Article

The Comparison of Searching Strategies for Genes Related to Ischemic Stroke: Case-control Human and Model Animal Studies

Gennady Khvorykh[1*], Andrey Khrunin[1], Ivan Filippenkov[1], Lyudmila Dergunova[1], Svetlana Limborska[1]

[1] National Research Centre "Kurchatov Institute", Moscow, Russia
* Correspondence: gennady_khvorykh@gmail.com
gennady_khvorykh@gmail.com, https://orcid.org/0000-0001-8927-5921 (G.K.)
khrunin-img@yandex.ru, https://orcid.org/0000-0002-7848-4688 (A.K.);
filippenkov-ib.img@yandex.ru, https://orcid.org/0000-0002-6964-3405 (I.F.);
dergunova-lv.img@yandex.ru, https://orcid.org/0000-0003-2789-2419 (L.D.);
limbor.img@yandex.ru, https://orcid.org/0000-0002-1697-6820 (S.L).

Abstract: The genetic basis of ischemic stroke (IS) remains unexplored. In this research we compared the lists of candidate genes obtained with three approaches: classical genome-wide association studies (GWAS), cluster-based GWAS and transfer of transcriptome data from rat to human subjects. The risk genes of IS downloaded from three online repositories were also included into consideration. Human orthologues of rat genes demonstrated good presence in public repositories thus pointing the potentials of rat data transfer approach. Different search strategies resulted in almost unique sets of candidate-genes. We assumed the approaches considered complement each other. The studies of genetic basis of multifactorial diseases can benefit from multiple research strategies.

Keywords: ischemic stroke, genome-wide association studies, animal models, genes

Citation: Khvorykh G., Khrunin A., Filippenkov I., Dergunova L., Limborska S. The Comparison of Searching Strategies for Genes Related to Ischemic Stroke: Case-control Human and Model Animal Studies. Journal of Clinical Physiology and Pathology (JISCPP) 2024; 3 (2): 4-7.

https://doi.org/10.59315.JISCPP.2024-3-2.4-7

Academic Editor: Igor Kastyro

Received: 15.04.24
Revised: 07.05.24
Accepted: 30.05.24
Published: 28.06.24

Publisher's Note: International Society for Clinical Physiology and Pathology (ISCPP) stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

1. Introduction.

Ischemic stroke (IS) is a multifactorial disorder with heritability reaching up to 40% depending on its subtypes [1]. About 80 genes are found to be associated with IS [2] but its genetic basis remains underexplored [3]. The key approaches to identify risk genes are linkage analysis, candidate gene studies and genome-wide association studies (GWAS), among which the last one was the most productive. Nevertheless, it has some limitations, which consequences are incomplete set of genetic markers and low reproducibility. Previously we introduced two promising extensions of GWAS and candidate gene approaches. Firstly, we demonstrated that statistical tests of individual single nucleotide polymorphisms (SNPs) can be elaborated with clustering approaches resulting in blocks of linked SNPs [4]. Secondly, we proposed and applied the protocol for the translation of the results obtained from rat models of IS into humans [5-8]. Here we present the results of comparative analysis of genes obtained with traditional and cluster-based GWAS and with transcriptome analysis of rat brains under ischemic conditions (Figure 1). Risk genes retrieved from three public repositories were also included into comparisons.

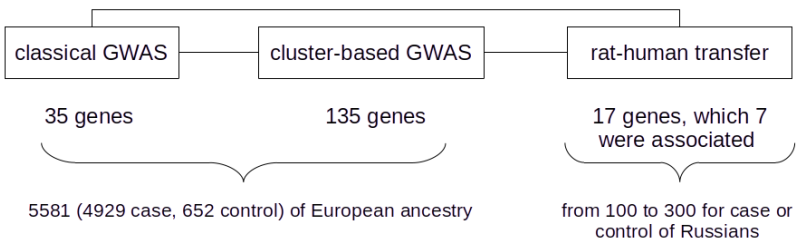

Figure 1. Candidate gene searching strategies

2. Patients and Methods

We explored SNPs in 17 genes obtained by transferring rat genes expressed differentially under tMCAO into human genome, genotyped them by real-time PCR in a cohort of individuals self-identified as Russians [5-8]. Seven of these genes were found to be associated with IS. The classical

and cluster based GWAS were made for 5581 individuals with European ancestry (4929 cases and 652 controls). The first one consisted of statistical testing of individual SNPs under different models of inheritance, the second one utilized SNP grouping with density-based spatial clustering algorithms DBSCAN [9] and HDBSCAN [10] followed by haplotype inference and statistical testing [4]. Previously we did not include the intergenic SNPs in downstream analysis. Now all SNPs were annotated and genes thus obtained considered. The annotation was made with snpEff software [11]. We also considered 131 genes associated with IS from Monarch Initiative (monarchinitiative.org, accessed on 13 June 2023) [12], 1159 genes from DisGeNET (disgenet.org, accessed on 7 October 2022) [13], and 400 genes from GWAS Central (gwascentral.org, accessed on 2 April 2023) [14].

3. Results

The classical GWAS revealed 29 SNPs significantly associated with IS, while cluster-based GWAS detected 666 and 892 SNPs from blocks associated significantly for DBSCAN and HDBSCAN, respectively [4]. The p-values in both approaches were ‹ 0.05 after Bonferroni correction. These SNPs can potentially affect 35 (classical GWAS), 1035 (DBSCAN) and 1362 (HDBSCAN) genes. The number of common genes for both algorithms of clusterization consisted of 135. They were further analyzed. This resulted in 13 common genes between classical and cluster-based approaches. The gene RUNX1 detected with classical GWAS and seven genes (USF1, CD34, KIF26B, MSX2, LHFPL3, RUNX1, and LGALS2) identified with cluster-based approach were presented in online repositories. Seventeen genes analyzed within rat-human approach contained 6 genes (CCL23, HSPB1, PTX3, CD14, LGALS3, and TSPO) from DisGeNET and RGS9 from GWAS Central (Figure 2)

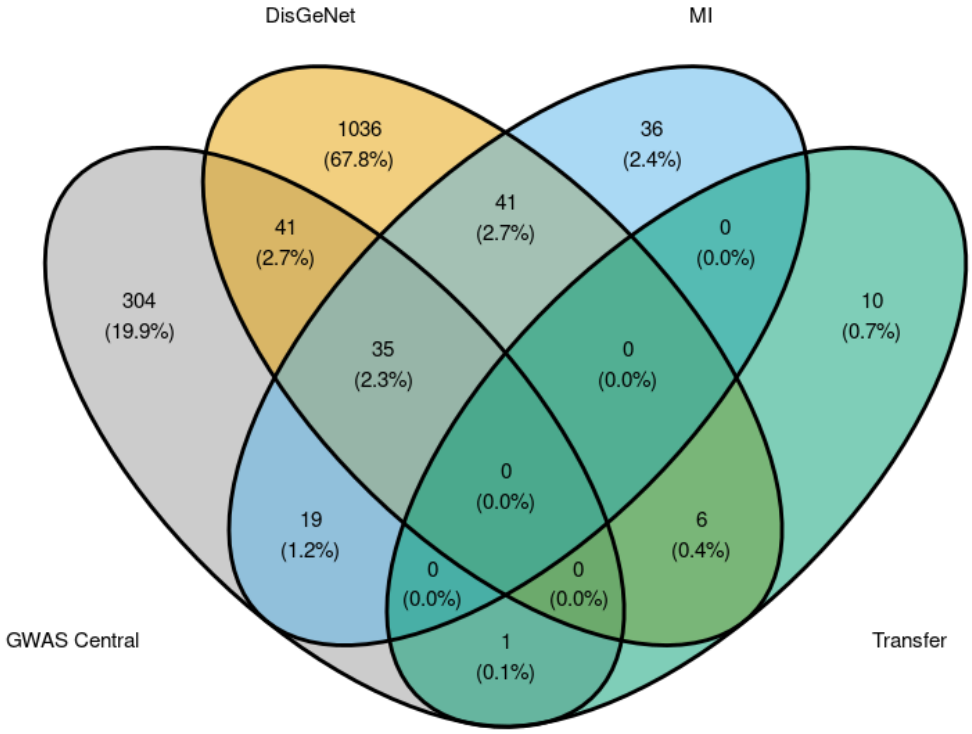

Figure 2 The intersections of candidate genes determined by rat-human transfer ap-proach with known genes in online repositories.

Neither of 7 genes validated by rat-human transfer protocol were presented in the results of classical or cluster-based GWAS. However, two of such genes, LGALS3 and PTX3, were presented in DisGeNET. Among three online resources of genes associated with IS, DisGeNET had the highest number of unique genes, that is 68.6%, while Monarch Initiative has the lowest number (2.4%).

4. Discussion

The number of candidate genes obtained previously for European cohort with cluster-based approach in-creased from 88 to 135 because of inclusion into consideration the intergenic SNPs.

Classical and cluster-based GWAS resulted in 13 common genes, thus demonstrating 14% and 78% of unique genes, respective-ly. We hypothesized cluster-based GWAS detects the genes missed by classical one since it is less influ-enced by multiple testing correction. We saw cluster-based approach resulted in more candidate genes than classical GWAS and it had greater fraction of genes in common with public resources. Therefore, it makes sense to consider the results of both approaches together. Last years, we have elaborated the protocol that allowed transferring the results of transcriptome analysis of rats under model ischemia into human studies. It is interesting to compare the results obtained with human genomic and rat transcriptomic data analysis. Previously we examined 17 human orthologues of rat genes expressed differentially under tMCAO [5-8]. All of these genes except CHRM4 were presented in GWAS data since the SNPs affecting them, according to snpEff, were genotyped and tested. Now we found that seven genes validated with rat-human transfer protocol were not reproduced by classical or cluster-based GWAS. Nine genes that were not verified with rat-human approach were also absent among the significant results of both GWAS approaches. It is clear that this comparison is preliminary, and more genes processed with transfer protocol should be considered. This is supported by the presence of 7 out of 17 genes considered in online repositories as being associated with IS. We believe this indicates a possibility for rat transfer protocol being further applied. For example, other model animals can be analyzed in similar way, allowing new genes associated with IS to be identified.

5. Conclusions

The comparison of three strategies for searching the candidate genes of IS on the level of gene lists showed that they complement each other. A combination of these methods can reinforce the studies of genetic underpinnings of ischemic stroke and other multifactorial diseases.

6. Prognosis and Conclusion

The natural history of Moyamoya disease tends to be progressive in children and adults. In studies with long-term follow-up of untreated patients, progression of neurological deficit and poor outcome were reported in 50-66% of cases. Radiographic progression within five years of diagnosis was noted in 36% of children with moyamoya. Vascular pathology is usually aggravated by extensive occlusion of intracranial large arteries and collateral circulation. Patients often suffer from cognitive and neurological decline due to recurrent ischemic stroke or hemorrhage.

Application of artificial intelligence: The review is written without the use of artificial intelligence technologies.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Funding: This research was supported by the Thematic plan of the National Research Centre "Kurchatov Institute" (5f.5.9.) (gene set comparisons) and by the Russian Science Foundation, grant number 23-14-00131 (per-forming classical and cluster-based GWAS).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, Jackson C, Farrall M, Rothwell PM, Sudlow C, Dichgans M, Markus HS. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke. 2012; 43(12):3161-7.
2. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, Kamanu FK, Koido M, Le Grand Q, Shi M, He Y, Georgakis MK, Caro I, Krebs K, Liaw YC, Vaura FC, Lin K, Winsvold BS, Srinivasasainagendra V, Parodi L, Bae HJ, Chauhan G, Chong MR, Tomppo L, Akinyemi R, Roshchupkin GV, Habib N, Jee YH, Thomassen JQ, Abedi V, Cárcel-Márquez J, Nygaard M, Leonard HL, Yang C, Yonova-Doing E, Knol MJ, Lewis AJ, Judy RL, Ago T, Amouyel P, Armstrong ND, Bakker MK, Bartz TM, Bennett DA, Bis JC, Bordes C, Børte S, Cain A, Ridker PM, Cho K, Chen Z, Cruchaga C, Cole JW, de Jager PL, de Cid R, Endres M, Ferreira LE, Geerlings MI, Gasca NC, Gudnason V, Hata J, He J, Heath AK, Ho YL, Havulinna AS, Hopewell JC, Hyacinth HI, Inouye M, Jacob MA, Jeon CE, Jern C, Kamouchi M, Keene KL, Kitazono T, Kittner SJ, Konuma T, Kumar A, Lacaze P, Launer LJ, Lee KJ, Lepik K, Li J, Li L, Manichaikul A, Markus HS, Marston NA, Meitinger T, Mitchell BD, Montellano FA, Morisaki T, Mosley TH, Nalls MA, Nordestgaard BG, O'Donnell MJ, Okada Y, Onland-Moret NC, Ovbiagele B, Peters A, Psaty BM, Rich SS, Rosand J, Sabatine MS, Sacco RL, Saleheen D, Sandset EC, Salomaa V, Sargurupremraj M, Sasaki M, Satizabal CL, Schmidt CO, Shimizu A, Smith NL, Sloane KL, Sutoh Y, Sun YV, Tanno K, Tiedt S, Tatlisumak T, Torres-Aguila NP, Tiwari HK, Trégouët DA, Trompet S, Tuladhar AM, Tybjærg-Hansen A, van Vugt M, Vibo R, Verma SS, Wiggins KL, Wennberg P, Woo D, Wilson PWF, Xu H, Yang Q, Yoon K; COMPASS Consortium; INVENT Consortium; Dutch Parelsnoer Initiative (PSI) Cerebrovascular Disease Study Group; Estonian Biobank; PRECISE4Q Consortium; FinnGen Consortium; NINDS Stroke Genetics Network (SiGN); MEGASTROKE Consortium; SIREN Consortium; China Kadoorie Biobank Collaborative Group; VA Million Veteran Program; International Stroke Genetics Consortium (ISGC); Biobank Japan; CHARGE Consortium; GIGASTROKE Consortium; Millwood IY, Gieger C, Ninomiya T, Grabe HJ, Jukema JW, Rissanen IL, Strbian D, Kim YJ, Chen PH, Mayerhofer E, Howson JMM, Irvin MR, Adams H, Wassertheil-Smoller S, Christensen K, Ikram MA, Rundek T, Worrall BB, Lathrop GM, Riaz M, Simonsick EM, Kõrv J, França PHC, Zand R, Prasad K, Frikke-Schmidt R, de Leeuw FE, Liman T, Haeusler KG, Ruigrok YM, Heuschmann PU, Longstreth WT, Jung KJ, Bastarache L, Paré G, Damrauer SM, Chasman DI, Rotter JI, Anderson CD, Zwart JA, Niiranen TJ, Fornage M, Liaw YP, Seshadri S, Fernández-Cadenas I, Walters RG, Ruff CT,

Owolabi MO, Huffman JE, Milani L, Kamatani Y, Dichgans M, Debette S. Stroke genetics informs drug discovery and risk prediction across ancestries. Nature. 2022 ;611(7934):115-123.
3. Zhang K, Loong SSE, Yuen LZH, Venketasubramanian N, Chin H-L, Lai PS, Tan BYQ. Genetics in Ischemic Stroke: Current Perspectives and Future Directions. Journal of Cardiovascular Development and Disease. 2023; 10(12):495.
4. Khvorykh GV, Sapozhnikov NA, Limborska SA, Khrunin AV. Evaluation of Density-Based Spatial Clustering for Identifying Genomic Loci Associated with Ischemic Stroke in Genome-Wide Data. International Journal of Molecular Sciences. 2023;24(20):15355.
5. Khrunin AV, Khvorykh GV, Rozhkova AV, Koltsova EA, Petrova EA, Kimelfeld EI, Limborska SA. Examination of Genetic Variants Revealed from a Rat Model of Brain Ischemia in Patients with Ischemic Stroke: A Pilot Study. Genes (Basel). 2021;12(12):1938.
6. Khrunin AV, Khvorykh GV, Arapova AS, Kulinskaya AE, Koltsova EA, Petrova EA, Kimelfeld EI, Limborska SA. The Study of the Association of Polymorphisms in LSP1, GPNMB, PDPN, TAGLN, TSPO, and TUBB6 Genes with the Risk and Outcome of Ischemic Stroke in the Russian Population. International Journal of Molecular Sciences. 2023;24(7):6831.
7. Khvorykh G, Khrunin A, Filippenkov I, Stavchansky V, Dergunova L, Limborska S. A Workflow for Selection of Single Nucleotide Polymorphic Markers for Studying of Genetics of Ischemic Stroke Outcomes. Genes (Basel). 2021; 12(3):328.
8. Khrunin AV, Khvorykh GV, Gnatko ED, Filippenkov IB, Stavchansky VV, Dergunova LV, Limborska SA. Study of polymorphism of human genes, orthologues of which are functionally involved in the response to experimental brain ischemia in model systems. Medical Genetics. 2020;19(5):83-85.
9. Ester M, Kriegel HP, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Proceedings of the KDD; Simoudis E, Han J, Fayyad UM, Eds.; AAAI Press: Cambridge. 1996; 226–231.
10. Campello RJGB, Moulavi D, Sander J. Density-Based Clustering Based on Hierarchical Density Estimates. In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Springer: 2013; 7819: 160–172
11. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92.
12. McMurry JA, Köhler S, Washington NL, Balhoff JP, Borromeo C, Brush M, Carbon S, Conlin T, Dunn N, Engelstad M, Foster E, Gourdine JP, Jacobsen JO, Keith D, Laraway B, Xuan JN, Shefchek K, Vasilevsky NA, Yuan Z, Lewis SE, Hochheiser H, Groza T, Smedley D, Robinson PN, Mungall CJ, Haendel MA. Navigating the Phenotype Frontier: The Monarch Initiative. Genetics. 2016;203(4):1491-5.
13. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Research. 2020;48(1):845-855.
14. Beck T, Rowlands T, Shorter T, Brookes AJ. GWAS Central: an expanding resource for finding and visualising genotype and phenotype data from genome-wide association studies. Nucleic Acids Research. 2023;51(1):986-993.