

# inCNV: An Integrated Analysis Tool for Copy Number Variation on Whole Exome Sequencing

Saowwapak Chanwigoon<sup>1</sup>, Sakkayaphab Piwluang<sup>1</sup> and Duangdao Wichadaku<sup>2</sup>

<sup>1</sup>Software Engineering Program, Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand. <sup>2</sup>Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand.

Evolutionary Bioinformatics  
Volume 16: 1–10  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1176934320956577



**ABSTRACT:** The detection of copy number variations (CNVs) on whole-exome sequencing (WES) represents a cost-effective technique for the study of genetic variants. This approach, however, has encountered an obstacle with high false-positive rates due to biases from exome sequencing capture kits and GC contents. Although plenty of CNV detection tools have been developed, they do not perform well with all types of CNVs. In addition, most tools lack features of genetic annotation, CNV visualization, and flexible installation, requiring users to put much effort into CNV interpretation. Here, we present “inCNV,” a web-based application that can accept multiple CNV-tool results, then integrate and prioritize them with user-friendly interfaces. This application helps users analyze the importance of called CNVs by generating CNV annotations from Ensembl, Database of Genomic Variants (DGV), ClinVar, and Online Mendelian Inheritance in Man (OMIM). Moreover, users can select and export CNVs of interest including their flanking sequences for primer design and experimental verification. We demonstrated how inCNV could help users filter and narrow down the called CNVs to a potentially novel CNV, a common CNV within a group of samples of the same disease, or a *de novo* CNV of a sample within the same family. Besides, we have provided in CNV as a docker image for ease of installation (<https://github.com/saowwapak/inCNV>).

**KEYWORDS:** Copy number variation, CNV, whole-exome sequencing, visualization

**RECEIVED:** April 24, 2020. **ACCEPTED:** August 13, 2020.

**TYPE:** Software or database Review

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work has been partially supported by the Chulalongkorn Academic Advancement into its 2nd Century (CUAAC) Project and by the Ratchadaphiseksomphot Endowment Fund Part of the “Research Grant for New Scholar CU Researcher’s Project” (RGN\_2559\_025\_06\_21).

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Duangdao Wichadaku, Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand. Email: duangdao.w@chula.ac.th

## Introduction

With the emergence of next-generation sequencing (NGS), the study of structural variants is now in great demand. Research on copy number variations (CNVs) is one of the major areas of genetics. In particular, the study of CNV on whole-exome sequencing (WES) is much more popular<sup>1</sup> and affordable<sup>2</sup> than the study on whole-genome sequencing (WGS). The study of CNVs on WES provides insights into human genome diversity and predisposition to diseases,<sup>3</sup> such as autism,<sup>4</sup> schizophrenia,<sup>5</sup> type-2 diabetes,<sup>6</sup> congenital heart disease (CHD),<sup>7</sup> non-syndromic hearing loss<sup>8,9</sup> and inherited retinal degenerations,<sup>10</sup> among others. Nevertheless, issues with the accuracy and complexity of captured CNVs as well as their interpretation still exist.

To date, no CNV detection tools have been able to accurately detect all of the different types of CNVs. Although various tools have been developed, each of them is suitable for capturing different kinds of CNVs, and their accuracy overall remains insufficient.<sup>11–13</sup> For example, the perfect usage of CONTRA<sup>14</sup> is to detect CNVs on tiny target regions, CNVnator<sup>15</sup> is used for all sizes of CNVs with high true-positive rates (TPR) but has variable false discovery rates (FDR),<sup>16</sup> and CODEX<sup>17</sup> is suitable for identifying rare CNVs.

Moreover, determining the significance of the identified CNVs is not straightforward. The output of CNV detection

tools (e.g., CONTRA, CoNIFER,<sup>17</sup> and CNVnator), in general, is provided in text format and does not include genetic or clinical data annotations. Therefore, those tools require extensive efforts to interpret and prioritize the obtained CNVs. For instance, users need to compare all the positions of CNVs on their own and use any annotation tools to determine the genetic meaning of CNVs.

Furthermore, the tools used for CNV annotation and visualization have essential limitations. First, they cannot provide any graphical user interface (GUI). For example, GenVisR,<sup>18</sup> CNView,<sup>19</sup> and iCopyDAV<sup>20</sup> require users to write code themselves to achieve the desired results. Second, although tools such as DeAnnCNV<sup>21</sup> and Ginkgo<sup>22</sup> have GUIs, they still need to apply their pre-built algorithms to detect CNVs. For example, to take advantage of CNV annotation and visualization of DeAnnCNV, the tool has to apply its algorithm and cannot cooperate with others.

Besides, the installation of many CNV tools is complicated for users without programming skills. Several tools (e.g., GenVisR and CNView) need to be installed via the command line. Some other tools are available only on their websites; if developers stop hosting their CNV tools, users will lose access to them. Last, there is a problem with the version of annotation data, in terms of which versions they are or whether they can be updated.



To meet those challenges, we present inCNV, a standalone web-based application for integrating the CNV results from multiple detection tools or samples with user-friendly interfaces. It annotates and visualizes CNV results and allows users to query, filter, and sort captured CNVs. Furthermore, inCNV can prioritize the obtained CNVs to help users in several ways: (1) finding whether the identified CNVs have been reported and/or whether they are pathogenic based on the incorporated databases consisting of the Database of Genomic Variants (DGV),<sup>23</sup> ClinVar<sup>24</sup> and OMIM<sup>25</sup>; (2) narrowing down candidates for *de novo* CNVs of a sample within the same family; (3) finding common CNVs across the samples having the same disease; (4) finding whether a sample has CNVs that may cause a specific disease. These features are highlighted in the present paper by analyzing ten samples from The Cancer Genome Atlas (TCGA).<sup>26</sup> Breast cancer genes (e.g., BRCA1, RAD51C and TP53, as reported by Zare *et al.*<sup>11</sup>) were used to demonstrate how the CNVs of these genes can be systematically identified and visualized via inCNV. From the technical aspect, users can easily install in CNV on a desktop or a server by using Docker. Regarding the system maintenance, the public databases in inCNV are automatically updated via the data source API provided at <https://github.com/saowwapark/inCNV-datasource>.

## Materials and methods

The overall architecture of inCNV (Figure 1) was divided into five main modules: (A) input data preparation; (B) management of built-in databases; (C) analysis configuration; (D) integrated CNV analyses; (E) exporting results.

### Input data preparation

inCNV provides users with a flexible upload component (Figure 2a). Users can upload CNV results from any CNV detection tool that matches a template predefined by the “CNV tool mapping” and “sample set” components. It is worth noting that users can also add new templates. The “CNV tool mapping” component (Figure 2b) allows users to define input file formats, which are used to map onto the CNV results generated from any CNV detection tools. In other words, inCNV can interpret CNV result files from any tools (see information on “input data preparation” at <https://github.com/saowwapark/inCNV>). The “sample set” component (Figure 2c) allows users to define a group of samples of CNV results they plan to upload into the system. With this component, users can define a name, a description, and sample names of the sample set. Moreover, inCNV provides a login process that allows users to analyze their files at ease without the need to upload them again. Also, users can delete their uploaded files anytime.

### Management of built-in databases

We acquired biological data from many sources in order to attach biological significance to the captured CNVs. These data were

reformatted and linked together for CNV annotations. We used them to align and visualize the given CNVs in the analysis result page. in CNV obtained data related to annotations from two sources: (1) public databases pre-built into the system; (2) additional data linked to external websites. The public databases included the DGV database released on 25-Feb-2020 for both GRCh37 and GRCh38, the ClinVar database released on 17-Mar-2020 for both GRCh37 and GRCh38, the Ensembl database with release 87 of GRCh37, and release 99 of GRCh38, as well as human reference genomes obtained from UCSC<sup>27</sup> released on 17-Jan-2020 for GRCh37 and 10-Aug-2018 for GRCh38. The DGV was used as the primary reference to known CNVs. All variants from DGV with the type of CNV were extracted and collected into the inCNV database. Furthermore, we incorporated clinical variants with OMIM from ClinVar and gene symbols from Ensembl and HUGO Gene Nomenclature Committee at the European Bioinformatics Institute (HGNC)<sup>28</sup> into the inCNV database. Human reference genomes were used to identify the flanking regions of a given CNV. To give users more details about how detected CNVs are related to DGV, ClinVar, and Ensembl, in CNV also provided the uniform resource locator (URL) linking to their websites.

### Analysis configuration

An analysis configuration module provides the process of choosing CNV results for the next procedure, “integrated CNV analyses.” The configuration module is divided into two submodules:

- (1) the configuration of individual-sampled analysis showing the integration of CNV results of one sample from numerous CNV detection tools.
- (2) the configuration of multiple-sampled analysis showing the integration of CNV results of many samples from a single CNV detection tool.

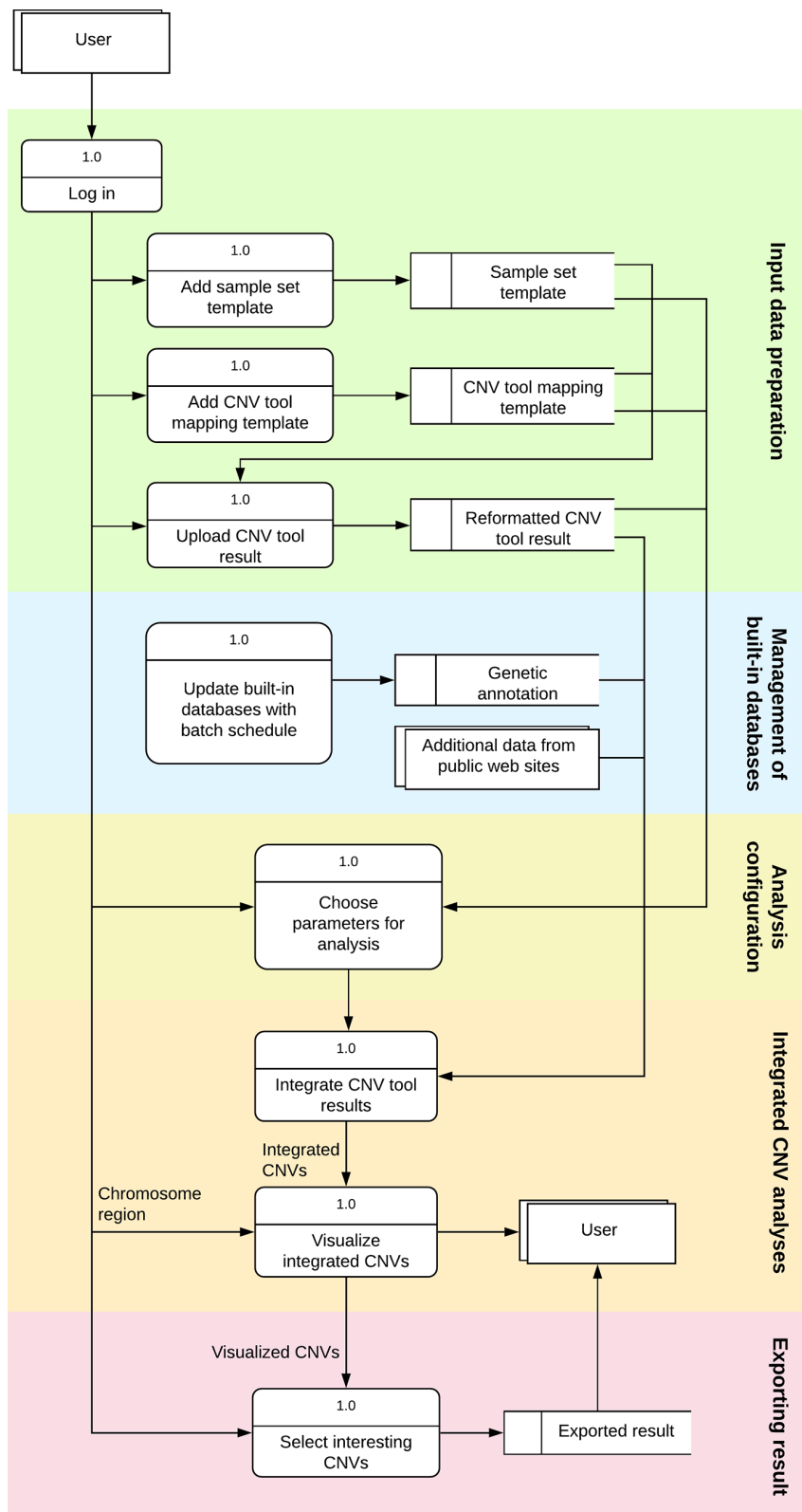
In both submodules, users filter the CNV results by a reference genome, a sample set, and a sample name. Then, users choose the results they want to integrate and specify a CNV type and a chromosome they want to see details (Figure 3).

### Integrated CNV analyses

inCNV provides two main analyses: (1) individual-sampled analysis; (2) multiple-sampled analysis. Both of them have similar features but are suitable for different use cases (details on their usage are provided in the “results and discussion” section).

The overall features of this module can be categorized into four components: (1) overview configuration; (2) overview chart; (3) all merged CNVs; (4) main chart;

*Overview configuration.* This component shows the summary information from the “analysis configuration” module, for example, a selected reference genome, a sample name, a



**Figure 1.** Overall architecture and workflow of inCNV: (1) input data preparation; (2) management of built-in databases; (3) analysis configuration; (4) integrated CNV analyses; (5) exporting results.

chromosome, a CNV type, and file names of uploaded CNV results etc (Supplemental Figure S1).

*Overview chart.* The overview chart of inCNV illustrates the distribution of CNVs along a chromosome (Figure 4a). This

chart displays the integrated result of CNVs in a broader view with a particular type of CNVs namely, duplication or deletion. The x-axis represents the base positions on a chromosome, and the y-axis signifies the number of overlapping CNVs. It is important to note that the y-axis of individual-sampled analysis

**Figure 2.** Input data preparation includes (a) an upload file component, (b) a CNV tool mapping component, (c) a sample set component.

indicates the overlap of CNV detection tools, but the y-axis of multiple-sampled analysis shows the overlap of samples. Moreover, this chart allows users to scroll the mouse wheel to zoom in and out along the chromosome and highlight an area to select a region of interest. The details of the selected region are provided in the “Main Chart” component (Figure 4b).

*All merged CNVs.* This component displays all CNVs integrated from the uploaded CNV result file(s) of multiple tools or samples, in a table form. Users can filter the results and see the details of the integrated CNVs via: (a) the filtering component; (b) the detailed table component (as described below).

*Filtering component.* Regarding “individual-sampled analysis,” users can filter the integrated CNVs with gene symbols from Ensembl, variant accession from DGV, OMIM and phenotypes from ClinVar as well as filtering by CNV detection tools (Figure 5a and Supplemental Figure S2). As regards “multiple-sampled analysis,” users can filter the integrated CNVs with the same attributes except for the CNV detection tool, replaced with the sample name.

*Detailed table component.* This component is the table in “All Merged CNVs” (Figure 5b). The table represents all information on the integrated CNVs. The information can be expanded by clicking on a particular row, consisting of annotations from DGV, Ensembl, and ClinVar, as well as flanking sequences. Users can click on any icon of these annotation sources to see more details linked to external data via the Internet. Furthermore, users can click on the table header to sort the results by such columns as start position, end position, and the number of overlapping tools (for individual-sampled analysis) or the number of overlapping samples (for multiple-sampled analysis).

*Main chart.* The purpose of the “Main chart” component is to visualize CNV details, allowing users to explore and manage identified CNVs at ease. This component comprises two parts: (1) regional selection on a chromosome (Figure 4b1) and (2) visualized charts (Figure 4b2–b5). The regional selection part allows users to specify their chromosome region of interest. The visualized charts consist of a “DGV” chart, a group of inputted-CNV charts, a “Merged CNVs” chart, and a “Selected CNVs” chart. Their x-axes represent base positions of a chromosome; moreover, regions of x-axes vary, depending on a highlight in the “Overview Chart,” selected CNVs in the “All Merged CNVs” table, or base positions in the regional selection part.

The DGV chart represents the location of CNVs based on the DGV database acquiring five CNV types: “duplication,” “deletion,” “gain,” “loss,” and “gain+loss” (Figure 4b2). Users can view the DGV information namely chromosome name, variant name, variant type, and variant subtype by hovering their mouse over the chart.

The inputted-CNV charts show the CNV results of each CNV detection tool or sample (Figure 4b3). Regarding individual-sampled analysis, the y-axis expresses the CNV tool names; in contrast, with multiple-sampled analysis, the y-axis represents the sample names. Users can hover their mouse over the inputted-CNV charts to view the corresponding base positions on a chromosome. By clicking on them, users can see an inputted-CNV dialog (Supplemental Figure S3) containing the CNV annotations, namely Ensembl, DGV, and ClinVar, as well as the flanking regions of the CNV.

The “Merged CNVs” chart visualize the integrated CNV results with a heatmap corresponding to the selected region (Figure 4b4). The heatmap uses color to indicate the overlap of CNV detection tools or samples; the more intense the color is, the

Individual Sample Analysis

Choose Reference Genome

Choose Sample Set

Choose Sample Name

4 Choose Uploaded CNV Results

Please choose at least 2 files.

Search...

Selected Files

File Name	File Info.	CNV Tool
ADTEX_TCGA-BH-A0E0.txt	threshold=+/-0.2	ADTEX
cn.MOPS_TCGA-BH-A0E0.txt	threshold=+/-0.2	cn.MOPS
CONTRA_TCGA-BH-A0E0.txt	threshold=+/-0.2	CONTRA
ExomeCNV_TCGA-BH-A0E0.txt	threshold=+/-0.2	ExomeCNV
VarScan2_TCGA-BH-A0E0.txt	threshold=+/-0.2	VarScan2

<input checked="" type="checkbox"/>	File Name	File Info.	CNV Tool	Tag Description	Created Date
<input checked="" type="checkbox"/>	ADTEX_TCGA-BH-A0E0.txt	threshold=+/-0.2	ADTEX	test	Mar 13, 2020
<input checked="" type="checkbox"/>	cn.MOPS_TCGA-BH-A0E0.txt	threshold=+/-0.2	cn.MOPS	test	Mar 13, 2020
<input checked="" type="checkbox"/>	CONTRA_TCGA-BH-A0E0.txt	threshold=+/-0.2	CONTRA	test	Mar 13, 2020
<input checked="" type="checkbox"/>	ExomeCNV_TCGA-BH-A0E0.txt	threshold=+/-0.2	ExomeCNV	test	Mar 13, 2020
<input checked="" type="checkbox"/>	VarScan2_TCGA-BH-A0E0.txt	threshold=+/-0.2	VarScan2	test	Mar 13, 2020

Items per page: 10 1 - 5 of 5

BACK NEXT

5 Choose CNV Type

6 Choose Chromosome

7 Done

**Figure 3.** Main screen for the analysis configuration of individual-sampled analysis.

more possible the CNVs becomes. However, intense color may represent a false positive depending on the algorithms of CNV detection tools. As a result, to effectively identify CNVs, users need to be aware of the characteristics of CNV detection tools suitable for different kinds of CNVs. Hovering or clicking on the “Merged CNVs” chart gives the same information as doing on the inputted-CNV chart except for having more information, namely the “overlapping CNV tools” element (for individual-sampled analysis) or the “overlapping sample names” element (for multiple-sampled analysis) and a “select CNV” checkbox element (see Supplemental Figure S4). Users can easily select the merged CNV by checking the checkbox. After that, the selected CNV will be shown in the “Selected CNVs” chart and the “All Selected CNVs” table, described as part of the “Exporting results” module.

The “Selected CNVs” chart represents the selected CNV regions (Figure 4b5). Hovering and clicking on a bar within this chart will provide the same information as doing on the “Merged CNVs” chart (Supplemental Figure S5).

### Exporting results

This module collects the selected CNVs into the “All Selected CNVs” table and exports them into a tab-delimited text file (Supplemental Figure S6). Users can select or unselect a CNV into or from the “All Selected CNVs” table via such a few components as the “All Merged CNVs” table, the “Merged CNVs” chart, and the “Selected CNVs” chart, as described previously.





**Figure 4.** Visualization of individual-sampled analysis after integrating CNV tool results from multiple CNV detection tools: ADTE<sub>x</sub>, cn.MOPS, CONTRA, ExomeCNV and VarScan2 on chromosome 17 with deletion-typed CNV. (a) Overview chart—the distribution of overlapping CNVs along a chromosome from multiple CNV detection tools. (b) Main chart. (b1) Regional selection on a chromosome—users can select the range along the chromosome to see CNVs in charts. (b2) DGV chart depicting the CNV positions of DGV along the chromosome region. (b3) A group of inputted-CNV charts—each chart represents the results of each CNV detection tool. (b4) Merged CNV chart providing integrated CNV tool results. (b5) Selected CNV chart showing the selected CNVs. (b6) Region of interesting CNVs—CNVs that are agreed by the largest number of CNV detection tools but some part of them does not match with the DGV chart.

## Implementation

inCNV was designed as a three-layer architecture comprised of (1) frontend, (2) backend, and (3) database. The frontend was developed by the Angular framework (version 9.0.0). With respect to UIs, inCNV used the Angular material UI component library (version 9.0.0) and d3.js library (version 5.14.2). The current version of inCNV supports Chrome, Opera, and Safari browsers.

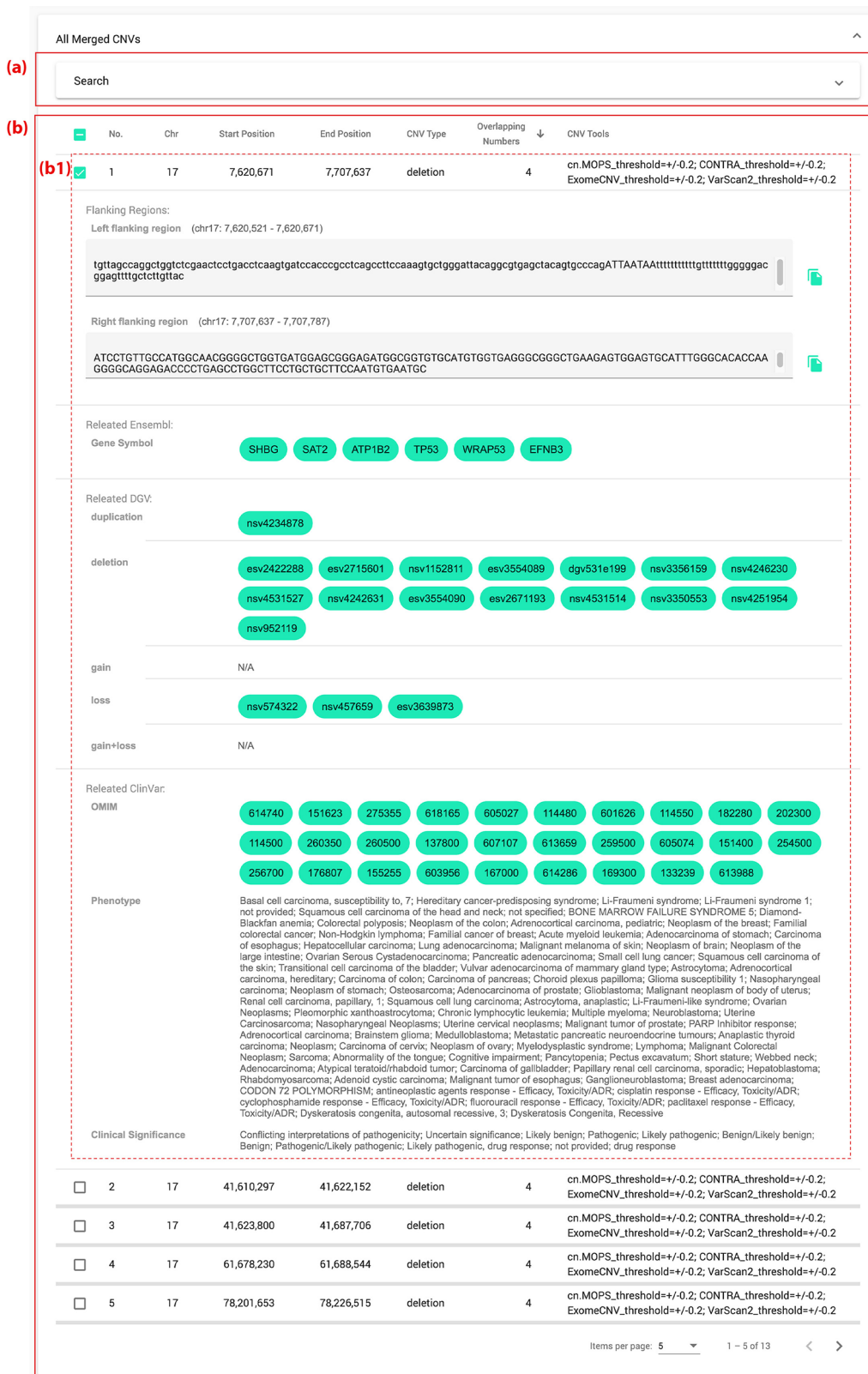
We used Node.js with Typescript and Express framework for backend development. The backend incorporated the indexed-fasta-js package (version 1.0.12) from JBrowse<sup>29</sup> to read FASTA file format. Besides, we used MySQL as our database management system (DBMS). We also provided inCNV as a docker image, which can be cloned from <https://github.com/saowwarpark/inCNV> and installed on any host with a Docker engine. The installed inCNV on a localhost will then automatically send a request to retrieve the data sources provided at <https://github.com/saowwarpark/inCNV-datasource>

in order to configure the initial database and other data sources of the system. Detailed steps for the installation can be found on our git repository.

## Data sets

To demonstrate the usage of inCNV, we used CNV data provided by Zare *et al*<sup>11</sup>. The data stemmed from running several CNV detection tools, namely ADTE<sub>x</sub>,<sup>30</sup> cn.MOPS,<sup>31</sup> CONTRA, ExomeCNV,<sup>32</sup> and VarScan2<sup>33</sup> against the WES of ten breast cancer patients in a BRCA project obtained from The Cancer Genome Atlas (TCGA).<sup>26</sup>

According to Zare *et al*<sup>11</sup> the result files from ADTE<sub>x</sub>, cn.MOPS, and ExomeCNV described the types of CNVs with standard numbers. The numbers “1,” “2,” and “3” represented a CNV deletion, no CNV (normal), and CNV duplication respectively, and numbers higher than “3” expressed amplification. For



**Figure 5.** “All Merged CNVs” of individual-sampled analysis. This shows the details of all CNVs integrated from multiple CNV tools (ADTEX, cn.MOPS, CONTRA, ExomeCNV, and VarScan2) on chromosome 17 with deletion-typed CNV and after searching for CNVs covering a region of breast cancer genes (BIRC5, BRCA1, BRIP1, ERBB2, GRB7, KPNA2, KRT17, RAD51C, or TP53). (a) Expandable searching component. (b) A detailed table. (b1) An CNV with the greatest number of overlapping tools. (b2) An expanded row showing annotation data of captured CNVs, including related genes from Ensembl, reported CNVs from DGV and clinical-related CNVs from ClinVar.

our datasets, we used the number “1” to denote CNV deletion and “3” or above to denote CNV duplication.

The result files from CONTRA and VarScan2 represented the CNV types in log<sub>2</sub> ratio. *Zare et al*<sup>11</sup> used the thresholds of  $\pm 0.2$  for the classification of CNV types. Therefore, to be consistent with *Zare et al*,<sup>11</sup> in our data sets we used log<sub>2</sub> ratio  $> +0.2$  as the criteria for CNV duplication and log<sub>2</sub> ratio  $< -0.2$  as the criteria for CNV deletion.

We modified the results files from *Zare et al*<sup>11</sup> before uploading them into inCNV: (1) Filter out unnecessary data for inCNV analysis; (2) Reformat the remaining data to match with CNV tool templates, predefined by using the “CNV tool mapping” component of the input data preparation module. The modified files were provided as “demo data” at <https://github.com/saowwapark/inCNV/tree/master/demo-data>.

## Results and discussion

inCNV provides three types of analyses: (A) individual-sampled analysis; (B) multiple-sampled analysis; (C) combined-processed analysis.

### *Individual-sampled analysis*

The individual-sampled analysis focuses on the intersection filtering of integrated CNVs from numerous CNV detection tools on one sample. To demonstrate this type of analysis, we followed the steps: (1) Integrate the CNV results of the TCGA-BH-A0E0 sample from five tools, namely ADTE<sub>x</sub>, cn.MOPS, CONTRA, ExomeCNV, and VarScan2 with the deletion type of CNV on chromosome 17 (see the “analysis configuration” module); (2) Filter the results with breast cancer genes on chromosome 17, namely BIRC5, BRCA1, BRIP1, ERBB2, GRB7, KPNA2, KRT17, RAD51C, and TP53;<sup>34</sup> (3) Sort the results by the overlapping numbers. As a consequence, thirteen CNVs overlapped between those CNV detection tools (Figure 5): Five CNVs overlapped across four tools; seven CNVs were common among three tools; one CNV overlapped across two tools. This proved that inCNV could integrate, filter, and prioritize the CNVs with multiple algorithms from several CNV detection tools at ease. Moreover, there was an interesting case in which inCNV found the overlapping CNVs in agreement with several tools but not matching with DGV as shown in Figure 4b6. This case allowed users to narrow the scope of the identification of novel CNVs. Furthermore, inCNV provides a feature of CNV flanking region extraction. This feature can extract the left and right flanking sequences of a CNV, supporting biologists to use them for primer design.

### *Multiple-sampled analysis*

The multiple-sampled analysis focuses on intersection filtering from the integrated CNV results of multiple samples identified by one CNV detection tool. To demonstrate this type of analysis, we used the CNV results obtained from CONTRA of ten

breast cancer samples. Based on the integrated CNV results by inCNV, we searched for CNVs that were only related to cancer genes like the way we did in the individual-sampled analysis. As a result, we found fourteen CNVs within or covering the region of the breast cancer genes (BIRC5, BRCA1, BRIP1, ERBB2, GRB7, KPNA2, KRT17, RAD51C, and TP53). One CNV overlapped across the seven samples. Three CNVs were common among six samples. Two CNVs overlapped across five samples, and four CNVs overlapped across three samples. These results indicated that inCNV helped determine the relationship between samples of the same disease (Figure 6 and Supplementary Figure S7). Also, inCNV can be used in several other cases. For example, inCNV can be used to find *de novo* CNVs of a sample within the same family.<sup>4,35</sup> To this end, users can filter out common or known CNVs and explore whether the remaining CNVs of the proband are different from their parents. Moreover, we can use inCNV to determine whether the sample has potential to develop a specific disease by the following steps: (1) Integrate the CNVs of the sample of interest with the CNVs of a sample set of the same disease; (2) Search for genes that are specifically related to the disease; (3) Search for the sample of interest; (4) Sort the results by the number of overlapping samples. Accordingly, if there are numerous remaining CNVs with a high frequency of overlapping, the sample of interest is prone to the disease. This determination, however, requires further experimental confirmation via wet labs.

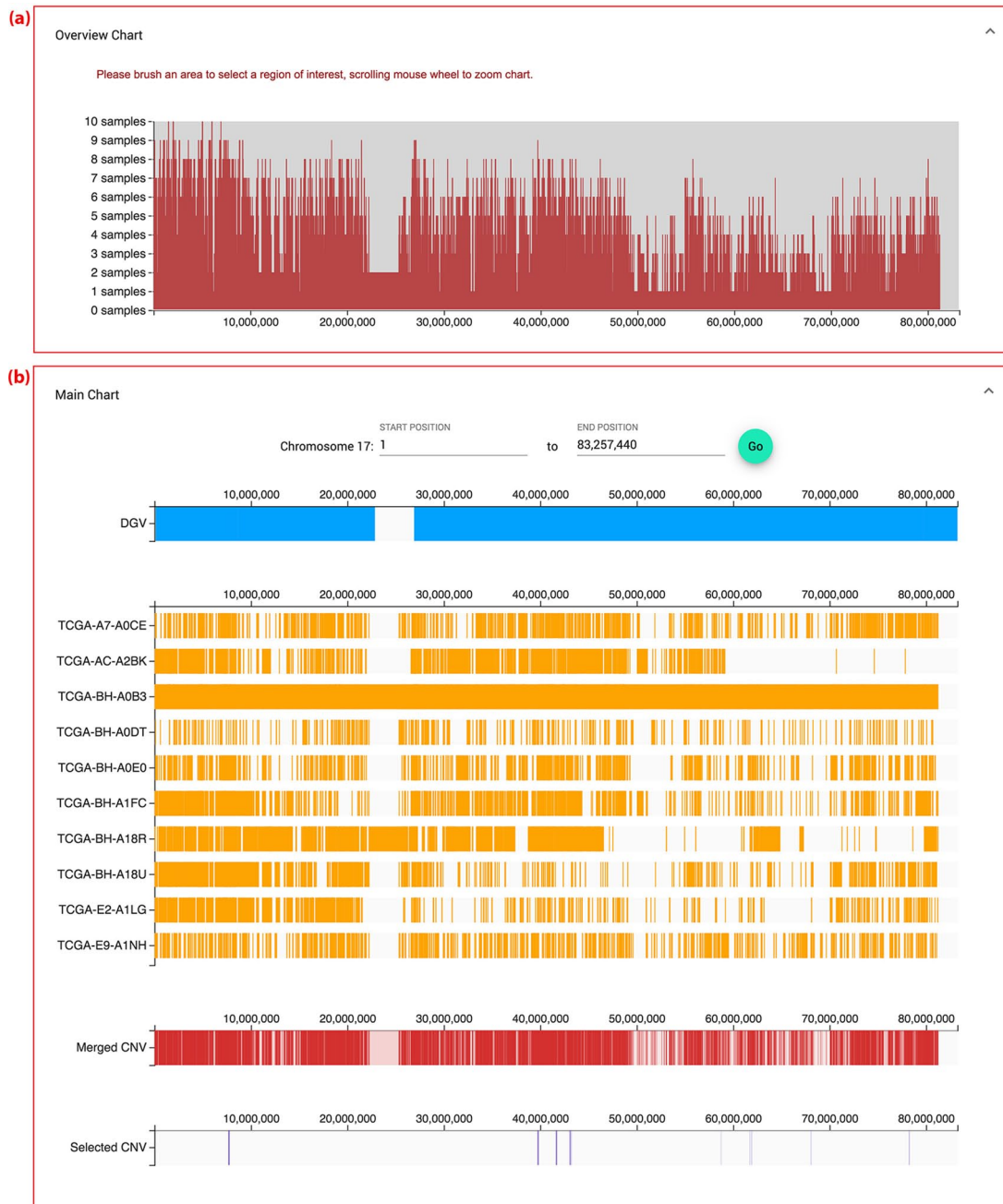
### *Combined-processed analysis*

Users can combine the results of individual-sampled and multiple-sampled analyses to identify the interesting CNV. For example, users can apply multiple-sampled analysis to find a patient who is at risk of developing a disease. Then, to prioritize CNVs with multiple algorithms, users can use individual-sampled analysis with the integration of CNV detection tools to validate the final result before conducting wet-lab experiments.

### *Minimum system requirement*

The current version of inCNV only supports real-time processing. The response time of the server with 4 CPU cores of Intel® Xeon® Gold 6140 CPU @ 2.30 GHz for multiple-sampled analysis of ten samples as described above, was about 28 seconds using 2.232 GB of RAM. From the user's perspective, based on a MacBook Pro, Dual-Core Intel Core i5 using the Opera web browser, the time and memory used to render the results were ~1 seconds with 188 MB of RAM. In regard to the multiple-sampled analysis of one hundred samples, the response time of the same server was 458 seconds with 4.495 GB of RAM, while the web browser on the user side required 8 seconds and 1 GB of RAM to render all of the results. Since the current version of inCNV only supports real-time





**Figure 6.** Visualization of multiple-sampled analysis after integrating the CNV results of ten samples of breast cancer called with CONTRA on chromosome 17 with the deletion-typed CNV. (a) Overview chart as described for the multiple-sampled analysis. (b) Main chart with a group of inputted-CNV charts, each of which represents the CNV results of each sample.

processing, we recommend installing inCNV on a server with at least 4 CPU cores and 8 GB of RAM for good user experience. We plan to also allow for analysis via batch processing in future versions.

## Conclusion

The results presented here demonstrate that inCNV can integrate CNV results from several CNV detection tools to identify CNVs with more coverage for all types of CNVs; however, the correction depends on the algorithms of the chosen CNV detection tools. Besides, inCNV can help users filter, prioritize,

select, annotate, and interpret potential CNVs for further wet-lab experiments with much less effort. To be sustainable, we have made re-formatted annotation data public so that users can see the provided online data and its released versions, and inCNV can automatically update the genome annotation data from our online resources. Moreover, inCNV is suitable for users who lack knowledge of programming as it can be installed via a docker, and users can easily run it with graphical user interfaces (GUIs). Finally, inCNV can also be used for whole genome sequencing (WGS) by following the same steps described in this paper.

## Acknowledgements

The authors would like to thank Natsuda Navamajiti for inspiring the key concept of this study on integrating CNV results by focusing on ease of understanding, rather than merely providing genetic information that is hard to interpret.

## Author Contributions

DW and SC conceptualized the manuscript. SC investigated the manuscript. SC methodologized the manuscript. SC implemented the software for this manuscript. SP provided technical support for this manuscript. Updated data sources designed by SP and implemented by SC. SP provided docker for installation. DW supervised the manuscript. SC drafted the original manuscript. DW wrote, reviewed and edited the manuscript.

## ORCID iD

Saowwapark Chanwigoon  <https://orcid.org/0000-0002-0608-5434>

## Supplemental material

Supplemental material for this article is available online.

## REFERENCES

- Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet.* 2011;48:580.
- National Human Genome Research Institute. The cost of sequencing a human genome. <https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/>. Accessed March 20, 2020.
- Pfundt R, Del Rosario M, Vissers L, et al. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. *Genet Med.* 2017;19:667–675.
- Bacchelli E, Cameli C, Viggiano M, et al. An integrated analysis of rare CNV and exome variation in Autism Spectrum Disorder using the Infinium PsychArray. *Sci Rep.* 2020;10:3198.
- Szatkiewicz JP, Neale BM, O'Dushlaine C, et al. Detecting large copy number variants using exome genotyping arrays in a large Swedish schizophrenia sample. *Mol Psychiatry.* 2013;18(11):1178–1184.
- Dajani R, Li J, Wei Z, et al. CNV analysis associates AKNAD1 with type-2 diabetes in Jordan Subpopulations. *Sci Rep.* 2015;5:13391.
- Glessner JT, Bick AG, Ito K, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ Res.* 2014;115(10):884–896.
- Park G, Gim J, Kim AR, et al. Multiphasic analysis of whole exome sequencing data identifies a novel mutation of ACTG1 in a nonsyndromic hearing loss family. *BMC Genomics.* 2013;14:191.
- Shearer AE, Kolbe DL, Azaiez H, et al. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med.* 2014;6:37.
- Zampaglione E, Kinde B, Place EM, et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet Med.* 2020;22:1079–1087.
- Zare F, Dow M, Monteleone N, Hosny A, Nabavi S. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinformatics.* 2017;18:286.
- Hong CS, Singh LN, Mullikin JC, Biesecker LG. Assessing the reproducibility of exome copy number variations predictions. *Genome Med.* 2016;8:82.
- Yao R, Zhang C, Yu T, et al. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. *Mol Cytogenet.* 2017;10:30.
- Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics.* 2012;28(10):1307–1313.
- Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;21:974–984.
- Zhang L, Bai W, Yuan N, Du Z. Comprehensively benchmarking applications for detecting copy number variation. *PLoS Comput Biol.* 2019;15:e1007069.
- Jiang Y, Oldridge DA, Diskin SJ, Zhang NR. CODEX: a normalization and copy number variation detection method for whole exome sequencing. *Nucleic Acids Res.* 2015;43:e39.
- Skidmore ZL, Wagner AH, Lesurf R, et al. GenVisR: Genomic Visualizations in R. *Bioinformatics.* 2016;32(19):3012–3014.
- Collins RL, Stone MR, Brand H, Glessner JT, Talkowski ME. CNView: a visualization and annotation tool for copy number variation from whole-genome sequencing. *bioRxiv.* 2016.
- Dharanipragada P, Vogeti S, Parekh N. iCopyDAV: integrated platform for copy number variations-Detection, annotation and visualization. *PLoS One.* 2018;13:e0195334.
- Zhang Y, Yu Z, Ban R, et al. DeAnnCNV: a tool for online detection and annotation of copy number variations from whole-exome sequencing data. *Nucleic Acids Res.* 2015;43(W1):W289–W294.
- Garvin T, Aboukhalil R, Kendall J, et al. Interactive analysis and assessment of single-cell copy-number variations. *Nat Methods.* 2015;12(11):1058–1060.
- MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research.* 2014;42(D1):D986–D992.
- Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research.* 2014;42(D1):D980–D985.
- Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015;43(Database issue):D789–D798.
- Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70.
- Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Research.* 2002;12:996–1006.
- HUGO Gene Nomenclature Committee at the European Bioinformatics Institute (HGNC). <https://www.genenames.org>. Accessed January 25, 2020.
- Buels R, Yao E, Diesh CM, et al. JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 2016;17:66.
- Amarasinghe KC, Li J, Hunter SM, et al. Inferring copy number and genotype in tumour exome data. *BMC Genomics.* 2014;15:732.
- Klambauer G, Schwarzbauer K, Mayr A, et al. cn.MOPS: mixture of poisson for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* 2012;40:e69.
- Sathirapongsasuti JF, Lee H, Horst BAJ, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: exomeCNV. *Bioinformatics.* 2011;27(19):2648–2654.
- Reble E, Castellani CA, Melka MG, O'Reilly R, Singh SM. VarScan2 analysis of de novo variants in monozygotic twins discordant for schizophrenia. *Psychiatr Genet.* 2017;27:62–70.
- GeneCards. Breast cancer related genes - Search Results. <https://www.genecards.org/Search/Keyword?queryString=breast%20cancer>. Accessed March 13, 2020.
- Leppa VM, Kravitz SN, Martin CL, et al. Rare Inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. *Am J Hum Genet.* 2016;99:540–554.