

DSNetwork: An integrative approach to visualize predictions of variants' deleteriousness

Audrey Lemaçon¹, Marie-Pier Scott-Boyer¹, Penny Soucy¹, Régis Ongaro-Carcy¹, Jacques Simard¹, Arnaud Droit^{1*}

¹ Genomics Center, Centre Hospitalier Universitaire de Québec - Université Laval Research Center, Quebec, Quebec, Canada

* arnaud.droit@crchudequebec.ulaval.ca

Abstract

One of the most challenging tasks of the post-genome-wide association studies (GWAS) research era is the identification of functional variants among those associated with a trait for an observed GWAS signal. Several methods have been developed to evaluate the potential functional implications of genetic variants. Each of these tools has its own scoring system which forces users to become acquainted with each approach to interpret their results. From an awareness of the amount of work needed to analyze and integrate results for a single locus, we proposed a flexible and versatile approach designed to help the prioritization of variants by aggregating the predictions of their potential functional implications. This approach has been made available through a web interface called DSNetwork which acts as a single-point of entry to almost 60 reference predictors for both coding and non-coding variants and displays predictions in an easy-to-interpret visualization. We confirmed the usefulness of our methodology by successfully identifying functional variants in four breast cancer susceptibility loci. DSNetwork is an integrative web application implemented through the Shiny framework and available at: <http://romix.genome.ulaval.ca/dsnetwork>.

Author summary

Over the past years, GWAS have enabled the identification of numerous susceptibility loci associated with complex traits (<https://www.ebi.ac.uk/gwas/>). However, many of those signals contain hundreds or even thousands of significantly associated variants among which only a few are really responsible of the phenotype. Substantial efforts have been made in the development of prediction methods to prioritize variants within GWAS-associated regions to go from statistical associations, to the identification of functional variants modulating gene expression, in order to ultimately gain insight into disease pathophysiology. Unfortunately, these numerous prediction tools generate contradictory predictions rendering the interpretation of results challenging. Some tools such as VEP [McLaren et al., 2016] report their scores using a color scheme, thus acknowledging the need to assist the user in the interpretation of predictor results. Nonetheless, the multiplication of approaches can often result in an extensive amount of data that is hard to synthesize. Aware of the challenge of evaluating the potential deleteriousness of variants in the context of fine mapping analyses, we created a customizable visualization approach that was implemented in the decision support tool called DSNetwork for **D**ecision **S**upport **N**etwork. This tool enables quick access

to gold standard and new predictors for both coding and non-coding variants through an easily interpretable visualization of these predictions for a set of variants.

Introduction

Since 2006, thousands of susceptibility loci have been identified through genome-wide association studies (GWAS) for numerous traits and complex diseases, including breast cancer [MacArthur et al., 2017]. GWAS build on the concept of linkage disequilibrium (LD) to identify statistical associations between genetic variants and diseases [Visscher et al., 2017]. While this approach is powerful for locus discovery, it cannot distinguish between truly causal variants and non-functional highly correlated neighboring variants. Thus, for the vast majority of these loci, the causal variant(s) and their functional mechanisms have not yet been elucidated.

Statistical fine-mapping analyses combined with the functional annotation of genetic variants can help pinpoint the genetic variant (or variants) responsible for complex traits, or at least narrow down the number of variants underlying the observed association for further functional studies. In this regard, tremendous efforts have been put forth to assist the functional assessment of variants at risk loci and numerous scoring methods and tools have been developed to predict the deleteriousness of variants based on a number of characteristics such as sequence conservation, characteristics of amino acid substitution, and location of the variant within protein domains or 3-dimensional protein structure.

In recent years, efforts have been made towards the aggregation of many different functional annotations resulting from these scoring methods in a single integrative value called metascore [Feng, 2017, Ionita-Laza et al., 2016], an approach which seems to yield better performances than any predictor individually [Dong et al., 2015]. Although these methods demonstrate themselves to be useful, they have some limitations, notably they are not directly comparable to each other and their prediction results are sometimes contradictory.

In order to allow a quick survey of a wide range of predictors for a given list of variants and assist in the interpretation of the resulting prediction scores, we propose a flexible and integrative method capable of gathering information from multiple sources in an easy-to-interpret representation rather than a static new metascore. For this purpose, we created a single-point of entry fetching predictors for coding and non-coding variants and presenting them as a network, where the nodes represent the variants of interest and the edges the linkage disequilibrium between variants. The network is built with the aim of rendering the predictor results easier to peruse during analyses involving multiple variants, and therefore, assist in the variant prioritization process in the context of fine-mapping analyses.

This approach has been made available through a web interface called DSNetwork at: <http://romix.genome.ulaval.ca/dsnetwork>.

Materials and methods

Annotations retrieval

Variant annotations and scoring data are fetched on-the-fly from MyVariant.info high-performance web services [Xin et al., 2016] using their third-party R package. SNPnexus [Dayem Ullah et al., 2018] scorings are fetched upon request through a Python script kindly provided by the SNPnexus team. Due to their novelty and relevance for our purpose, three complementary whole-genome resources are included : LINSIGHT [Huang et al., 2017], BayesDel [Feng, 2017] and predictions and sequence

constraint data [di Iulio et al., 2018] which can be used as a proxy to score functionality and the consequences of mutations. BayesDel, LINSIGHT and Context-Dependent Tolerance scores were extracted from a local copy. LD data are computed from 1000 Genomes Phase 3 [1000 Genomes Project Consortium et al., 2015].

Visual integration

Once fetched for the variants of interest, prediction results are displayed as a network, whose components, namely the edges and nodes, are used to convey different types of information in an easy-to-comprehend way.

Nodes correspond to annotated variants and their color scheme displays prediction scores as a pie chart, where each slice represents the score of a variant for a particular predictor. For each predictor, the selected variants are ranked according to their deleteriousness and the rankings are reflected using a color gradient, ranging from green to red, where a red slice indicates a variant more likely to be damaging with regard to a particular predictor. The edges between the nodes can be used to map Linkage Disequilibrium (LD) levels between two variants. LD (squared correlation r^2) is based on a user-chosen 1000 genomes population and is represented by an absolute color gradient ranging from yellow to red. Red indicates a high disequilibrium.

In addition to standard individual predictors, our approach includes overall measures called “metascores”. We provided two types of metascores. The first type consists in an average ranking of all selected predictors, which enables a quick visualization of the global ranking of variants across a particular region. The second type of metascore incorporated several existing integrative approaches namely BayesDel, LINSIGHT, Eigen, Eigen-PC [Ionita-Laza et al., 2016] and an integrative weighted scoring for variations in the noncoding genome called IW-Scoring. This type of metascore provides absolute values comparable between analyses and gives insight into the relevance of each variant regardless of the other candidates. Figure 1 illustrates some networks related to the case studies which will be detailed hereafter.

Fig 1. Networks displaying impact prediction scores for a subset of variants of interest. A) network representing all the available predictions; B) relative metascores with missing values ranked at the end; C) absolute metascores showing rs7484123 as the best candidate with regards to deleteriousness predictions.

Implementation

The DSNetwork was created using the Shiny framework [Chang et al., 2017]. This tool provides the users with deleteriousness predictions for a selected set of coding and non-coding human variants (hg19 build) and generates a set of prioritized results for further analysis. These prediction scores are recovered from several trusted sources and presented in a user-friendly web interface. The interface is organized in three sections, namely Input, Selection and Visualization, as illustrated and described in Figure 2. For complete user guide, see S1 File.

Fig 2. Architecture overview. The first panel is dedicated to user input and parameters for data retrieval. The middle panel presents a relevant subset of annotations for each submitted variant and enables the selection of variants to be integrated in the final visualization. The bottom part on the interface is dedicated to the integrated visualization of the deleteriousness predictions under the form of a network.

Case studies

We chose to demonstrate the utility of DSNetwork in the context of the functional analysis of four breast cancer susceptibility loci identified through the latest published breast cancer association study (full description in [Michailidou et al., 2017]). This paper reports the discovery of 65 new breast cancer risk loci and deepens the functional characterisation for four regions namely 1p36, 1p34, 7q22 and 11p15. For each of these regions, the authors defined sets of credible risk variants and investigated their impact through functional assays in order to to identify the functional variants.

Results and Discussion

Prioritization of four breast cancer susceptibility loci

The original analysis defined for each of the 65 regions, a set of credible risk variants (CRV) containing variants with P-values within two orders of magnitude of the most significant SNPs in this region. They selected four loci for further evaluation namely 1p36, 1p34, 7q22 and 11p15. Initially, those regions contained respectively 54, 13, 19 and 85 significantly associated variants. The p-value cut off enabled to reduce the number of variants to respectively 1, 4, 6 and 19 CRVs. The list of variants for these loci was extracted from the original paper's Supplementary Tables 8 and 13 in the context the present analysis. Following data extraction, the analysis procedure was: 1) upload the variants of interest on the web tool, 2) fetch the annotations, 3) Visualise the variants through the overview plot, 4) visualise the available deleteriousness scores in the decision network, 5) use relative metascore visualizations to quickly identified the best candidates and finally 6) conclude.

Locus 1p36

This region contains a single CRV, rs2992756 ($P = 1.6 \times 10^{-15}$). For demonstration purposes, we selected the 30 most associated variants to put to the test. Amongst these 30 variants, 2 variants (rs200439143, rs71018084) weren't annotated by DSNetwork because of their absence from MyVariant.info service, 24 were identified as regulatory variants and 4 as non-synonymous variants. We focused our analysis on the regulatory variants.

Based on the deleteriousness scores available for this subset of variants, a quick overview of variant nodes has allowed to easily identify rs2992756 as the best candidate. Indeed, the node for this variant contained the largest proportion of red, indicating a good ranking for most of the scoring approaches (Figure 3.A). To confirm this observation, we used the relative metascore visualization (Figure 3.B). The mean rankings, clearly materialized by both the color code and the values, enabled the confirmation of rs2992756 as best candidate among the 30 most breast cancer-associated variants at the 1p36 locus. Using reporter assays, Michailidou et al. [Michailidou et al., 2017] demonstrated that the presence of the risk T-allele of this variant within *KLHDC7A* promoter significantly lowers its activity.

Fig 3. Prioritization analysis of locus 1p36. Networks representing for the 4 CRVs associated variants with breast cancer at the 1p34 locus. A) all available predictions ; B) relative metascore with missing values ranked at the end.

Locus 1p34

This region contains 4 CRVs among 13 significantly associated variants. All the variants were found by DSNetwork and identified as regulatory variants. Based on the deleteriousness scores available for this subset of variants, a quick overview of variant nodes has allowed to easily identify two variants, rs42334486 and rs7554973 as the best candidates. Indeed, the nodes for these variants contained the largest proportion of red, indicating a good ranking for most of the scoring approaches (Figure 4.A). The visualization of the mean ranking confirms rs42334486 as the most credible candidate among the CRVs (Figure 4.B). This observation is in accordance with results from Michailidou et al. [Michailidou et al., 2017], which demonstrated, using reporter assays, that the presence of the risk T-allele of this variant within a putative regulatory element (PRE) reduce *CITED4* promoter activity.

Fig 4. Prioritization analysis of locus 1p34. Networks representing for the 4 CRVs associated variants with breast cancer at the 1p34 locus. A) all available predictions ; B) relative metascore with missing values ranked at the end.

Locus 7q22

This region contains 6 CRVs among 19 significantly associated variants. All the variants were found by DSNetwork and identified as regulatory variants. Based on the deleteriousness scores available for this subset of variants, a quick overview of variant nodes has allowed to easily identify two variants, rs6961094 and rs71559437 as the best candidates. Indeed, the nodes for these variant contained the largest proportion of red, indicating a good ranking for most of the scoring approaches (Figure 5.A). The visualization of the mean ranking confirms rs6961094 and rs71559437 as the most credible candidates among the CRVs (Figure 5.B). This observations are supported by the functional experiments performed by Michailidou et al. [Michailidou et al., 2017], which demonstrated, using allele-specific Chromatin conformation capture (3C) assays, that the presence of the risk-haplotype (rs6961094 combined with rs71559437) is associated with chromatin looping between *CUX1*, *RASA4* and *PRKRIP1* promoters suggesting that the protective alleles abrogate this phenomenon.

Fig 5. Prioritization analysis of locus 7q22. Networks representing for the 19 CRVs associated variants with breast cancer at the 11p15 locus. A) relative metascore with missing values ranked at the end ; B) absolute metascores show rs7484123 and rs11246314 as the best candidates with regard to deleteriousness predictions.

Locus 11p15

This region contains 19 CRVs among 85 candidate variants. All the variants were found by DSNetwork and 18 were identified as regulatory variants and 1 as a non-synonymous variant. Among the 19 CRVs, five variants, located in the proximal promoter of *PIDD1* (a gene implicated in DNA-damage-induced apoptosis and tumorigenesis [Lin et al., 2000], namely rs7484123, rs7484068, rs11246313, rs11246314, rs11246316 were further analysed by [Michailidou et al., 2017]. They demonstrated, using reporter assays, that these variants, incorporated in a construct, significantly increased *PIDD1* promoter activity.

A quick overview of the relative and absolute metascores visualization allowed to easily prioritize the 19 CRVs (Figure 6 A and B). First, the prioritised list based on the metascores confirms the selection of these five variants as functional credible SNPs.

Indeed they are ranked at the first, second, third, fifth and eighth place out of nineteen. Moreover, we notice that variants rs7484123 and rs11246314 demonstrate a higher level of coloration, confirming them as the best candidates among the variants located in the proximal promoter of *PIDD1*. The variant rs7484123 particularly stands out as a very promising candidate for subsequent experiments.

Fig 6. Prioritization analysis of locus 11p15. Networks representing for the 19 CRVs associated variants with breast cancer at the 11p15 locus. A) relative metascore with missing values ranked at the end ; B) absolute metascores show rs7484123 and rs11246314 as the best candidates with regard to deleteriousness predictions.

Conclusion

We analysed the four regions through DSNetwork and were able to pinpoint the same most plausible causal variants than those proposed in the original paper. DSNetwork provides a user-friendly interface integrating predictors for both coding and non-coding variants in an easy-to-interpret visualization to assist the prioritization process. The use of DSNetwork greatly facilitates the selection process by aggregating the results of nearly sixty prediction approaches and easily highlights the best candidate variants for further functional analysis.

Supporting information

S1 File. User guide

Acknowledgments

We thank the BCAC for providing the impetus to create DSNetwork. We thank the SNPnexus team (Barts Cancer Institute, Queen Mary University of London) for their assistance in integrating SNPnexus data. We also thank Pr. Bing-Jian Feng (University of Utah School of Medicine) for his help and advice regarding BayesDel integration. We are also grateful to all the personnel of Arnaud Droit's lab and particularly Gwenaëlle Lemoine and Benjamin Vittrant for their advice and assistance in the preparation of this article. We wish to extend our thanks to all the current and future DSNetwork users.

Funding

The PERSPECTIVE and PERSPECTIVE I&I projects were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-1293344, grant GP1-155865), the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the Quebec Breast Cancer Foundation.

References

- 1000 Genomes Project Consortium et al., 2015. 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., and Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571):68–74.

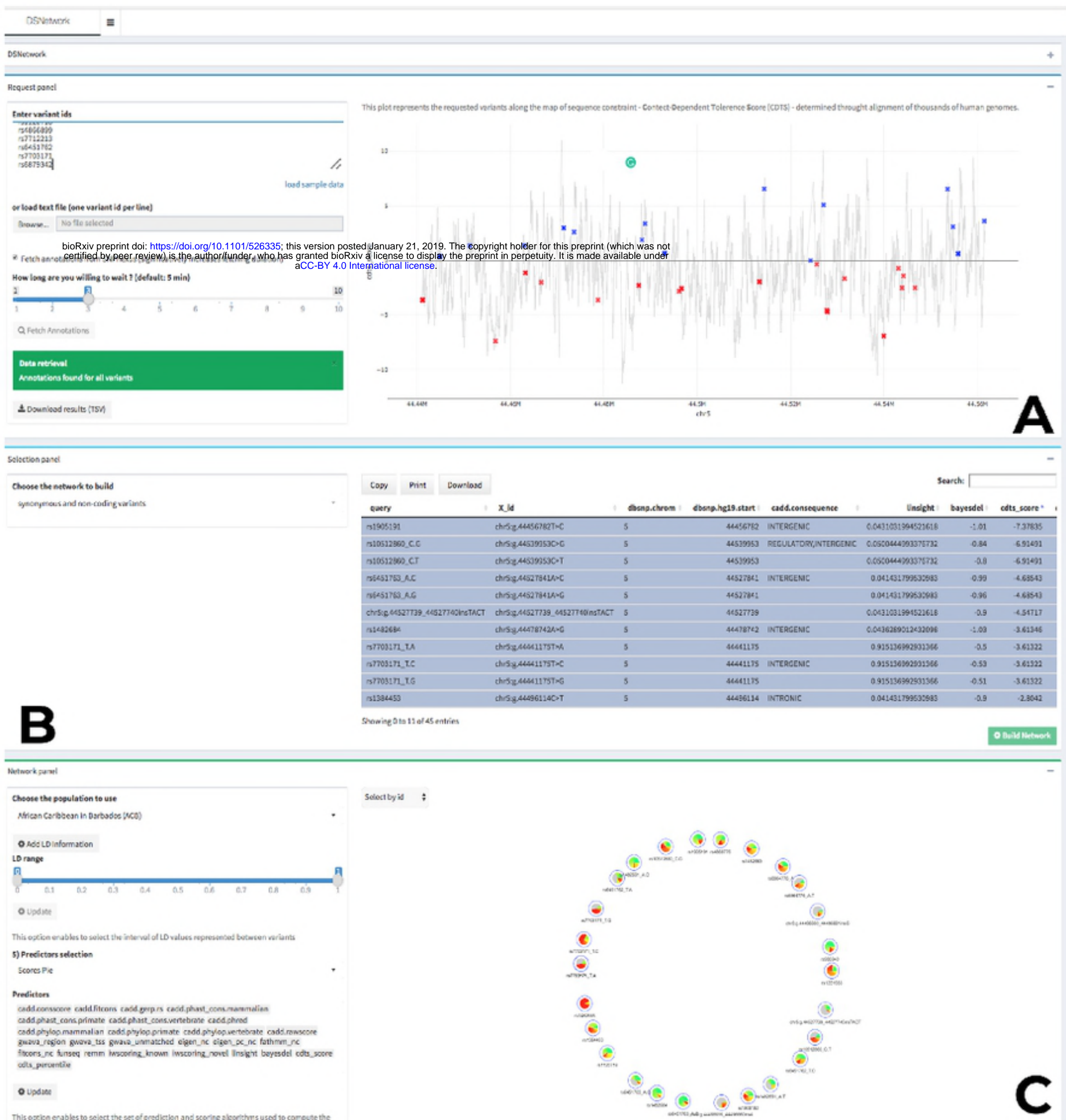
- Chang et al., 2017. Chang, W., Cheng, J., Allaire, J. J., Xie, Y., and McPherson, J. (2017). shiny: Web application framework for R [computer software]. URL <http://CRAN.R-project.org/package=shiny> (R package version 1.0.0).
- Dayem Ullah et al., 2018. Dayem Ullah, A. Z., Oscanoa, J., Wang, J., Nagano, A., Lemoine, N. R., and Chelala, C. (2018). SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.*, 46(W1):W109–W113.
- di Iulio et al., 2018. di Iulio, J., Barthä, I., Wong, E. H. M., Yu, H.-C., Lavrenko, V., Yang, D., Jung, I., Hicks, M. A., Shah, N., Kirkness, E. F., Fabani, M. M., Biggs, W. H., Ren, B., Venter, J. C., and Telenti, A. (2018). The human noncoding genome defined by genetic diversity. *Nat. Genet.*, 50(3):333–337.
- Dong et al., 2015. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.*, 24(8):2125–2137.
- Feng, 2017. Feng, B.-J. (2017). PERCH: A unified framework for disease gene prioritization. *Hum. Mutat.*, 38(3):243–251.
- Huang et al., 2017. Huang, Y.-F., Gulko, B., and Siepel, A. (2017). Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat. Genet.*, 49(4):618–624.
- Ionita-Laza et al., 2016. Ionita-Laza, I., McCallum, K., Xu, B., and Buxbaum, J. D. (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.*, 48(2):214–220.
- Lin et al., 2000. Lin, Y., Ma, W., and Benchimol, S. (2000). Pidd, a new death-domain-containing protein, is induced by p53 and promotes apoptosis. *Nat. Genet.*, 26(1):122–127.
- MacArthur et al., 2017. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., and Parkinson, H. (2017). The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res.*, 45(D1):D896–D901.
- McLaren et al., 2016. McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biol.*, 17(1):122.
- Michailidou et al., 2017. Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., French, J. D., Qing Chen, X., Fachal, L., McCue, K., McCart Reed, A. E., Ghoussaini, M., Carroll, J. S., Jiang, X., Finucane, H., Adams, M., Adank, M. A., Ahsan, H., Aittomäki, K., Anton-Culver, H., Antonenkova, N. N., Arndt, V., Aronson, K. J., Arun, B., Auer, P. L., Bacot, F., Barndahl, M., Baynes, C., Beckmann, M. W., Behrens, S., Benitez, J., Bermisheva, M., Bernstein, L., Blomqvist, C., Bogdanova, N. V., Bojesen, S. E., Bonanni, B., Børresen-Dale, A.-L., Brand, J. S., Brauch, H., Brennan, P., Brenner, H., Brinton, L., Broberg, P., Brock, I. W., Broeks, A., Brooks-Wilson, A., Brucker, S. Y., Brüning, T., Burwinkel, B., Butterbach, K., Cai, Q., Cai, H., Caldés, T., Canzian, F.,

Carracedo, A., Carter, B. D., Castela, J. E., Chan, T. L., David Cheng, T.-Y., Seng Chia, K., Choi, J.-Y., Christiansen, H., Clarke, C. L., NBCS Collaborators, Collée, M., Conroy, D. M., Cordina-Duverger, E., Cornelissen, S., Cox, D. G., Cox, A., Cross, S. S., Cunningham, J. M., Czene, K., Daly, M. B., Devilee, P., Doheny, K. F., Dörk, T., Dos-Santos-Silva, I., Dumont, M., Durcan, L., Dwek, M., Eccles, D. M., Ekici, A. B., Eliassen, A. H., Ellberg, C., Elvira, M., Engel, C., Eriksson, M., Fasching, P. A., Figueroa, J., Flesch-Janys, D., Fletcher, O., Flyger, H., Fritsch, L., Gaborieau, V., Gabrielson, M., Gago-Dominguez, M., Gao, Y.-T., Gapstur, S. M., García-Sáenz, J. A., Gaudet, M. M., Georgoulas, V., Giles, G. G., Glendon, G., Goldberg, M. S., Goldgar, D. E., González-Neira, A., Grenaker Alnæs, G. I., Grip, M., Gronwald, J., Grundy, A., Guénel, P., Haeberle, L., Hahnen, E., Haiman, C. A., Håkansson, N., Hamann, U., Hamel, N., Hankinson, S., Harrington, P., Hart, S. N., Hartikainen, J. M., Hartman, M., Hein, A., Heyworth, J., Hicks, B., Hillemanns, P., Ho, D. N., Hollestelle, A., Hooning, M. J., Hoover, R. N., Hopper, J. L., Hou, M.-F., Hsiung, C.-N., Huang, G., Humphreys, K., Ishiguro, J., Ito, H., Iwasaki, M., Iwata, H., Jakubowska, A., Janni, W., John, E. M., Johnson, N., Jones, K., Jones, M., Jukkola-Vuorinen, A., Kaaks, R., Kabisch, M., Kaczmarek, K., Kang, D., Kasuga, Y., Kerin, M. J., Khan, S., Khusnutdinova, E., Kiiski, J. I., Kim, S.-W., Knight, J. A., Kosma, V.-M., Kristensen, V. N., Krüger, U., Kwong, A., Lambrechts, D., Le Marchand, L., Lee, E., Lee, M. H., Lee, J. W., Neng Lee, C., Lejbkowitz, F., Li, J., Lilyquist, J., Lindblom, A., Lissowska, J., Lo, W.-Y., Loibl, S., Long, J., Lophatananon, A., Lubinski, J., Luccarini, C., Lux, M. P., Ma, E. S. K., MacInnis, R. J., Maishman, T., Makalic, E., Malone, K. E., Kostovska, I. M., Mannermaa, A., Manoukian, S., Manson, J. E., Margolin, S., Mariapun, S., Martinez, M. E., Matsuo, K., Mavroudis, D., McKay, J., McLean, C., Meijers-Heijboer, H., Meindl, A., Menéndez, P., Menon, U., Meyer, J., Miao, H., Miller, N., Taib, N. A. M., Muir, K., Mulligan, A. M., Mulot, C., Neuhausen, S. L., Nevanlinna, H., Neven, P., Nielsen, S. F., Noh, D.-Y., Nordestgaard, B. G., Norman, A., Olopade, O. I., Olson, J. E., Olsson, H., Olswold, C., Orr, N., Pankratz, V. S., Park, S. K., Park-Simon, T.-W., Lloyd, R., Perez, J. I. A., Peterlongo, P., Peto, J., Phillips, K.-A., Pinchev, M., Plaseska-Karanfilska, D., Prentice, R., Presneau, N., Prokofyeva, D., Pugh, E., Pylkäs, K., Rack, B., Radice, P., Rahman, N., Rennert, G., Rennert, H. S., Rhenius, V., Romero, A., Romm, J., Ruddy, K. J., Rüdiger, T., Rudolph, A., Ruebner, M., Rutgers, E. J. T., Saloustros, E., Sandler, D. P., Sangrajang, S., Sawyer, E. J., Schmidt, D. F., Schmutzler, R. K., Schneeweiss, A., Schoemaker, M. J., Schumacher, F., Schürmann, P., Scott, R. J., Scott, C., Seal, S., Seynaeve, C., Shah, M., Sharma, P., Shen, C.-Y., Sheng, G., Sherman, M. E., Shrubsole, M. J., Shu, X.-O., Smeets, A., Sohn, C., Southey, M. C., Spinelli, J. J., Stegmaier, C., Stewart-Brown, S., Stone, J., Stram, D. O., Surowy, H., Swerdlow, A., Tamimi, R., Taylor, J. A., Tengström, M., Teo, S. H., Beth Terry, M., Tessier, D. C., Thanassitichai, S., Thöne, K., Tollenaar, R. A. E. M., Tomlinson, I., Tong, L., Torres, D., Truong, T., Tseng, C.-C., Tsugane, S., Ulmer, H.-U., Ursin, G., Untch, M., Vachon, C., van Asperen, C. J., Van Den Berg, D., van den Ouweland, A. M. W., van der Kolk, L., van der Luijt, R. B., Vincent, D., Vollenweider, J., Waisfisz, Q., Wang-Gohrke, S., Weinberg, C. R., Wendt, C., Whittemore, A. S., Wildiers, H., Willett, W., Winqvist, R., Wolk, A., Wu, A. H., Xia, L., Yamaji, T., Yang, X. R., Har Yip, C., Yoo, K.-Y., Yu, J.-C., Zheng, W., Zheng, Y., Zhu, B., Ziogas, A., Ziv, E., ABCTB Investigators, ConFab/AOCS Investigators, Lakhani, S. R., Antoniou, A. C., Droit, A., Andrulis, I. L., Amos, C. I., Couch, F. J., Pharoah, P. D. P., Chang-Claude, J., Hall, P., Hunter, D. J., Milne, R. L., García-Closas, M.,

- Schmidt, M. K., Chanock, S. J., Dunning, A. M., Edwards, S. L., Bader, G. D., Chenevix-Trench, G., Simard, J., Kraft, P., and Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678):92–94.
- Visscher et al., 2017. Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.*, 101(1):5–22.
- Xin et al., 2016. Xin, J., Mark, A., Afrasiabi, C., Tsueng, G., Juchler, M., Gopal, N., Stupp, G. S., Putman, T. E., Ainscough, B. J., Griffith, O. L., Torkamani, A., Whetzel, P. L., Mungall, C. J., Mooney, S. D., Su, A. I., and Wu, C. (2016). High-performance web services for querying gene and variant annotation. *Genome Biol.*, 17(1):91.



Figure 1



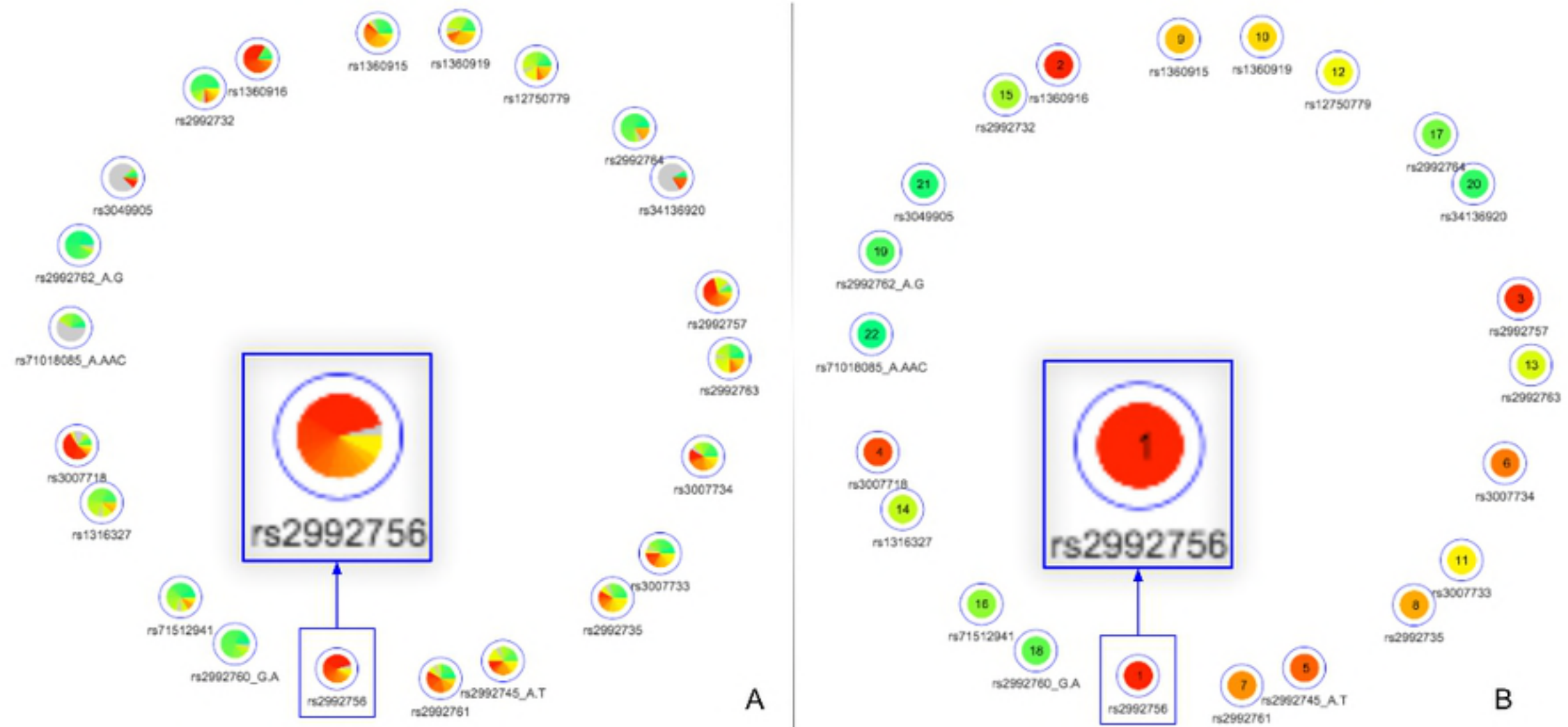


Figure 3



rs35054111_T.TA



rs4233486



rs11804913



rs7554973_T.C

A



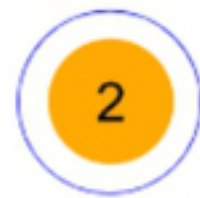
rs35054111_T.TA



rs4233486



rs11804913



rs7554973_T.C

B

Figure 4

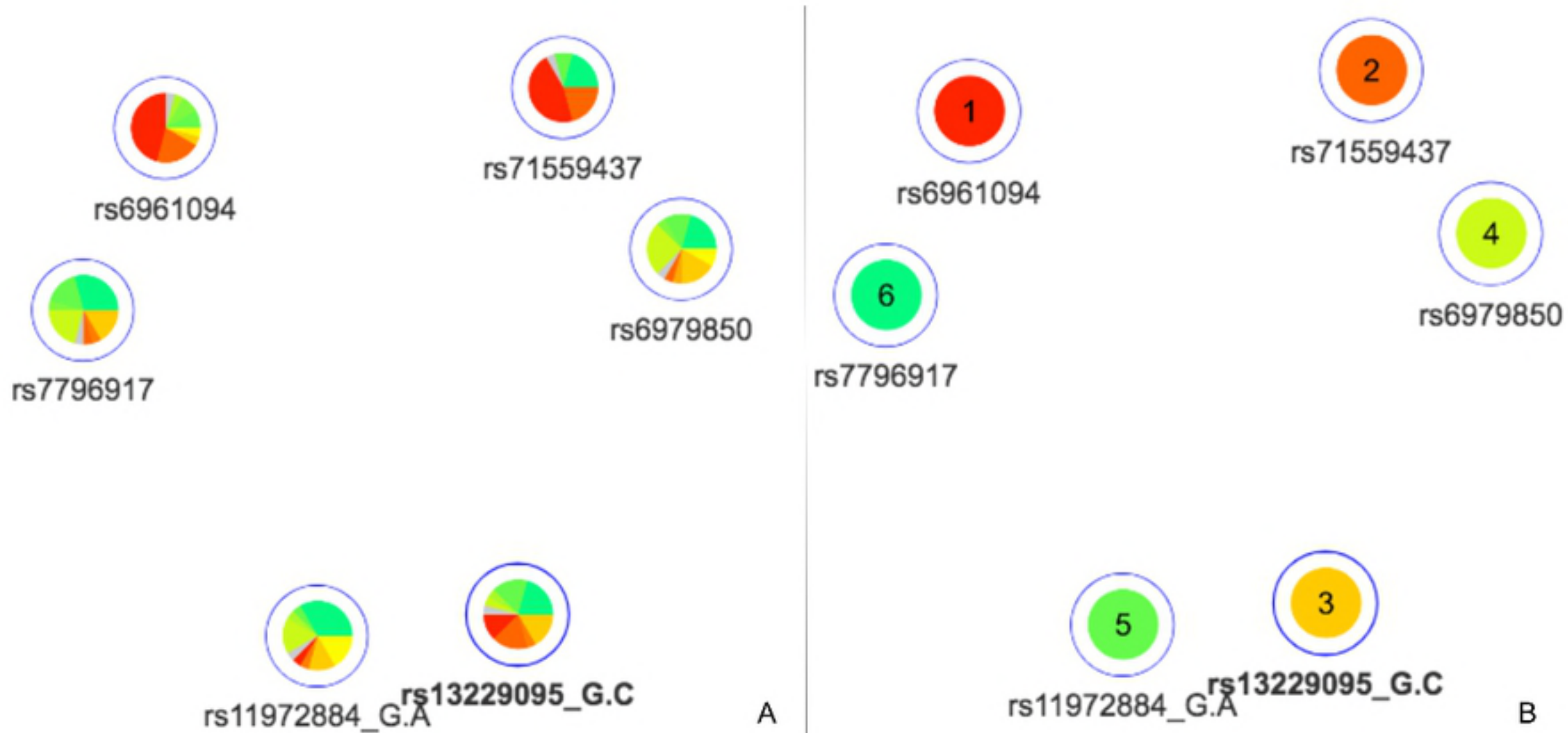


Figure 5

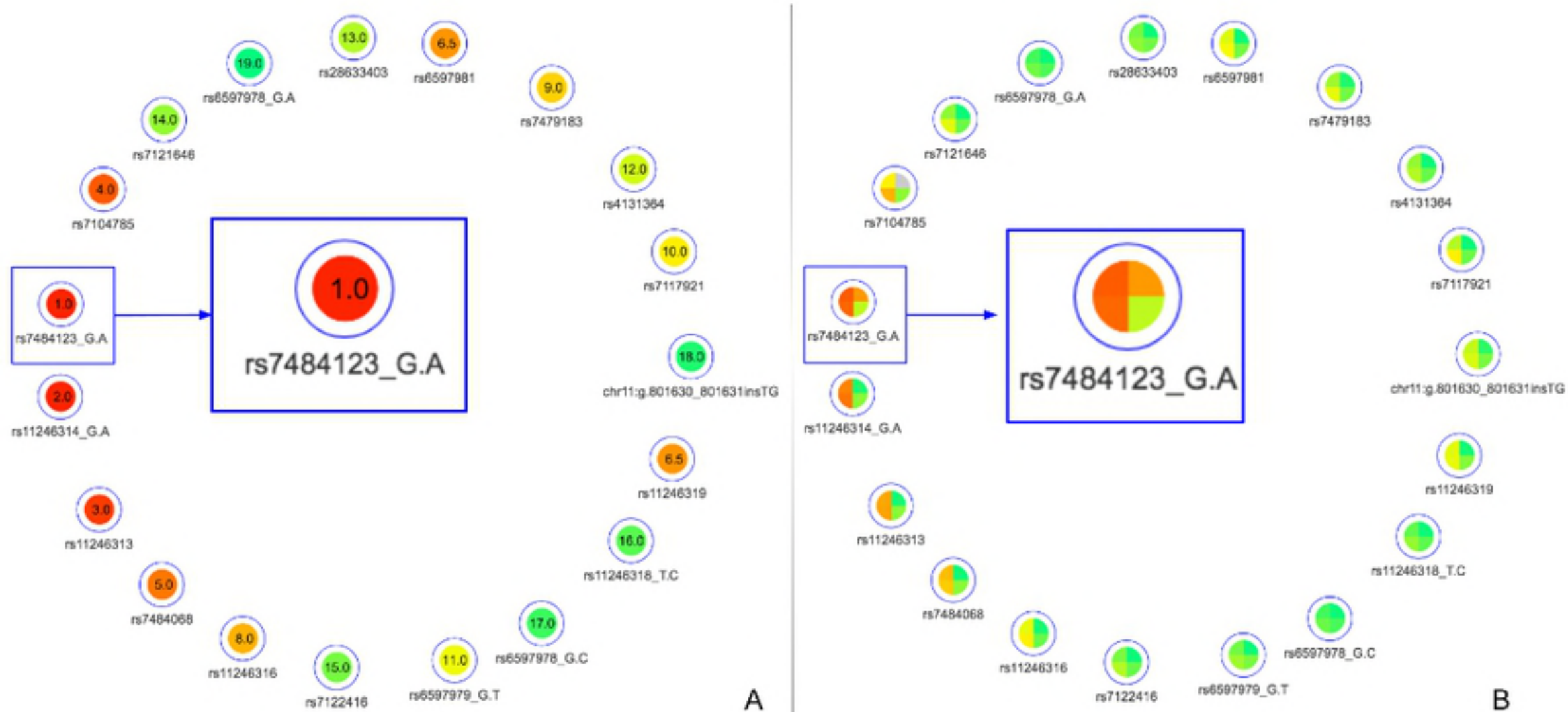


Figure 6