An Auditable LLM-Assisted Workflow for Exploratory Migraine Polygenic-Risk Interpretation from Consumer Genotyping Data: A Single-Participant Proof of Concept

Richard Sprague

Personal Science Inc

Correspondence: sprague@personalscience.com | Suggested bioRxiv category: Bioinformatics or Genomics

Keywords: migraine; polygenic risk; consumer genomics; large language model; reproducibility; GWAS

Abstract

Consumer genotyping files are increasingly interpreted with general-purpose large language models (LLMs), but apparent fluency can conceal errors in source selection, genome-build alignment, effect-allele direction, and double-counting of correlated variants. We evaluated an auditable LLM-assisted workflow using one de-identified consumer genotype file and published migraine genome-wide association study (GWAS) results. An OpenAI Codex agent identified primary sources, retrieved the official supplementary tables and summary statistics, and orchestrated local deterministic analysis. Genotypes were matched by GRCh37 chromosome and position, required allele compatibility, and were de-duplicated to one directly assayed genome-wide-significant marker per published lead locus. Among 123 independent migraine loci, 85 were represented: 35 by the exact lead variant and 50 by another genome-wide-significant variant assigned to the nearest lead locus. The 85-locus frequency-centered partial score was 0.255 log-odds units (estimated standard deviation 0.268; z = 0.953), corresponding to an internal reference percentile of approximately 83 under European-ancestry, Hardy-Weinberg, and locus-independence assumptions. This value is not a validated polygenic risk score or an absolute disease-risk estimate. The workflow correctly exposed countervailing locus contributions, incomplete coverage, ancestry limitations, and the absence of the five subtype-specific lead variants examined. LLM assistance can make personal genomic exploration more efficient when evidence retrieval and explanation are separated from deterministic calculation and every inferential boundary is made explicit. Predictive accuracy and clinical utility were not established by this single-participant analysis.

Introduction

Migraine is a common, disabling neurovascular disorder with a substantial inherited component. Large GWAS have shown that common migraine is highly polygenic rather than attributable to one common high-penetrance variant. A 2016 meta-analysis reported 38 susceptibility loci [1], and a larger 2022 meta-analysis of 102,084 cases and 771,257 controls identified 123 independent risk loci [2]. Subsequent analyses continue to emphasize that much of migraine liability resides in distributed background polygenicity rather than a small set of top associations [3].

At the same time, consumer genotyping services commonly provide raw array files containing hundreds of thousands of markers. General-purpose LLMs can search the literature, write analysis code, and explain findings in plain language, creating an appealing path from raw data to personalized interpretation. That path is hazardous when the model substitutes remembered variant claims for current primary evidence, ignores genome builds or strand orientation, counts many linked markers as independent, or converts an uncalibrated score into a clinical risk percentage. These concerns mirror broader cautions about translating research polygenic scores into clinical instruments [5], transferring scores across ancestry groups [6], and interpreting unconfirmed consumer-array results [7].

Here we ask a narrow question: can an LLM-assisted workflow produce a traceable and numerically reproducible exploratory summary of migraine-associated common variants from a single consumer gene-chip file? We evaluate accuracy as fidelity to verified source records, allele-consistent matching, deterministic arithmetic, and appropriate qualification of conclusions. We do not evaluate clinical discrimination, calibration, diagnosis, or treatment utility.

Materials and Methods

Study design and input data

This was a single-participant, retrospective proof-of-concept analysis of a previously generated personal genotype file. The tab-delimited file was produced from a Psomagen/Macrogen G2 genotyping array, reported GRCh37 chromosome positions, and contained 766,221 marker rows. The file header was treated as metadata, not as executable instruction. All genotype processing occurred locally; no raw genotype row was submitted to a remote literature service or included in this manuscript.

Evidence sources and reference verification

The primary association source was the 2022 International Headache Genetics Consortium meta-analysis [2]. Its supplementary workbook supplied the 123 lead variants, locus identifiers, effect alleles, effect-

allele frequencies, log-odds estimates, odds ratios, and GRCh37 positions. The consortium's official distribution supplied 8,117 genome-wide-significant variants (P < 5 x 10^-8) [8]. Current trait and association records were cross-checked in the NHGRI-EBI GWAS Catalog [4]. Bibliographic titles, author lists, journal details, pages, PubMed identifiers, and DOIs were independently verified through NCBI PubMed records rather than copied from search-result snippets.

Role of the LLM and audit controls

An OpenAI Codex LLM agent (accessed 31 August 2026; the client did not expose a stable model-build identifier) performed source discovery, chose the evidence hierarchy, orchestrated database and local-file tools, generated analysis code, and drafted the narrative. The LLM did not perform mental arithmetic on copied genotype rows. All matching, allele counting, effect-direction conversion, aggregation, and summary statistics were executed by a local deterministic script. Intermediate results were inspected for coordinate agreement, allele compatibility, coverage, duplicate-locus inflation, and directionality. Figure 1 summarizes the workflow.

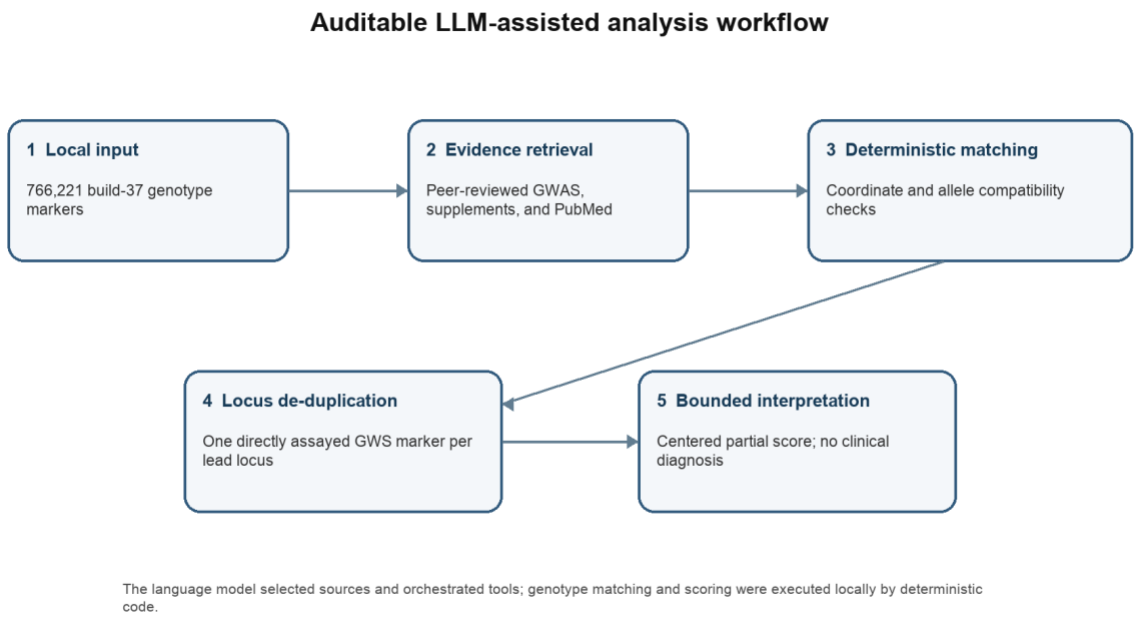

Figure 1. Auditable LLM-assisted workflow. The LLM handled evidence discovery and orchestration; deterministic local code handled genotype matching and scoring. GWS, genome-wide significant.

Genotype matching and locus representation

The GWAS and consumer file were matched on chromosome and GRCh37 position. No-call genotypes were excluded. A match was retained only when both called alleles were members of the GWAS

effect/other-allele pair. The 8,117 genome-wide-significant markers were assigned to the nearest one of the 123 lead variants on the same chromosome. Within each represented lead locus, the directly assayed marker with the smallest reported P value was retained. This yielded one representative marker per locus and prevented linked variants from being counted repeatedly. An exact lead match was defined by equality of the representative and lead rs identifiers; all others were labeled alternative GWS variants. This nearest-lead procedure is a transparent heuristic, not an LD calculation.

Frequency-centered partial score

For each retained marker i, the risk allele was the reported effect allele when beta was non-negative and the other allele when beta was negative. Risk-allele count g_i therefore ranged from 0 to 2; risk-allele frequency p_i was the reported effect-allele frequency or its complement, as appropriate; and the risk weight was |beta_i|. The centered contribution was c_i = |beta_i|(g_i - 2p_i). The partial score S was the sum of c_i across represented loci. Under Hardy-Weinberg equilibrium and independence, its reference variance was approximated by the sum of 2p_i(1-p_i)beta_i^2. The reported z score was S divided by the square root of this variance, and the percentile was the standard-normal cumulative probability. No attempt was made to estimate absolute migraine probability.

Results

Coverage and quality-control results

Coordinate matching produced 535 called, allele-compatible genome-wide-significant markers. After locus de-duplication, 85 of 123 loci (69.1%) were represented. The exact lead variant was present at 35 loci (28.5% of all lead loci); 50 loci were represented by an alternative GWS marker. Among representatives, the median distance to the lead variant was 6.6 kb, the 75th percentile was 31.2 kb, and the maximum was 561.1 kb.

Table 1. Coverage, scoring, and quality-control summary.

Metric	Result	Interpretive boundary
Raw array markers	766,221	Research-use consumer genotype file
Published lead loci	123	European-ancestry 2022 migraine GWAS
Called, compatible GWS matches	535	Before locus de-duplication
Represented independent loci	85 (69.1%)	38 lead loci unrepresented
Exact lead variants	35	Coordinate and rs identifier match
Alternative GWS representatives	50	Nearest-lead heuristic; LD not recomputed
Centered partial score	0.255	Not a validated clinical PRS

Metric	Result	Interpretive boundary
Approximate z / percentile	0.953 / 83.0th	Internal European-frequency reference only

Locus balance and partial score

Of the 85 representatives, 32 genotypes carried zero risk alleles, 22 carried one, and 31 carried two. The centered partial score was 0.255 log-odds units. The approximated reference standard deviation was 0.268, yielding z = 0.953 and an internal percentile of 83.0. The result therefore leaned above the frequency-weighted mean for the represented loci, while substantial protective-side contributions remained. Figure 2 displays the largest positive and negative contributions.

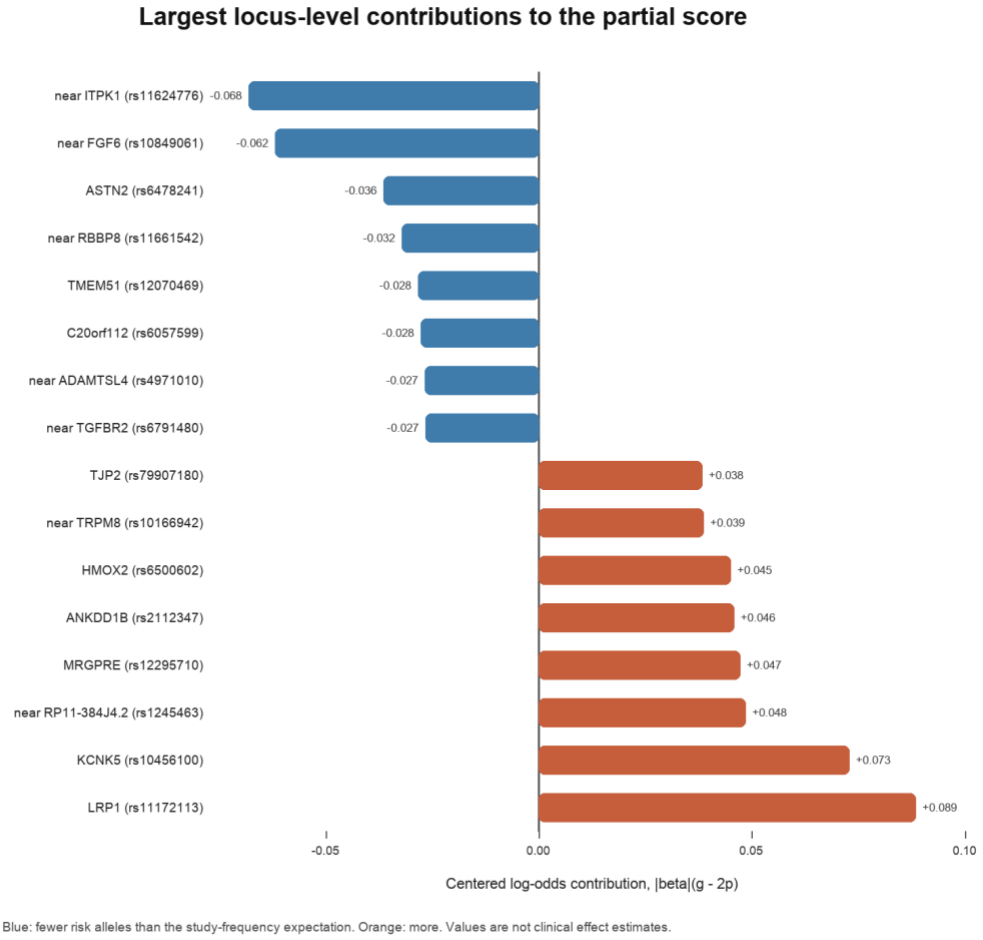

Figure 2. Largest centered locus contributions. Positive values indicate more risk alleles than the study-frequency expectation; negative values indicate fewer. These are components of an incomplete research score, not clinical effect estimates.

Notable positive and negative contributors

The largest positive centered contributions arose at LRP1, KCNK5, the RP11-384J4.2 locus, MRGPRE, and ANKDD1B. The largest negative contributions arose near ITPK1 and FGF6 and at ASTN2, near RBBP8, and at TMEM51. Locus labels identify the study's named region and must not be interpreted as proof that the named gene is causal.

Table 2. Largest negative and positive centered contributions among represented loci.

Direction	Locus	Marker	Genotype	Risk copies	OR/copy	Contribution
Below mean	near ITPK1	rs11624776	CC	0	1.051	-0.068
Below mean	near FGF6	rs10849061	TT	0	1.067	-0.062
Below mean	ASTN2	rs6478241	GG	0	1.051	-0.036
Below mean	near RBBP8	rs11661542	AA	0	1.034	-0.032
Below mean	TMEM51	rs12070469	CC	0	1.041	-0.028
Above mean	ANKDD1B	rs2112347	GG	2	1.037	+0.046
Above mean	MRGPRE	rs12295710	TT	2	1.046	+0.047
Above mean	near RP11-384J4.2	rs1245463	AA	2	1.040	+0.048
Above mean	KCNK5	rs10456100	TT	2	1.052	+0.073
Above mean	LRP1	rs11172113	TT	2	1.113	+0.089

Migraine subtype loci

The study reported three lead variants that appeared specific to migraine with aura (HMOX2 rs12598836, CACNA1A rs10405121, and MPPED2 rs11031122) and two that appeared specific to migraine without aura (near SPINK2 rs7684253 and near FECH rs8087942) [2]. None of these five exact lead variants was assayed in the consumer file. Alternative markers were available in some corresponding regions, but they were not used to claim subtype-specific risk.

Discussion

Principal finding

This proof of concept shows that an LLM can assist a technically sound personal-genomics workflow when its role is restricted to evidence discovery, orchestration, code generation, and explanation, while deterministic tools perform the genotype calculations. The resulting interpretation was more

conservative than a typical list of 'risk genes': it quantified missing coverage, balanced risk-raising and risk-lowering contributions, prevented hundreds of linked genome-wide-significant markers from being counted as independent, and declined to produce an absolute disease probability.

What 'accuracy' means in this setting

Accuracy had three testable components. Bibliographic accuracy meant that titles, dates, identifiers, and effect records were checked against PubMed, publisher supplements, the consortium distribution, and the GWAS Catalog. Computational accuracy meant that genome build, coordinates, called alleles, effect direction, and formulas were explicit and reproducible. Semantic accuracy meant that the final language distinguished association from causation, partial score from validated PRS, relative genetic burden from absolute risk, and common-variant susceptibility from rare monogenic disease. The analysis did not test predictive accuracy because no independent phenotype, comparator population, or held-out validation cohort was available.

Interpretation of the partial score

The approximately 83rd internal percentile should be read only as an algebraic summary of 85 represented loci under the published European-ancestry allele frequencies. It is not equivalent to the percentile of a genome-wide PRS. It excludes 38 lead loci and the much larger number of sub-threshold variants that contribute to migraine polygenicity [3]. It also lacks ancestry principal components, independent genotype quality control, imputation, LD-aware score construction across the genome, calibration to a target population, and phenotype validation. The result therefore supports the qualitative description 'above the study-frequency mean among covered loci,' not a clinical label such as high risk.

Clinical and ancestry boundaries

Research PRS generally require population-specific validation before clinical interpretation [5]. Scores derived primarily in European-ancestry samples can lose accuracy and amplify inequity in other populations [6]. In addition, consumer-array findings should not be used to diagnose rare familial hemiplegic migraine or other monogenic disorders; suspected clinically consequential variants require confirmation in an accredited clinical laboratory [7]. No treatment choice, medication response, or need for surveillance can be inferred from the present analysis.

Limitations

The analysis has several limitations. First, it concerns one participant and has no case-control inference. Second, the consumer array may contain calling errors and was not re-genotyped or clinically confirmed.

Third, ancestry was not inferred, yet the centering frequencies came from a European-ancestry GWAS. Fourth, only published genome-wide-significant variants were considered, which omits most polygenic signal. Fifth, alternative representatives were assigned by nearest lead position rather than a recomputed ancestry-matched LD matrix; the maximum representative-to-lead distance was 561 kb. Sixth, the approximate percentile assumed Hardy-Weinberg equilibrium and independence. Seventh, the LLM workflow was performed in one session and was not benchmarked against other models, repeated runs, or expert analysts. Finally, the client did not expose a stable model-build identifier, limiting exact computational reproduction of the LLM's narrative behavior even though the numerical analysis is deterministic.

Conclusions

An LLM can help transform consumer genotype data into a source-backed exploratory migraine analysis, but fluency is not evidence of validity. A defensible workflow must privilege primary sources, preserve genome-build and allele semantics, collapse linked associations to independent units, perform arithmetic outside the language model, and state what the data cannot support. In this single participant, the covered migraine loci produced an above-average frequency-centered partial score with substantial countervailing contributions. The result is hypothesis-generating and non-clinical. Future work should benchmark repeated LLM runs against expert-curated pipelines in ancestrally diverse validation cohorts using pre-specified scoring and calibration procedures.

Declarations

Ethics approval and consent to participate. This was a secondary analysis of the corresponding author's own previously generated genotype data. No recruitment, intervention, or third-party health record was involved. The data owner initiated the analysis and consented to publication of aggregate derived results. Raw genotypes are not shared because they are inherently identifying.

Funding. No external funding was reported for this analysis.

Competing interests. The author should declare any competing interests before submission.

Author contributions. Richard Sprague supplied the data, defined the research question, reviewed the interpretation, and is responsible for the manuscript. The LLM is not an author and cannot take responsibility for the work.

AI-use disclosure. An OpenAI Codex LLM agent assisted with literature discovery, workflow orchestration, code generation, analysis explanation, and manuscript drafting. All cited references were checked against

authoritative records, numerical results were generated by deterministic local code, and the human author remains responsible for verification and final submission.

Data availability. The raw genotype file is not publicly deposited because genome-wide genotype data are identifying. Aggregate locus-level results are provided in Supplementary Table S1. The public GWAS summary statistics were obtained from the International Headache Genetics Consortium [8].

Code availability. The deterministic analysis code is retained by the author and should be deposited in a public, versioned repository before or at preprint submission. A repository URL or DOI should be added here when available.

References

1. Gormley P, Anttila V, Winsvold BS, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet. 2016;48(8):856-866. doi:10.1038/ng.3598. PMID:27322543.

2. Hautakangas H, Winsvold BS, Ruotsalainen SE, et al. Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. Nat Genet. 2022;54(2):152-160. doi:10.1038/s41588-021-00990-0. PMID:35115687.

3. Torok D, Petschner P, Baksa D, Juhasz G. Improved polygenic risk prediction in migraine-first patients. J Headache Pain. 2024;25(1):161. doi:10.1186/s10194-024-01870-8. PMID:39333847.

4. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51(D1):D977-D985. doi:10.1093/nar/gkac1010. PMID:36350656.

5. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med. 2020;12(1):44. doi:10.1186/s13073-020-00742-5. PMID:32423490.

6. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-591. doi:10.1038/s41588-019-0379-x. PMID:30926966.

7. Tandy-Connor S, Guiltinan J, Krempely K, et al. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. Genet Med. 2018;20(12):1515-1521. doi:10.1038/gim.2018.38. PMID:29565420.

8. International Headache Genetics Consortium. Datasets and cohorts: Hautakangas et al. (2022) migraine summary statistics. https://www.headachegenetics.org/datasets-cohorts. Accessed 31 August 2026.

Supplementary material

Supplementary Table S1 is supplied as a separate tab-delimited file containing one directly assayed genome-wide-significant representative per covered lead locus. 'Alternative GWS' denotes a marker assigned to the nearest lead on the same chromosome and does not assert a newly computed LD relationship.