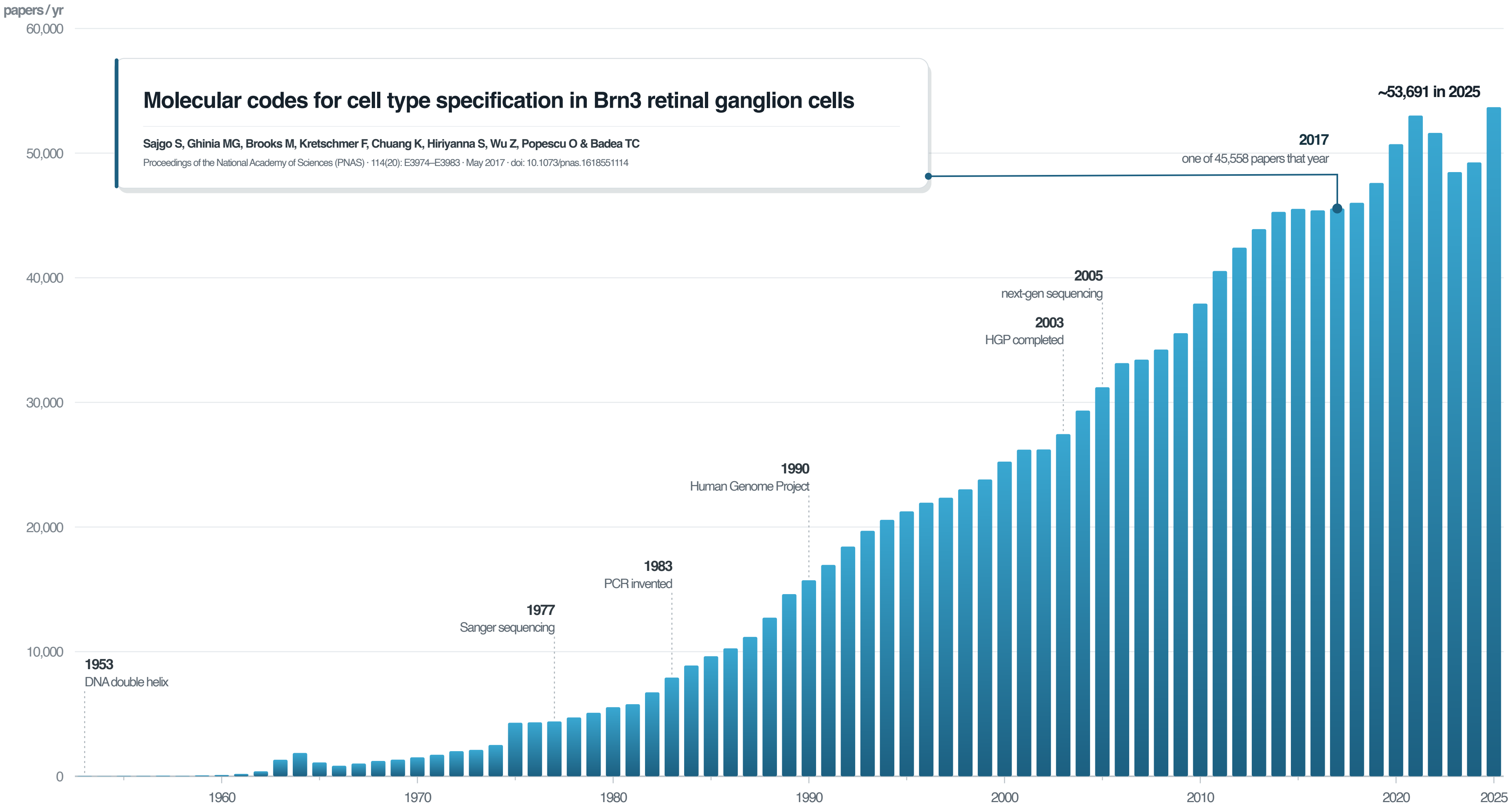
Bridging promise and reality

16 June 2026
Futura Camp, Berlin

Péter Szilágyi
Dark Bio Founder



Exploding breadth and depth



Molecular codes for cell type specification in Brn3 retinal ganglion cells

Sajgo S, Ghinia MG, Brooks M, Kretschmer F, Chuang K, Hiriyanna S, Wu Z, Popescu O & Badea TC

Proceedings of the National Academy of Sciences (PNAS) · 114(20): E3974–E3983 · May 2017 · doi: 10.1073/pnas.1618551114

Source: PubMed / MEDLINE via NCBI E-utilities, queried 2026-06-10 · papers per year with “DNA” in title or abstract.

The promise

"Hey Claude, write a genetic analysis for alcohol tolerance!" **~Me, ETHPrague hackathon**

Tolerance:

- **rs671** – how fast you metabolize acetaldehyde (alcohol gunk)
- **rs1229984** – how fast you turn alcohol into acetaldehyde (Asia)
- **rs2066702** – how fast you turn alcohol into acetaldehyde (Africa)

Pleasure:

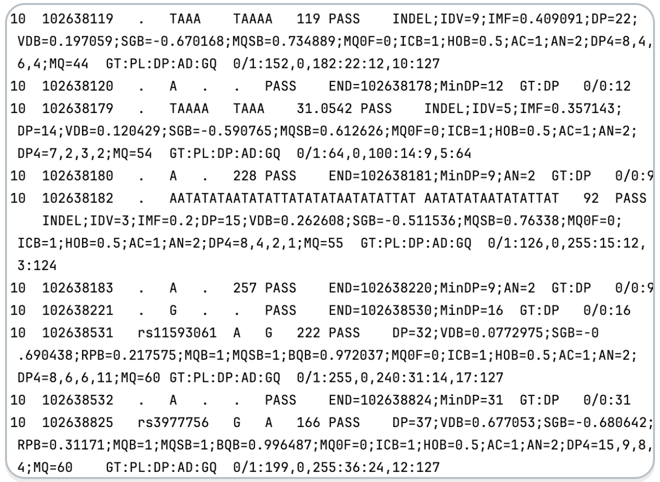
- **rs1799971** – higher endorphin sensitivity
- **rs1800497** – lower dopamine sensitivity
- **rs279858** – elevated calming effect

The reality

- Bug #1 – Mixed up the two DNA strands (dataset vs. literature)
- Bug #2 – One marker's position was off by 25K bases
- Bug #3 – Interpreted the mutation as the issue, not what it was
- Bug #4 – Missing data interpreted as all is good
- Bug #5 – Uncertain data interpreted as all is good
- Bug #6 – Multiple-data points collapsed into first
- Bug #7 – Loaded multi-letter sequence, computed on first

I only caught these because results "looked" weird...

Problem: Complex inputs, no feedback loop

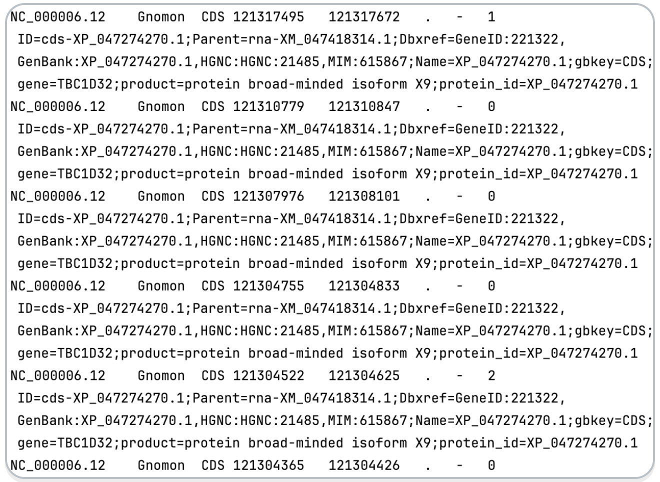


```
10  102638119   .   TAAA    TAAAA   119 PASS    INDEL;IDV=9;IMF=0.409091;DP=22;
 VDB=0.197059;SGB=-0.670168;MQSB=0.734889;MQ0F=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=8,4,
 6,4;MQ=44  GT:PL:DP:AD:GQ  0/1:152,0,182:22:12,10:127
10  102638120   .   A   .   .   PASS    END=102638178;MinDP=12  GT:DP   0/0:12
10  102638179   .   TAAAA   TAAA    31.0542 PASS    INDEL;IDV=5;IMF=0.357143;
 DP=14;VDB=0.120429;SGB=-0.590765;MQSB=0.612626;MQ0F=0;ICB=1;HOB=0.5;AC=1;AN=2;
 DP4=7,2,3,2;MQ=54  GT:PL:DP:AD:GQ  0/1:64,0,100:14:9,5:64
10  102638180   .   A   .   228 PASS    END=102638181;MinDP=9;AN=2  GT:DP   0/0:9
10  102638182   .   AATATATAATATATTATATATAATATATTAT AATATATAATATATTAT   92  PASS
    INDEL;IDV=3;IMF=0.2;DP=15;VDB=0.262608;SGB=-0.511536;MQSB=0.76338;MQ0F=0;
 ICB=1;HOB=0.5;AC=1;AN=2;DP4=8,4,2,1;MQ=55  GT:PL:DP:AD:GQ  0/1:126,0,255:15:12,
 3:124
10  102638183   .   A   .   257 PASS    END=102638220;MinDP=9;AN=2  GT:DP   0/0:9
10  102638221   .   G   .   .   PASS    END=102638530;MinDP=16  GT:DP   0/0:16
10  102638531   rs11593061  A   G   222 PASS    DP=32;VDB=0.0772975;SGB=-0
 .690438;RPB=0.217575;MQB=1;MQSB=1;BQB=0.972037;MQ0F=0;ICB=1;HOB=0.5;AC=1;AN=2;
 DP4=8,6,6,11;MQ=60 GT:PL:DP:AD:GQ  0/1:255,0,240:31:14,17:127
10  102638532   .   A   .   .   PASS    END=102638824;MinDP=31  GT:DP   0/0:31
10  102638825   rs3977756   G   A   166 PASS    DP=37;VDB=0.677053;SGB=-0.680642;
 RPB=0.31171;MQB=1;MQSB=1;BQB=0.996487;MQ0F=0;ICB=1;HOB=0.5;AC=1;AN=2;DP4=15,9,8,
 4;MQ=60    GT:PL:DP:AD:GQ  0/1:199,0,255:36:24,12:127
```

Your genome variants · **1.45GB**

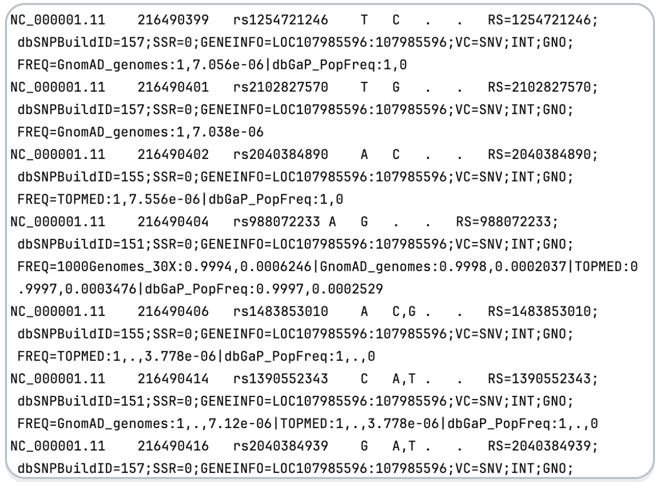
```
GGGAGGTGGCTTCTCTATGGAGGATGGGTTGGAATAAGTCGCAAGCAAGAAGGGTGTGAGCAGACACACGGGAAGAAGCT
ATTGCAGTGCTCCAGTGAGAGGGGCTGCAGCAGGAACTGAGGCAGCAGTGGAGCTGGAGAGAAGTAATCCCTGTGGGACA
TGCTTGGAGATGGAGGCACAGGACCTGAGGGTGCATTGAGTGTGGggatgagaaaaagagagaagcaagGATTCTTTTT
GACTTGGGCAACTGGGTGGATGGGACCTGAGGATGCATTGAGTGTGGggatgagaaaaagagagaagcaagGATTCTTT
TTGACTTGGGAAACTGGGTGGATGGGACCTGAGGATGCATTGAGTGTGGggatgagaaaaagagagaagcaagGATTCT
TTTTGACTTGGGCAACTGGGTGGATGGTGGTTGGCTGGAGGTGAAGTACTCATGGGGGTGGCAGGAGTGAGAAGGAAATC
AAGAGCTCCTTTTTGCATGTGTTAGGTTTGAGATGCCCATCAGACATCCAGGTAGGTGCGCTGAGTCAGACTTGGATACG
TGCGTCCAGAATTTCAGGAAGAGCCAGACCAGGAGACTTCCACTTGGGAGTCATCAACCTATAGGTGGTATTTAAAGCCA
GAGGATGAGGTCACTTAGGGATGGTGTGTAGctagaagaaagaaatgggcCAAAGATGGGGCTCTAGGGCACTACACTGC
AACATTCGGAGGTCAAGCAGGGGAGAAGGGCAGCAAAGAGGTAGAGAAGATGCAGTCGGTTGAAGGGCATTCAGGAGAGA
GGAGGACCATGAAGCCCAGGGAGGGGGGCTTGCTCAACTGTAGGGAATGTGGAGGGATCTGGTCCGGTAAGAACAGAGAA
GTGACCACCATTGACTTTGGCCACATGGAGGTTACCAGCGCTCTTAACAAGCACAGGCTCAATGTGGGGCTGGGAGAGAA
ATCCAGTTAGAATAGGTTCAGGAAAGACTGTGAGGCAAGGACCAGAAATGATGgaatagaaaaatggaacagTAGCTGTA
AAGGAAGGTGGCATTAAGggaagaagtttttttaaaaaaaaaataggagatatGAAAACGTGTATGTGATTCACTGGTGT
AGTAGAGGGACAGAGAAAGATTGCAGGTGCAGGAAAAAGTAAATTCAGCAGCAAAGACCTTGAGAAGGTAAGGTGTGTGT
GTCAGGGGTGGCTTTTGCCCTGCCCAGTGGAGGGGTGGTCTTTGCCTTGCACAGGTGGAGGGGTGGTCTTTGCCTTGCAC
AGGTGGTGGGGTGGTCTTTGCCTTGCACAGGTGGAGAGGCGGCCTTTGCCTTGGGGCAGGGACAGTTCAGCCATAGAAGT
GGGGGAAAGTGGGGAGTGTGGGTCCAGATGGAGGTGCCAGTGGTGGGAAGCCCTTGTTTTGCTGCTTTCATTTGCTTAGT
GAAGAATGAGACAGTTGCATGCCAAAAAGAGAGAAGGTGAAGGGAGGTGAGTCCTGGaatgagagggaaggagggagggc
agcAGATAGGCTGGGGGAGTGAGGATTGCTGCGATGTGTCTGATGCCTGTAGAGATGTGTGACCATAGAGGCTCAGTGAG
TCCCATCCTTGCTGTTGTGTGATCTTTCTCTTACCACTCTCAGGAGCTCGAGTGCGATTGGAGTAGAAGGTTGCCGGGAT
```

Reference genome · **3.34GB**



```
NC_000006.12    Gnomon  CDS 121317495   121317672   .   -   1
 ID=cds-XP_047274270.1;Parent=rna-XM_047418314.1;Dbxref=GeneID:221322,
 GenBank:XP_047274270.1,HGNC:HGNC:21485,MIM:615867;Name=XP_047274270.1;gbkey=CDS;
 gene=TBC1D32;product=protein broad-minded isoform X9;protein_id=XP_047274270.1
NC_000006.12    Gnomon  CDS 121310779   121310847   .   -   0
 ID=cds-XP_047274270.1;Parent=rna-XM_047418314.1;Dbxref=GeneID:221322,
 GenBank:XP_047274270.1,HGNC:HGNC:21485,MIM:615867;Name=XP_047274270.1;gbkey=CDS;
 gene=TBC1D32;product=protein broad-minded isoform X9;protein_id=XP_047274270.1
NC_000006.12    Gnomon  CDS 121307976   121308101   .   -   0
 ID=cds-XP_047274270.1;Parent=rna-XM_047418314.1;Dbxref=GeneID:221322,
 GenBank:XP_047274270.1,HGNC:HGNC:21485,MIM:615867;Name=XP_047274270.1;gbkey=CDS;
 gene=TBC1D32;product=protein broad-minded isoform X9;protein_id=XP_047274270.1
NC_000006.12    Gnomon  CDS 121304755   121304833   .   -   0
 ID=cds-XP_047274270.1;Parent=rna-XM_047418314.1;Dbxref=GeneID:221322,
 GenBank:XP_047274270.1,HGNC:HGNC:21485,MIM:615867;Name=XP_047274270.1;gbkey=CDS;
 gene=TBC1D32;product=protein broad-minded isoform X9;protein_id=XP_047274270.1
NC_000006.12    Gnomon  CDS 121304522   121304625   .   -   2
 ID=cds-XP_047274270.1;Parent=rna-XM_047418314.1;Dbxref=GeneID:221322,
 GenBank:XP_047274270.1,HGNC:HGNC:21485,MIM:615867;Name=XP_047274270.1;gbkey=CDS;
 gene=TBC1D32;product=protein broad-minded isoform X9;protein_id=XP_047274270.1
NC_000006.12    Gnomon  CDS 121304365   121304426   .   -   0
```

Gene annotations · **1.6GB**



```
NC_000001.11    216490399   rs1254721246    T   C   .   .   RS=1254721246;
 dbSNPBuildID=157;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
 FREQ=GnomAD_genomes:1,7.056e-06|dbGaP_PopFreq:1,0
NC_000001.11    216490401   rs2102827570    T   G   .   .   RS=2102827570;
 dbSNPBuildID=157;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
 FREQ=GnomAD_genomes:1,7.038e-06
NC_000001.11    216490402   rs2040384890    A   C   .   .   RS=2040384890;
 dbSNPBuildID=155;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
 FREQ=TOPMED:1,7.556e-06|dbGaP_PopFreq:1,0
NC_000001.11    216490404   rs988072233 A   G   .   .   RS=988072233;
 dbSNPBuildID=151;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
 FREQ=1000Genomes_30X:0.9994,0.0006246|GnomAD_genomes:0.9998,0.0002037|TOPMED:0
 .9997,0.0003476|dbGaP_PopFreq:0.9997,0.0002529
NC_000001.11    216490406   rs1483853010    A   C,G .   .   RS=1483853010;
 dbSNPBuildID=155;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
 FREQ=TOPMED:1,.,3.778e-06|dbGaP_PopFreq:1,.,0
NC_000001.11    216490414   rs1390552343    C   A,T .   .   RS=1390552343;
 dbSNPBuildID=151;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
 FREQ=GnomAD_genomes:1,.,7.12e-06|TOPMED:1,.,3.778e-06|dbGaP_PopFreq:1,.,0
NC_000001.11    216490416   rs2040384939    G   A,T .   .   RS=2040384939;
 dbSNPBuildID=157;SSR=0;GENEINFO=LOC107985596:107985596;VC=SNV;INT;GNO;
```

Known variant database · **213.78GB**

Solution: Deterministic foundations, expert feedback loops

LLMs are a *last-mile* technology

- Good at combining horizontals, bad at cross-cutting verticals
- Any layer can be built, if everything below is deterministic

LLMs are hallucinating slop-machines

- Extremely capable, but always slightly inaccurate
- Error correction needs *some* feedback loop

Thank you

Péter Szilágyi
peter@dark.bio
@peter_szilagyi

Dark Bio Founder
https://dark.bio/whitepaper