

RESEARCH NOTE

Open Access



An exploratory assessment of the applicability of direct-to-consumer genetic testing to translational research in Japan

Masahiro Inoue¹, Shota Arichi¹, Tsuyoshi Hachiya¹, Anna Ohtera², Seok-Won Kim², Eric Yu², Masatoshi Nishimura³, Kazuhito Shiosakai³ and Takeshi Ohira^{3*}

Abstract

Objective: In order to assess the applicability of a direct-to-consumer (DTC) genetic testing to translational research for obtaining new knowledge on relationships between drug target genes and diseases, we examined possibility of these data by associating SNPs and disease related phenotype information collected from healthy individuals.

Results: A total of 12,598 saliva samples were collected from the customers of commercial service for SNPs analysis and web survey were conducted to collect phenotype information. The collected dataset revealed similarity to the Japanese data but distinguished differences to other populations of all dataset of the 1000 Genomes Project. After confirmation of a well-known relationship between *ALDH2* and alcohol-sensitivity, Phenome-Wide Association Study (PheWAS) was performed to find association between pre-selected drug target genes and all the phenotypes. Association was found between *GRIN2B* and multiple phenotypes related to depression, which is considered reliable based on previous reports on the biological function of GRIN2B protein and its relationship with depression. These results suggest possibility of using SNPs and phenotype information collected from healthy individuals as a translational research tool for drug discovery to find relationship between a gene and a disease if it is possible to extract individuals in pre-disease states by properly designed questionnaire.

Keywords: DTC, SNPs, Phenotype, PheWAS, Translational research, GRIN2B

Introduction

Clinical trials conducted by pharmaceutical companies are sometimes discontinued due to inability to reproduce treatment effects previously confirmed in animal studies. One possible reason is the difference in drug response between disease model animals and humans. Recent advances have led to a method to evaluate similarities in drug response between animals and humans based on gene expression profiles [1]. However, it does not

completely address species difference between animals and humans. Therefore, the importance of translational research is recognized in which the relationship between a target gene and a disease is validated using data obtained from humans [2]. Probably the most straightforward approach is to obtain patient samples and examine the expression of the target gene or protein in a disease of interest; however, this cannot easily be performed. On the other hand, the Genome-Wide Association Study (GWAS) method could estimate the association between a target gene and a disease without taking patient samples by comprehensively searching for single-nucleotide polymorphism (SNP) which differs in frequency between

*Correspondence: ohira.takeshi.cy@daiichisankyo.co.jp

³ Daiichi Sankyo Co. Ltd, Tokyo, Japan

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

patients and healthy individuals. However, it may be challenging to collect a sufficient number of samples.

A previous study introduced the association between customer survey data and genotypes using the direct-to-consumer (DTC) genetic testing for antidepressant efficacy [3]. Also, in the area of personalized precision medicine, the DTC genetic testing for big data is expected to develop new drugs [4]. Therefore, to assess the applicability of a DTC genetic testing to translational research for obtaining new knowledge on relationships between drug target genes and diseases in this study, we examined the possibility of DTC genetic testing by the relationship between a target gene and a disease using a panel-based web survey of healthy individuals. SNPs of the target gene were associated with a disease via various phenotypic information related to the disease. A set of target genes were selected in advance and the Phenome-Wide Association Study (PheWAS) was employed instead of the GWAS to perform a hypothesis-free cross-phenotype search [5].

Main text

Methods

This study was performed as shown in Additional file 1: Figure S1. All samples were supplied by participants extracted from the customers of Japanese Direct-to-Consumer (DTC) genetic testing service, HealthData Lab (Yahoo! Japan Corporation, Tokyo, Japan), with consent of opt-out. In total 12,596 subjects were approached and 2 subjects (0.016%) did not opt-out. All individuals were irreversibly anonymized for confidentiality. Two genotyping platforms were used at RIKEN GENESIS (Tokyo, Japan). Sample QC was performed both individual and SNP levels. A mandatory questionnaire was conducted on all the participants on the HealthData Lab website for testing individual phenotypes. The questionnaire consists of one-hundred sixty-one questions as shown in Additional file 1: Table 1. These questions were designed to collect comprehensive information of the participants related to lifestyle, taste preference, health checkup summary, medical history and mental health condition, and not specifically designed to collect disease related information. These data were used to divide participants into case and control groups, thus generating various phenotypes, based on the answer and categorization method with multiple thresholds. 447 phenotypes with case number > 100 was considered reliable and used for analysis. Using these phenotypes, association between *ALDH2* and alcohol sensitivity was tested as a positive control to assess the reliability of the data set and analysis method while other pre-selected five genes was used as negative controls. *ALDH2* has been known as a gene which is related to alcohol metabolism [6]. According to STRING

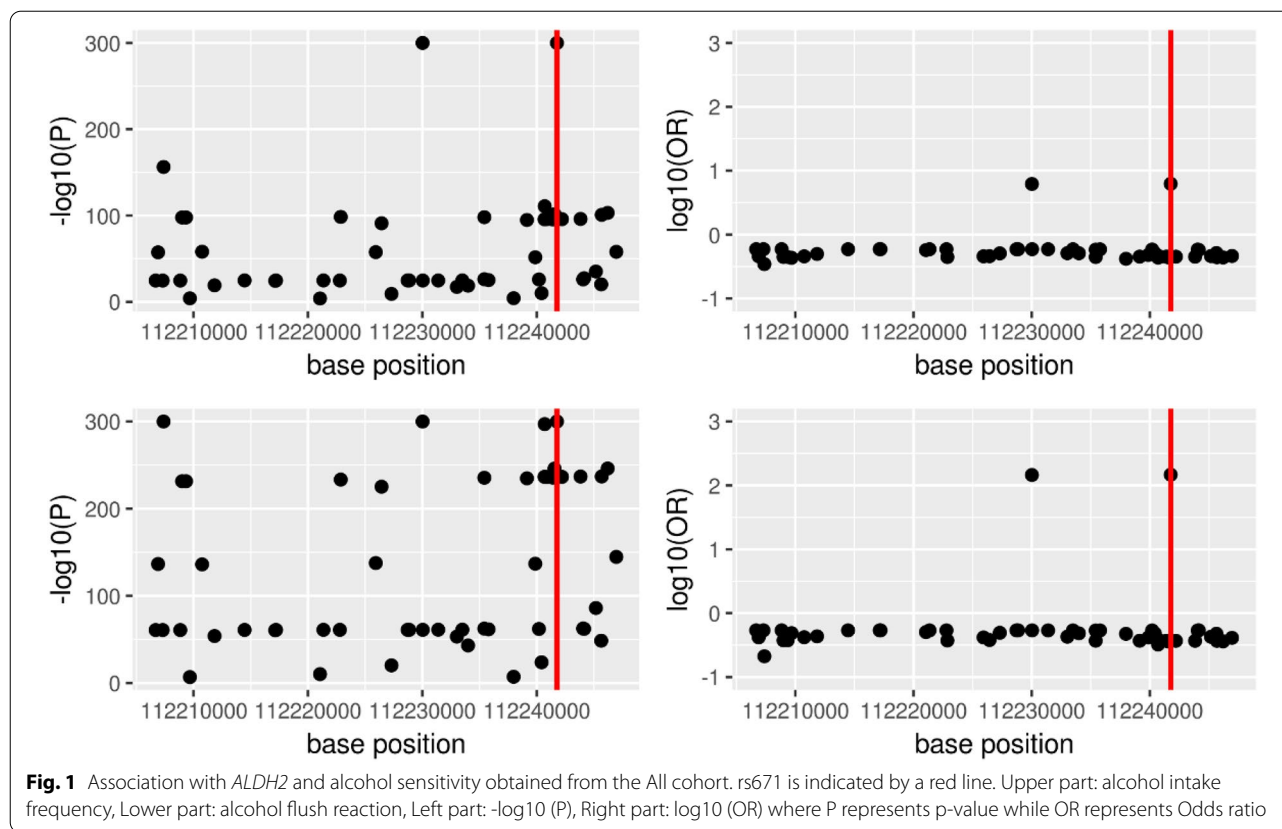
database [7], a couple of genes belonging to an alternative metabolic pathway are shown to have relationship with *ALDH2* but the degree of association between these genes and alcohol metabolism is much lower compared to *ALDH2* [6]. Therefore, *ALDH2* was used as a single and definite positive control. In this analysis, threshold of minor allele frequency was 0.03, which was used in the previous study related to neuroscience research [8]. These genes were/are targets of drugs development project in DaiichiSankyo Co., Ltd. for various disease areas including central nervous system diseases, ophthalmic diseases, immunological diseases, dyslipidemia. Then, Phenome-wide association study (PheWAS) was performed to test association between SNPs belonging to these drug target genes and all the 447 phenotypes. For these association analyses, logistic regression using an additive genetic model was employed for the following three cohorts, Male, Female, All (Male and Female). All method details are explained in the Additional file.

Results

We performed a principal component analysis (PCA) and investigated characteristics of this study data compared to currently published dataset. As shown in Additional file 1: Figure S3, our dataset revealed similarity to the Japanese data in the 1000 Genomes Project dataset, but distinguished differences to other populations of all and East Asian dataset of the 1000 Genomes Project dataset.

We compared the population of each prefecture with the Ministry of Internal Affairs and Communications (MIC) survey [9] to confirm the similarity between the population ratio used in this study and the national population. Additional file 1: Figure S4 indicates population ratio of data used in this study matched the general population investigated by the national survey.

In the examination of a relationship between *ALDH2* and alcohol-sensitive phenotypes as positive control, a significant association was found between SNPs including previously reported rs671 and multiple SNPs located around it and alcohol sensitivity represented as drinking frequency and flush reaction [10]. Figure 1 shows the result obtained from All cohort as a representative case. No significant difference was found in terms of p-value and OR between the All cohort and Male / Female cohorts. Odds ratio was higher for the flushing response than for drinking frequency. Only rs671 and rs4646776 showed very high values (147.0 and 145.6) and all the other SNPs showed values around 1 for the flush reaction. On the other hand, in all the five genes selected as drug target and used as negative controls, no p-value smaller than 1.1×10^{-7} , which is the Bonferroni adjusted threshold corresponding to p-value = 0.05, was obtained



from association analysis for both alcohol drinking frequency and flush reaction.

Among all associations between all the SNPs corresponding to the five drug target genes and all the phenotypes, no phenotype was found with p-value smaller than 1.1×10^{-7} for each model. Since all the SNPs are not mutually independent and all the phenotypes are not mutually independent either, requirement for the phenotypes to be considered significant was changed to p-value $< 1 \times 10^{-4}$, corresponding to FDR = 0.509, to generate a manageable number of phenotypes for manual investigation while keeping the ratio of false positive about 50%. As expected, even with this p-value, the number of associations was quite limited. Four genes out of five genes had a very limited number (zero or one) of association with the phenotypes which seem disease related and hence not considered reliable for further investigation. However, one of the genes (*GRIN2B*) had an association with a total of 22 phenotypes (Additional file 2: Table 5). As the major phenotype among these phenotypes, four of them are related to depression as shown in Table 1 (left part) and considered reliable enough for further investigation. Table 1 (right part) shows the number of case and control for all the cohorts (All, Male, Female) and the depression related phenotypes. It

was found that the significant association was observed mostly in Male cohort and not observed in Female cohort. This male specific association is also obvious in Fig. 2, which shows the plot of p-value and OR along the chromosome for the phenotypes and cohorts corresponding to Table 1 (right part).

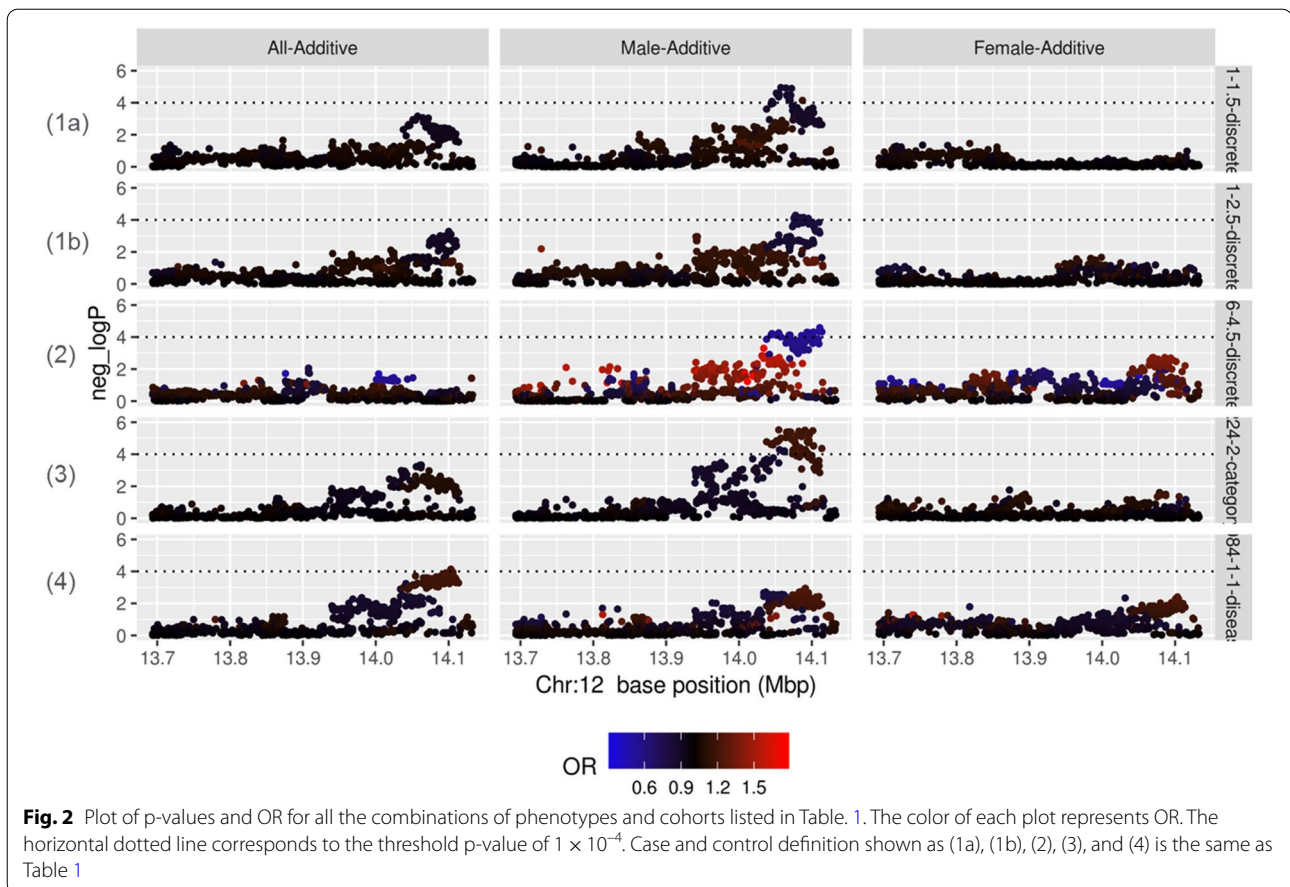
Discussion

Data generated by this study demonstrated similarity to Japanese genome dataset in the 1000 Genomes Project and regional characteristics compared to the national survey. This indicates that these data may be applied further to Japanese PheWAS studies with additional questions.

We examined the reliability of association between SNPs and the phenotype information collected by web-based survey using a known relationship as a positive control. The SNP, rs671 is known to produce an amino acid mutation (E504K) along with a nucleic acid mutation (1510G > A) to reduce its acetaldehyde metabolizing ability [10]. The strong association observed in this study between rs671 and alcohol sensitivity was consistent with this known relationship. In addition to rs671, an association between the SNP, rs4646776 and alcohol sensitivity was also observed possibly because rs4646776 is a tag

Table 1 Significant association observed between *GRIN2B* and phenotypes related to depression for certain cohorts (left part) and the number of case and control for all the cohorts (All, Male, Female) (right part). The cells enclosed by a thick line correspond to the cohort shown in the left part for which significant association is found

No	Question	Case (Definition)	Control (Definition)	Cohort	No. of associated SNPs	SNPs with the lowest p-value	P-value	OR	All Cohort		Male Cohort		Female Cohort	
									Case (N)	Control (N)	Case (N)	Control (N)	Case (N)	Control (N)
1a	Did you feel nervous?	Always Often Sometimes Rarely	Never	Male	22	rs11055680	1.13×10^{-5}	0.853 (0.794–0.916)	5,849	6,795	2,626	4,072	3,214	2,723
b	Did you feel nervous?	Always Often Sometimes	Rarely Never	Male	12	rs10845862	5.14×10^{-5}	0.801 (0.719–0.891)	2,106	10,538	897	5,801	1,200	4,737
2	Did you feel you were a worthless person?	Always	Often Sometimes Rarely Never	Male	23	rs7970022	2.62×10^{-5}	0.493 (0.355–0.686)	387	12,257	158	6,540	220	5,717
3	Did you feel uninterested in the things or could not really enjoy the things for the past one month?	Yes	No	Male	65	rs759619	8.51×10^{-7}	1.21 (1.12–1.31)	3,819	8,825	1,948	4,747	1,871	4,066
4	Did you have a history of depression?	Yes	No	All	2	rs4763360	7.34×10^{-5}	1.23 (1.11–1.35)	1,061	11,583	545	6,153	516	5,421



SNP of rs671 [11]. Associations between these SNPs and flushing reaction was stronger than associations between these SNPs and the drinking frequency. This result is interpreted to be appropriate as the flushing reaction is considered to reflect alcohol sensitivity more directly than the drinking frequency. Furthermore, the associations with alcohol sensitivity were observed only for the SNPs corresponding to *ALDH2* and not observed for the SNPs corresponding to the five drug target genes as negative controls. From the above, collection of reliable SNPs information and phenotype information is considered possible through web questionnaire if questionnaire is properly designed to define a phenotype of interest.

We conducted PheWAS cross-phenotype searches for the five drug target genes in a hypothesis-free manner and found association between *GRIN2B* and multiple depression-related phenotypes. This result appeared to be reliable, as *GRIN2B* protein mediates excitatory neurotransmission in the brain as a subunit of the NMDA receptor and there were already several previous studies about the association between *GRIN2B* and depression [12, 13]. rs1805502 was found associated with treatment-resistant depression in Han Chinese

population [12], while rs220549 was found associated with neuroticism, which is an endophenotype of MDD in European population [13]. In the present study, no association was found between these two SNPs and depression related phenotypes, which may be because of the difference of ethnicity. Furthermore, it was found that the significant association was observed mostly in Male cohort and not observed in Female cohort. The difference between the sexes in terms of association with depression-related phenotypes might be attributed to the difference in the causes of depressive mood. In female, hormonal state changes are known to have substantial impact on mood such as postpartum and postmenopausal depressions [14]. On the other hand, depressive mood caused by hormonal changes is less frequent in male [15]. The association between *GRIN2B* and depressive phenotypes was not detected in female potentially because hormonal changes might have a larger impact on depressions in female than *GRIN2B*, which might be involved in depression by a different mechanism. It may be possible to prove this hypothesis by confirming if association is found between *GRIN2B* genotype and depression related phenotypes

using female subpopulation in which hormonal state is considered stable. However, the questions used in this study were designed for gathering comprehensive information of the participants and specific questions to estimate the hormonal state change elicited during postpartum and postmenopausal period of female participants were not included. Therefore, further analysis using a newly designed set of questions to estimate female hormonal state is necessary to test the hypothesis.

In contrast, we identified a very limited disease associations in the other four genes. This is consistent with the original expectation because the questionnaire was not designed to collect disease related information, especially the target disease of these four genes (ophthalmic diseases, immunological diseases, dyslipidemia), and the number of participants were small. Therefore, it is conceivable that the association between *GRIN2B* and depression related phenotypes were obtained because the web survey contained multiple questions relating to depression-related mental states. Please note that this association was not obtained from standard case/control study where case is usually defined as patients having certain disease but from the data obtained from a larger number of healthy individuals before being diagnosed as the disease. In addition, while the number of SNPs was high for *GRIN2B* (784), the number of SNPs were medium or low for other four genes (184, 43, 1, and 1) which may worked favorably for *GRIN2B* compared to other genes in terms of detection of association. Since the relationship between *GRIN2B* and depression is much more complicated than the relationship between *ALDH2* and alcohol sensitivity, this type of association analysis could be applied to find association between a gene and a disease of some complexity with involvement of multiple genes and environmental factors when the above conditions regarding questionnaire are met.

Based on the above observation, although this study was conducted in a retrospective and hypothesis-free manner using already obtained data, it makes more sense to proactively collect phenotype data to test or refine a hypothesis between a gene and a disease as described above in the case of *GRIN2B* and depression by eliminating the effect of hormonal state change. In general, collection of phenotype information to test a hypothesis is feasible if it is possible 1) to define multiple pre-disease phenotypes of the target disease, 2) to collect these pre-disease phenotype information using a web-based questionnaire, and 3) to obtain a large number of cases. If such scheme would be implemented, analyzing SNPs and phenotype information using data obtained by DTC genetic testing and web-based questionnaire could be an effective translational research tool for drug discovery.

Conclusion

Using SNPs and phenotype information obtained from healthy individuals, well documented relationships between *ALDH2* and alcohol-sensitivity were confirmed and reproduced. Furthermore, associations between *GRIN2B* and multiple phenotypes related to depression were found. These results demonstrate the possibility of using DTC genetic testing service as a translational research tool for drug discovery to find relationship between a gene and a disease if it is possible to extract individuals in pre-disease states by properly designed questionnaire.

Limitation

This study has a limitation. We performed all analyses using commercial dataset with end of service. Furthermore, this dataset was used under the terms of use that all dataset cannot be public disclosure following the genome guideline, and all customers agreed on this terms of use; therefore, we cannot publicly provide dataset for readers to confirm reproducibility or any further information. Furthermore, we cannot disclose the name of genes except *GRIN2B* since drug development projects targeting these genes are still active.

Abbreviations

DTC: Direct-to-Consumer; FDR: False Discovery Rate; GWAS: Genome-Wide Association Study; MIC: Ministry of Internal Affairs and Communications; PheWAS: Phenome Wide Association Study; PCA: Principal Component Analysis; QC: Quality Control; SNP: Single Nucleotide Polymorphism.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13104-021-05696-4>.

Additional file 1: Method details for this study and supplementary Table 1, 2, 3 and 4, and supporting figure S1, S2, S3 and S4.

Additional file 2: Full list of the number of associated SNP of *GRIN2B* for 447 phenotypes.

Acknowledgements

We would like to thank Drs. Narahara, Yamada, and Kiyosawa (Daiichi Sankyo) for their insightful advice on the direction of this research, analysis procedure, and interpretation of the analysis result.

Authors' contributions

MI, SA, TH and performed PheWAS analysis. MN, KS and TO interpreted analysis results in PheWAS. AO, SWK and EY contributed in writing the manuscript. All authors reviewed and approved the final manuscript. All authors read and approved the final manuscript.

Funding

MI, SA, TH were employees in Yahoo! Japan, AO, SWK, EY are employees in IQVIA, and MN, KS, TO are employees in Daiichi Sankyo. This study was supported by all Daiichi Sankyo, Yahoo! Japan and IQVIA. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All authors read and approved the final manuscript.

Availability of data and materials

This study used commercial dataset with end of service; therefore, we cannot publicly provide dataset for readers to confirm reproducibility or any further information because all raw data has been deleted. However, de-identified analysis data could be provided with reasonable scientific reason. Please contact corresponding author directly.

Declarations

Ethics approval and consent to participate

This study complied with both the Act concerning protection of personal information and the Japanese ethical guidelines for human genome/gene analysis research [16, 17], and approved (reference number:17000017) by the ethics committee, the personal genetic information handling review committee (個人遺伝情報取扱審査委員会), of Yahoo! Japan Corporation. Written consent was obtained from the participants using the enclosed form within the kit where they consented for their data to be used for research.

Consent for publication

It is not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹HealthData Lab, Yahoo! Japan Corporation, Kioi Tower, Tokyo Garden Terrace Kioicho, 1-3, Kioi-cho, Chiyoda-ku, Tokyo 102-8282, Japan. ²Real World Evidence Solutions, IQVIA Solutions Japan K.K, Takanawa 4-10-18, Minato-ku, Tokyo 108-0074, Japan. ³Daiichi Sankyo Co. Ltd, Tokyo, Japan.

Received: 24 January 2021 Accepted: 14 July 2021

Published online: 23 July 2021

References

- Kiyosawa N, Manabe S. Data-intensive drug development in the information age: applications of Systems Biology/Pharmacology/Toxicology. *J Toxicol Sci*. 2016; 41(1):15–25.
- Sugiyama Y. Importance of Reverse Translational Research (rTR). *Yakugaku Zasshi*. 2017;137(6):673–9.
- Li QS, Tian C, Seabrook GR, Drevets WC, Narayan VA. Analysis of 23andMe antidepressant efficacy survey data: implication of circadian rhythm and neuroplasticity in bupropion response. *Transl Psychiatry* 2016;6(9):e889.
- Kim RS, Goossens N, Hoshida Y. Use of big data in drug development for precision medicine. *Expert Rev Precis Med Drug Dev*. 2016;1(3):245–53.
- Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205–10.
- Lind PA, Macgregor S, Heath AC, Madden PA, Montgomery GW, Martin NG, Whitfield JB. Association between in vivo alcohol metabolism and genetic variation in pathways that metabolize the carbon skeleton of ethanol and NADH reoxidation in the alcohol challenge twin study. *Alcohol Clin Exp Res*. 2012;36(12):2074–85.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.
- Pihlstrom L, Schottlaender L, Chelban V, Consortium MSAE, Meissner WG, Federoff M, Singleton A, Houlden H. Lysosomal storage disorder gene variants in multiple system atrophy. *Brain*. 2018;141(7):53.
- Prefecture, age (5 years old), population by gender-total population (as of October 1, 2015) [<https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200524&tstat=00000090001&cycle=7&year=20150&month=0&tclass1=000001011679>].
- Okada Y, Momozawa Y, Sakaue S, Kanai M, Ishigaki K, Akiyama M, Kishikawa T, Arai Y, Sasaki T, Kosaki K, et al. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nat Commun*. 2018;9(1):1631.
- Takeuchi F, Isono M, Nabika T, Katsuya T, Sugiyama T, Yamaguchi S, Kobayashi S, Ogihara T, Yamori Y, Fujioka A, et al. Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ J*. 2011;75(4):911–8.
- Zhang C, Li Z, Wu Z, Chen J, Wang Z, Peng D, Hong W, Yuan C, Wang Z, Yu S, et al. A study of N-methyl-D-aspartate receptor gene (GRIN2B) variants as predictors of treatment-resistant major depression. *Psychopharmacology*. 2014;231(4):685–93.
- Aragam N, Wang KS, Anderson JL, Liu X. TMRSS9 and GRIN2B are associated with neuroticism: a genome-wide association study in a European sample. *J Mol Neurosci*. 2013;50(2):250–6.
- Eid RS, Gobinath AR, Galea LAM. Sex differences in depression: Insights from clinical and preclinical studies. *Prog Neurobiol*. 2019;176:86–102.
- SchweizerSchubert S, Gordon JL, EisenlohrMoul TA, Meltzer-Brody S, Schmalenberger KM, Slopian R, Zietlow AL, Ehlert U, Ditzel B. Steroid Hormone Sensitivity in Reproductive Mood Disorders: On the Role of the GABAA Receptor Complex and Stress During Hormonal Transitions. *Front Med (Lausanne)* 2020;7:479646.
- Ethical Guidelines for Human Genome / Gene Analysis Research. In: Ministry of Health, Labor and Welfare. 2001.
- Tsutsumi M. Outline of revision of "Ethical Guidelines for Human Genome / Gene Analysis Research." *Organ Biology*. 2014;21(1):9–15.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

