Prioritizing Suggestive Candidate Genes in Migraine: An Opinion

Simona Denise Frederiksen *

Independent Researcher, Calgary, AB, Canada

Keywords: GWAS, bioinformatics, gene prioritization, headache, migraine, causation, pathophysiology, membrane potential

INTRODUCTION

The search for the underlying causes of migraine has been ongoing for decades, with genome-wide association studies (GWASs) enabling the discovery of common single nucleotide polymorphisms (SNPs) associated with this disorder, along with suggestive candidate genes (examples include PHACTR1, TRPM8 and PRDM16) (1–3). Suggestive candidate genes have predominantly been selected based on their genomic location and on expert knowledge. The term “suggestive candidate gene” reveals the level of evidence of the finding (i.e., suggesting a link between a gene and condition) thereby indicating that validation is necessary.

Gene prioritization, especially relying on computationally-intensive multi-omics analyses [e.g., weighing score based on evidence source (4)], has been used to help identify candidate genes truly associated with a condition (5–7). Gene prioritization is conducted to rank “genes according to their likelihood of being associated with the disease” and thus researchers can distinguish between credible and non-credible suggestive candidate genes, and thus select the most credible genes to further study (8). Meta-analyses of GWASs have also helped to confirm findings (2, 3). Even though some SNP-condition associations are non-reproducible, it is not enough evidence to rule out those findings. As we know, many common conditions are multifactorial in nature and the genetic architecture and regulatory networks differ between individual patients (9, 10).

Despite of the initial methodologies used to identify suggestive candidate genes being similar across migraine GWASs, the use of downstream gene prioritization varies. This may in part be explained by the continuous advancements of bioinformatics tools over time, but may also be explained by a lack of defined systematic gene prioritization efforts, particularly focused on causality. Adding an additional gene prioritization step in future GWASs to further prioritize identified suggestive candidate genes may (i) reduce the reporting of false positives, and (ii) enhance our etiological understanding of the disorder. As the number of suggestive candidate genes increases together with the number of publications, there is a growing need for valid gene prioritization (11). Here, the current and potential future state of gene prioritization in migraine GWASs will be discussed.

GENE PRIORITIZATION IN MIGRAINE

Depending on the study objective, gene prioritization might help to answer the question: “What is the likelihood of the suggestive candidate genes truly causing common migraine?” Yet, here it is important to keep in mind that evidence points to a multifactorial etiology of common migraine (12), and that the causes of common migraine are largely unknown.

Further Validation of Suggestive Candidate Genes Needed

GWASs have provided some clues about the migraine etiology, particularly at the SNP level. One limitation of this approach is the uncertainty of causality. For instance, the genotyped SNPs found
to be significantly associated with migraine might be in linkage disequilibrium (LD) with the causal variants, rather than being causal themselves, and the LD structure might contain numerous genes (13, 14). Researchers have sought to find out how these SNPs may be associated with the disorder, and have frequently looked into whether those SNPs are located in coding or non-coding regions. If located in a non-coding region, suggestive candidate genes have primarily been identified focusing on the genes located nearest to the SNPs or on functionally-relevant genes in the proximate genomic region of the SNPs. It has however been found that “only about one-third of causal genes are the nearest gene to the GWAS hit” (13), and the implication of non-coding variants is rarely studied. So, even though the GWAS methodology itself is hypothesis-free, the identification of suggestive candidate genes has predominantly been hypothesis-driven. Findings from GWAS in migraine have been discussed by van den Maagdenberg and colleagues (15).

When examining existing migraine GWASs, suggestive candidate genes have primarily been identified by (i) examining genes in proximate genomic region (2, 3, 16–21), (ii) reporting the genes for coding SNPs (2, 3, 19, 20, 22, 23), (iii) reporting nearest gene (21, 22), and (iv) using LD analysis outputs for guidance (2, 16–19, 24). These methodologies cannot be used to infer causality, and selection of suggestive candidate genes in the proximate genomic regions of SNPs of interest is generally based on expert knowledge (and today’s knowledge). Therefore, GWAS findings may be biased toward the perspectives held by those experts. For example, in addition to migraine being described as a neurovascular disorder, several other theories have been proposed throughout the years. Recently, researchers have started to describe migraine as a purely neurological condition (e.g., with “primarily neuronal origin with the vascular manifestations”) (25). Other theories have arisen throughout the past decade where researchers present migraine as a neuro-glio-vascular disorder (26) or dysfunctional neuromimic pain network (27).

To account for some of these limitations, additional gene prioritization has been conducted in some migraine GWASs. Examples of applied downstream gene prioritization methodologies include tissue-based gene expression analysis (3), and expression quantitative trait locus (eQTL) analysis using human control tissues [e.g., umbilical cords (16), cerebellum and frontal cortex (3), thyroid and brain (17)]. Despite the use of some advanced tools to prioritize suggestive candidate genes in migraine GWASs, there is still a gap in gene prioritization efforts that need to be addressed (e.g., causality is not thoroughly examined). The existence of this gap can in part be explained by the difficulty in obtaining relevant omics data of diseased tissue, especially for neurological conditions.

**Additional Gene Prioritization Step in Future GWASs**

Due to the emergence of advanced bioinformatics tools, gene prioritization in GWASs can be taken a step further. This opportunity is important to consider as the combination of GWAS and eQTL does not inform us about whether “gene expression and the trait are affected by the same underlying causal variant” as stated by Zhu and colleagues (28). Causality cannot be inferred. Referring to the disease-associated loci, Cano-Gamez and Trynka state that “it is unclear which genes they regulate” (29).

There are several other reasons why filtering of the list of suggestive candidate genes is important, including (i) evidence sources such as the GWAS catalog (30, 31) are used in downstream bioinformatics analysis to examine potential involvement of genes in disease and (ii) researchers want to reveal how the genetic background of an individual influences their biological functions and disease susceptibility. If the cause(s) of a disorder is known, health professionals can provide more targeted treatment instead of just trying to manage symptoms. Importantly, applying our knowledge about genetic causes of familial/monogenic migraine may help us separate signal from noise among the GWAS findings (as causality in these cases has been established), and thus examine the clinical relevance of suggestive candidate genes in common migraine.

Currently, we know of the following monogenic forms of migraine: Familial hemiplegic migraine type 1 (FHM1; mutations in the calcium channel gene CACNA1A), type 2 (FHM2; mutations in the sodium/potassium-transporting ATPase gene ATP1A2) and type 3 (FHM3; mutations in the sodium channel gene SCN1A) (32, 33). Those genes all seem to affect neurotransmission, susceptibility for cortical spreading depression and cognitive function (34–42). Among families with migraine, mutations in several other genes, such as KCNK18 (potassium channel gene), ATXN1 (chromatin-binding factor gene) and CACNA1B (calcium channel gene), have been found (32, 43, 44). These six genes are involved in regulation of membrane potential (GO:0042391), based on ToppGene [a candidate gene prioritization tool] phenotype and functional annotations (45).

The first step toward conducting additional gene prioritization in future GWASs is to understand each component of a gene prioritization tool. A gene prioritization tool “represents a unique combination of evidence sources, prioritization strategy and input requirements”, as defined by Zolotareva and Kleine (46). Testing data, training data and evidence sources are used as inputs. Training data (genes used to prioritize) has previously been created based on established genes underlying familial forms of a disease, for example for Alzheimer’s disease (46). To obtain training data, a list of genes previously linked to migraine (e.g., causative rather than susceptible) can be created based on the biomedical literature (46–48). Information stored in databases such as ClinVar (focus on genomic variation in human health) (49) and OMIM (based on reviews of the biomedical literature by experts) (50, 51) can also be utilized to enhance the decision process.

When using gene prioritization tools like ToppGene (45), ToppNet (45) or pBRIT (52), the user also needs to define the testing data (genes to prioritize). To obtain testing data, identified suggestive candidate genes in migraine GWAS can be used. Alternatively, the complete list of suggestive candidate genes can be identified through the NHGRI-EBI GWAS catalog (i.e., a curated collection of all published human GWAS) and corresponding R package gwasraptidp (30, 31). For some tools, the user can adjust training parameters (or use default settings). If using ToppGene, this includes choosing evidence
FIGURE 1 | Proposed gene prioritization for future GWASs. (A) Suggested gene prioritization workflow. According to the described approach (i.e., creating training data focused on known familial migraine genes), the gene prioritization strategy will seek to prioritize suggestive candidate genes in relation to the genes forming the training data. This prioritization may be based on features such as similarity (e.g., focused on genetic sequence, involvement in biological pathways, or accompanying (Continued)
sources/features (e.g., the Gene Ontology (GO) resource to explore gene functions (53), and PubMed to explore the biomedical literature). Overall, evidence sources (together with computational approaches) have been used to estimate gene similarity/proximity focused on the testing and training data (46).

This proposed additional gene prioritization step is visualized in Figure 1A.

**Gene Prioritization Efforts in Existing GWASs**

Use of established gene prioritization tools can help us to more confidently predict whether a suggestive candidate gene is credible or not or, more likely, to uncover how credible a suggestive candidate gene might be (i.e., ranking by score). This may help us to facilitate the selection of genes that are most likely to be associated with the migraine, beyond the capabilities of expression data.

The suggestive candidate genes GJA1 and KCNK5 (2) (rarely reported in migraine GWASs) ranked in the top 2 based on ToppGene and pBRIT outputs (Figure 1B, being mindful that databases continuously get updated). Both genes are involved in regulation of membrane potential (GO:0042391) as are the majority of genes known to cause familial/monogenic migraine. Yet, Gormley and colleagues stated that "loci identified to date do not support the idea of common variants in ion channel genes being strong susceptibility components in prevalent forms of migraine" (2). However, recent migraine GWAS findings point in another direction. Hautakangas and colleagues found a risk variant in CACNA1A that seemed to be specific for migraine with aura, and stated that "CACNA1A seems involved in both monogenic and polygenic forms of migraine" (55).

This indicates that the proposed gene prioritization step (Figure 1A) is likely to be beneficial for future migraine GWASs.

**DISCUSSION**

Here, use of gene prioritization to score and rank suggestive candidate genes in migraine was discussed. In some migraine GWASs, expression data from control human tissues (difficulty in obtaining diseased human brain tissues) have been used to prioritize suggestive candidate genes. Even if diseased human brain tissues were used, such analysis is not able to infer causality of the genetic variants. Hence, the overall goal with this opinion piece is to advance the conversation about gene prioritization in GWAS, presented from the perspective of migraine.

As we already know of genes implicated in the causation of familial/monogenic migraine, this information may have a role to play when prioritizing suggestive candidate genes in future migraine GWASs. Our knowledge about familial/monogenic migraine (e.g., hallmarks of less prevalent migraine types) can potentially help us to better understand underlying causes of common migraine. One question worth answering is "does common migraine share genetic risk factors with familial/monogenic migraine?". Recent evidence points to some degree of shared genetic risk factors (55).

When using gene prioritization approaches, one needs to pay attention to limitations. For example, there may be a difference in prioritization performance between monogenic and polygenic disorders (focusing on predicting novel disease genes) (56), potentially due to "the assumption of functional coherence among genes contributing to the same disease" and the fact that "complex diseases tend to perturb multiple biological processes", as stated by Linghu et al. (56). Moreover, the choice of gene prioritization tool(s) and the combination of gene prioritization components (evidence sources, prioritization strategy and input requirements) is key to enhance accuracy and precision. So, how do you best separate signal from noise?

The proposed gene prioritization approach is likely to be relevant for other fields, and could be used beyond that of causation. For example, the gene prioritization could be conducted from the perspective of disorder chronification or treatment effectiveness which then will guide the creation of training data (i.e., genes used to prioritize).

**AUTHOR CONTRIBUTIONS**

SF conceptualized the opinion piece and wrote the manuscript.

**REFERENCES**

1. Zhao Y, Zhu R, Xiao T, Liu X. Genetic variants in migraine: a field synopsis and systematic re-analysis of meta-analyses. J Headache Pain. (2020) 21:13. doi: 10.1186/s10194-020-01087-5
2. Gormley P, Anttila V, Winsvold BS, Palta P, Esko T, Pers TH, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet. (2016) 48:856–66. doi: 10.1038/ng.3598
3. Anttila V, Winsvold BS, Gormley P, Kurth T, Bettella F, McMahon G, et al. Genome-wide meta-analysis identifies new susceptibility loci for migraine. Nat Genet. (2013) 45:912–7. doi: 10.1038/ng.2676
4. Sun J, Jia P, Fanous AH, Webb BT, van den Oord EJCG, Chen X, et al. multi-dimensional evidence-based candidate gene prioritization approach for complex diseases–schizophrenia as a case. Bioinformatics. (2009) 25:2595–602. doi: 10.1093/bioinformatics/btp428
5. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet. (2021) 53:1527–33. doi: 10.1038/s41588-021-00945-5
6. Nicholls HL, John CR, Watson DS, Munroe PB, Barnes MR, Cabrera CP. Reaching the end-game for GWAS: machine learning approaches for the prioritization of complex disease loci. Front Genet. (2020) 11:350. doi: 10.3389/fgene.2020.00350
7. Gandikota S, Hershey GKK, Mersha TB, GENEASE. real time bioinformatics tool for multi-omics and disease ontology exploration, analysis and visualization. Bioinformatics. (2018) 34:3160–8. doi: 10.1093/bioinformatics/bty182
8. Gill N, Singh S, Aseri TC. Computational disease gene prioritization: an appraisal. J Comput Biol. (2014) 21:456–65. doi: 10.1089/cmb.2013.0158
9. van der Wijst MGP, de Vries DH, Brugge H, Westra H-J, Franke L. An integrative approach for building personalized gene regulatory networks for precision medicine. Genome Med. (2018) 10:96. doi: 10.1186/s13073-018-0608-4
10. Subramanian M, Wojtczusczyn A, Favre L, Boughorbel S, Shan J, Letain FB, et al. Precision medicine in the era of artificial intelligence: implications in chronic disease management. J Transl Med. (2020) 18:472. doi: 10.1186/s12967-020-02658-5
11. Guada D, Sonhammer ELL. A large-scale benchmark of gene prioritization methods. Sci Rep. (2017) 7:46598. doi: 10.1038/srep46598
12. Becker F, van El CG, Ibarreta D, Zika E, Hogarth S, Borri P, et al. Genetic testing and common disorders in a public health framework: how to assess relevance and possibilities. Eur J Hum Genet. (2011) 19:56–S44. doi: 10.1038/ejhg.2010.249
13. Tam V, Patel N, Turcotte M, Bossé Y, Parié G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. (2019) 20:467–8. doi: 10.1038/s41576-019-0127-1
14. Pavlides JM, Zhu Z, Gratten J, McRae AF, Wray NR, Yang J. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. Genome Med. (2016) 8:84. doi: 10.1186/s13073-016-0338-4
15. van den Maagdenberg AMJM, Nyholt DR, Anttila V, de Vries B, Schminke U, Launer LJ, et al. Common variants at 5q33.1 predispose to migraine in African-American children. Eur J Hum Genet. (2018) 26:5. doi: 10.1186/s10194-018-0956-x
16. Maizels M, Aurora S, Heinricher M. Beyond neurovascular: migraine as a dysfunctional neurolimbic pain network. Headache. (2012) 52:1553–65. doi: 10.1111/j.1526-4610.2012.02029.x
17. Zhu Z, Zhang F, Hu H, Bokshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. (2016) 48:481–7. doi: 10.1038/ng.3338
18. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front Genet. (2020) 11:424. doi: 10.3389/fgene.2020.00424
19. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. (2019) 47:D1005–12. doi: 10.1093/nar/gky1120
20. Magnó R, Maia A-T. gwasrappid: an R package to query, download and wrangle GWAS catalog data. Bioinformatics. (2020) 36:649–50. doi: 10.1093/bioinformatics/btaa124
21. Bron C, Sutherland HG, Griffiths LR. Exploring the Hereditary Nature of Migraine. Neuropsychiatr Dis Treat. (2021) 17:1183–94. doi: 10.2147/NDT.S282562
22. Vannomlott KRI, Kors EE, Turk U, Turkdogan D, Keyser A, Broos LAM, et al. Two of novo mutations in the Na,K-ATPase gene ATP1A2 associated with pure familial hemiplegic migraine. Eur J Hum Genet. (2006) 14:553–60. doi: 10.1038/sj.ejhg.5201607
23. Mallmann RT, Elgoeta C, Slemen F, Castonguay J, Wilmes T, van den Maagdenberg A, et al. Ablation of Cav2.1 voltage-gated Ca2+ channels in mouse forebrain generates multiple cognitive impairments. PLoS One. (2013) 8:e78598. doi: 10.1371/journal.pone.0078598
24. Terpolilli NA, Dolp R, Waehnen K, Schwarzmaier SM, Rumbnail E, Todorov B, et al. Ca(V)2.1 channel mutations causing familial hemiplegic migraine type 1 increase the susceptibility for cortical spreading depolarizations and seizures and worsen outcome after experimental traumatic brain injury. Elife. (2022) 11:e74923. doi: 10.7554/eLife.74923
25. Kaja S, Payne AJ, Nielsen EO, Thompson CL, van den Maagdenberg AJM, Koulon P, et al. Differential cerebellar GABAA receptor expression in mice with mutations in Cav2.1 (P/Q-type) calcium channels. Neuroscience. (2015) 304:196–208. doi: 10.1016/j.neuroscience.2015.07.044
26. Iure A de, Mazzocchetti P, Bastioli G, Picconi B, Costa C, Marchionni I, et al. Differential effect of FHFM2 mutation on synaptic plasticity in distinct hippocampal regions. Cephalalgia. (2019) 39:1333–8. doi: 10.1177/0333102419839967
27. Leo L, Gherardini L, Barone V, De Fusco M, Pietrobon D, Pizzorusso T, et al. Increased susceptibility to cortical spreading depression in the mouse model of familial hemiplegic migraine type 2. PLoS Genet. (2011) 7:e1002129. doi: 10.1371/journal.pgen.1002129
28. Grizz SM, Radcliffe RA. Genetic effects of ATP1A2 in familial hemiplegic migraine type II and animal models. Hum Genomics. (2013) 7:8. doi: 10.1186/1479-7364-7-8
29. Bahceci D, Anderson LL, Occelli Hanbury Brown CV, Zhou C, Arnold JC. Adolescent behavioral abnormalities in a Scn1a(+/-) mouse model of Dravet syndrome. Epilepsy Behav. (2020) 103:106842. doi: 10.1016/j.yebeh.2019.106842
30. Auffenberg E, Hedrich UB, Barbieri R, Miely D, Groschup B, Wittke TV, et al. Hyperexcitable interneurons trigger cortical spreading depression in an Scn1a migraine model. J Clin Invest. (2021) 131:e142202. doi: 10.1172/JCI142202
31. Lemaire L, Desroches M, Krupa M, Pizzamiglio L, Scalmari P, Mantegazza M. Modeling NaV1.1/SCN1A sodium channel mutations in a microcircuit with realistic ion concentration dynamics suggests differential GABAergic mechanisms leading to hyperexcitability in epilepsy and hemiplegic migraine. PLoS Comput Biol. (2017) 13:e1009239. doi: 10.1371/journal.pcbi.1009239
32. Lafrenière RG, Cader MZ, Poulin J-F, Andres-Enguix I, Simoneau M, Gupta N, et al. A dominant-negative mutation in the TREK potassium channel is linked to familial migraine with aura. Nat Med. (2010) 16:1157–60. doi: 10.1038/nm.2216
33. Rasmussen AH, Kogelman LJA, Kristensen DM, Chalmer MA, Olesen J, Hansen TF. Functional gene networks reveal distinct mechanisms segregating in migraine families. Brain. (2020) 143:2945–56. doi: 10.1093/brain/awa242
45. Chen J, Bardes EE, Aronow BJ, Jegga AG. TOPPGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* (2009) 37:W305–11. doi: 10.1093/nar/gkp427

46. Zolotareva O, Kleine MA. Survey of Gene Prioritization Tools for Mendelian and Complex Human Diseases. *J Integr Bioinform.* (2019) 16:20180069. doi: 10.1515/jib-2018-0069

47. Kumar A, Samanta D, Emmday PD, Arora R. Hemiplegic Migraine. Treasure Island (FL). (2022).

48. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet.* (2009) 10:392–404. doi: 10.1038/nrg2579

49. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* (2018) 46:D1062–7. doi: 10.1093/nar/gkx1153

50. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* (2015) 43:D789–98. doi: 10.1093/nar/gku1205

51. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* (2019) 47:D1038–43. doi: 10.1093/nar/gky1151

52. Kumar AA, Van Laer L, Alaerts M, Ardeshradavani A, Moreau Y, Laukens K, et al. pBRIT: gene prioritization by correlating functional and phenotypic annotations through integrative data fusion. *Bioinformatics.* (2018) 34:2254–62. doi: 10.1093/bioinformatics/bty079

53. Raj MR, Sreerja A. Analysis of Computational Gene Prioritization Approaches. *Procedia Comput Sci.* (2018) 143:395–410. doi: 10.1016/j.procs.2018.10.411

54. Tranchevent L-C, Ardestadavani A, ElShal S, Alcaide D, Aerts J, Auboeuf D, et al. Candidate gene prioritization with Endeavour. *Nucleic Acids Res.* (2016) 44:W117–21. doi: 10.1093/nar/gkw365

55. Hautakangas H, Winsvold BS, Ruotsalainen SE, Bjornsdottir G, Harder AVL, Kogelman LJA, et al. Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *Nat Genet.* (2022) 54:152–60. doi: 10.1038/s41588-021-00990-0

56. Linghu B, Snitkin ES, Hu Z, Xia Y, Delisi C. Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol.* (2009) 10:R91. doi: 10.1186/gb-2009-10-9-r91

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Copyright © 2022 Frederiksen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**