

RESEARCH ARTICLE

QuIN: A Web Server for Querying and Visualizing Chromatin Interaction Networks

Asa Thibodeau¹, Eladio J. Márquez², Oscar Luo², Yijun Ruan², Francesca Menghi², Dong-Guk Shin¹, Michael L. Stitzel^{2,3}, Paola Vera-Licona^{3,4}, Duygu Ucar^{2,3*}

1 Department of Computer Science, University of Connecticut, Storrs, Connecticut, United States of America, **2** The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, United States of America, **3** Institute of Systems Genomics, University of Connecticut Health Center, Farmington, Connecticut, United States of America, **4** Center for Quantitative Medicine, Department of Cell Biology, University of Connecticut Health Center, Farmington, Connecticut, United States of America

* Duygu.ucar@jax.org



Abstract

Recent studies of the human genome have indicated that regulatory elements (e.g. promoters and enhancers) at distal genomic locations can interact with each other via chromatin folding and affect gene expression levels. Genomic technologies for mapping interactions between DNA regions, e.g., ChIA-PET and HiC, can generate genome-wide maps of interactions between regulatory elements. These interaction datasets are important resources to infer distal gene targets of non-coding regulatory elements and to facilitate prioritization of critical loci for important cellular functions. With the increasing diversity and complexity of genomic information and public ontologies, making sense of these datasets demands integrative and easy-to-use software tools. Moreover, network representation of chromatin interaction maps enables effective data visualization, integration, and mining. Currently, there is no software that can take full advantage of network theory approaches for the analysis of chromatin interaction datasets. To fill this gap, we developed a web-based application, QuIN, which enables: 1) building and visualizing chromatin interaction networks, 2) annotating networks with user-provided private and publicly available functional genomics and interaction datasets, 3) querying network components based on gene name or chromosome location, and 4) utilizing network based measures to identify and prioritize critical regulatory targets and their direct and indirect interactions. **AVAILABILITY:** QuIN's web server is available at <http://quin.jax.org> QuIN is developed in Java and JavaScript, utilizing an Apache Tomcat web server and MySQL database and the source code is available under the GPLv3 license available on GitHub: <https://github.com/UcarLab/QuIN/>.

OPEN ACCESS

Citation: Thibodeau A, Márquez EJ, Luo O, Ruan Y, Menghi F, Shin D-G, et al. (2016) QuIN: A Web Server for Querying and Visualizing Chromatin Interaction Networks. *PLoS Comput Biol* 12(6): e1004809. doi:10.1371/journal.pcbi.1004809

Editor: Paul P Gardner, University of Canterbury, NEW ZEALAND

Received: February 5, 2016

Accepted: May 12, 2016

Published: June 23, 2016

Copyright: © 2016 Thibodeau et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data (or accession numbers to public data sources) are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

This is a *PLOS Computational Biology* Software paper.

Introduction

Chromatin structure plays a major role in basic cellular functions. Advances in genomic technologies have revealed information regarding three-dimensional (3D) chromatin conformation and have shown that many regulatory elements that are distal on the linear genome map are actually in close physical proximity with each other as a result of the 3D chromatin structure. Current technologies for capturing this 3D structure include Chromosome Conformation Capture based methods (3C) [1], 4C [2], 5C [3], Hi-C [4] and Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) [5]. These technologies identify chromatin interactions between promoters, enhancers, and other regulatory elements. The data generated by these technologies are the starting point from which we can infer distal regulatory interactions and their system-level effects by modeling them in the form of interaction networks. Moreover, integrating interaction datasets with additional data types and public repositories facilitate the discovery of regulatory elements and interactions that are critical for cellular functions and for disease biology, such as gene targets of non-coding regulatory elements harboring disease-causing Single Nucleotide Polymorphisms (SNPs). It is therefore imperative to have an easy-to-use software platform that enables biologists to model and study their chromatin interaction datasets under the light of other data sources, such as SNP databases and epigenetic marks.

Chromatin interaction data are typically visualized using a genome browser in a linear fashion, providing one-dimensional representation of the data (see [S1 Fig](#) for an example). A commonly used tool for this is the UCSC Genome Browser [6]. However, two and three dimensional representations of chromatin interactions in the form of networks and three-dimensional models can provide a global view of the interactions and facilitate the use of established network analysis methods and measures on these datasets [7]. For example, network representation of RNA Pol2 ChIA-PET data revealed that loci harboring disease-associated SNPs are differently connected in chromatin interaction networks [8]. Available tools for analyzing interaction data currently are unable to take advantage of a network representation; these tools include i) HOMER's HiC analysis suite [9], which enables filtering and calling interactions, as well as testing for the significance of the frequency of interactions between annotations using SIMA [10], ii) HiBrowse [11], which integrates regions of interest with 3D co-localized sites; iii) GWAS3D [12], which integrates user uploaded SNPs with interaction data from ENCODE [13] to identify loci interacting with these SNPs; and iv) GenomicInteractions [14], an R package, which allows users to annotate anchors of interactions and produce summaries of the interaction data. An existing tool for building and analyzing chromatin interaction networks is CytoHiC [15], a Cytoscape [16] plugin, tailored for HiC data. However, CytoHiC is limited in annotation and query capabilities and does not enable integrating interactions with other databases through the plugin itself.

To the best of our knowledge, there is no publicly available software for the analysis of chromatin interaction networks that is web-accessible and easy to use, making it suitable for molecular biologists with no programming experience to use. To overcome the limitations of current tools, we developed a single platform for **Querying** and visualizing Chromatin Interaction Networks (QuIN) (<http://quin.jax.org>) ([Fig 1](#)). QuIN enables: 1) building and visualizing chromatin interaction networks from ChIA-PET or HiC interactions; 2) annotating these networks with functional information from epigenetic datasets, SNPs, gene definitions, Gene Ontology terms, other interaction networks etc.; 3) querying network components for specific genes, loci, or disease-causing SNPs; and 4) utilizing network-based algorithms and measures to prioritize genomic sites for functional validation. QuIN mines chromatin interaction datasets such as those generated by the ENCODE consortium [13] (or user supplied ones) and integrates these datasets with other functional information such as chromatin states that can be inferred from

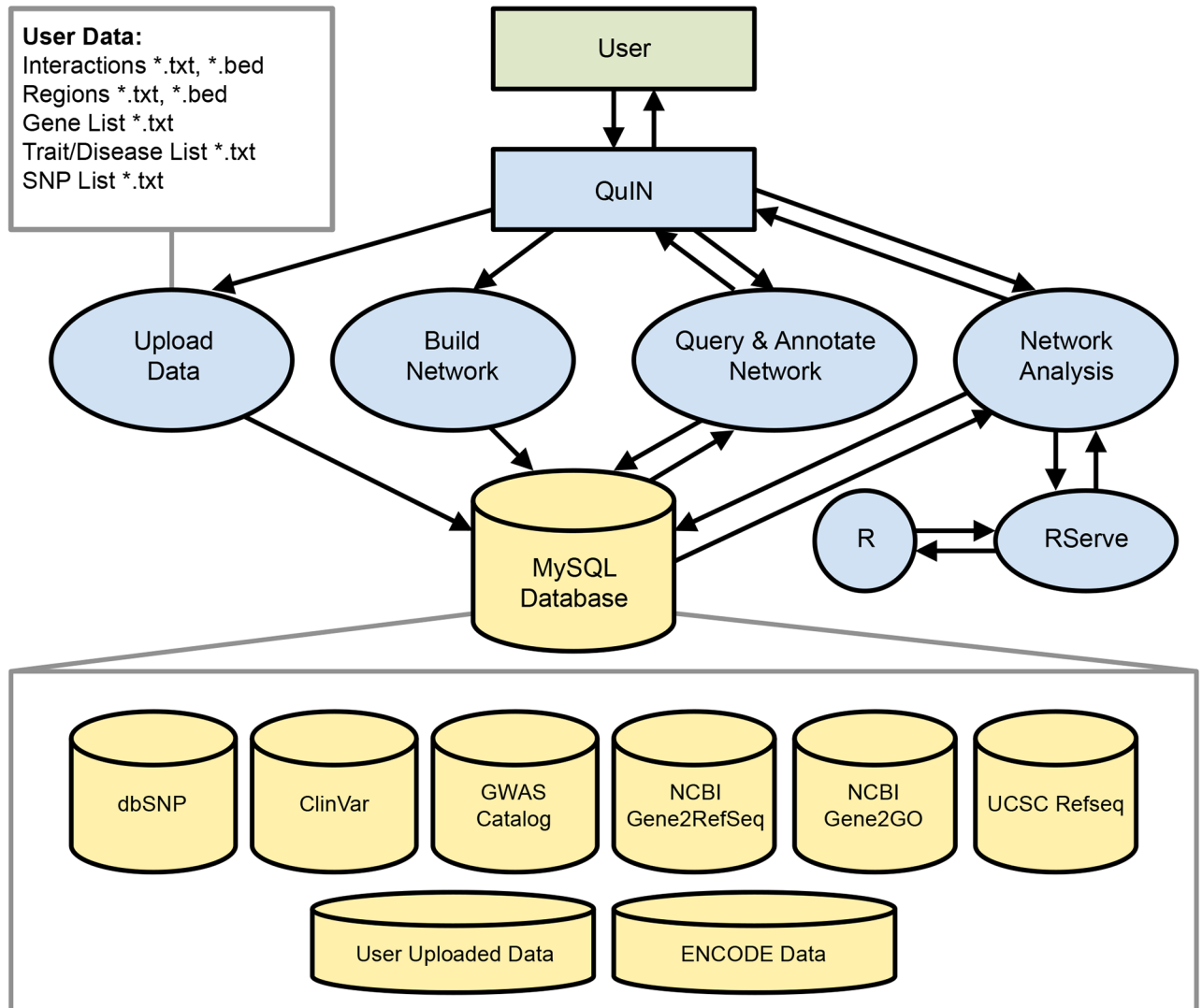


Fig 1. Data flow diagram of QuIN. QuIN allows users to upload diverse data types and formats and it enables building, querying, annotating, and analyzing chromatin interaction networks. QuIN also integrates publically available databases for network annotation and enrichment.

doi:10.1371/journal.pcbi.1004809.g001

histone modification datasets or SNP databases. We summarize the features of existing tools and QuIN in [S2 Fig](#). In summary, QuIN is designed to enable biologists to easily represent and annotate their chromatin-interaction datasets in the form of networks and to use these datasets for discovering important interactions or targets.

Design and Implementation

Software and Databases

QuIN is implemented in Java and JavaScript, with a MySQL database backend for storing user supplied data and local instances of the publicly available data including ClinVar [17], dbSNP [18], GWAS Catalog [19], NCBI's Gene2Refseq and Gene2GO databases[20] and UCSC's hg19 RefSeq database [21]. The graphical user interface is developed using Cytoscape JS (<http://js.cytoscape.org/>) for network visualization and JQuery/JQuery UI (<https://jquery.com/>)

) for other user interface elements. For network and other analyses, the web server communicates with R using RServe (<https://rforge.net/Rserve/>) to utilize existing R packages including topGO, vioplot and pheatmap to generate GO analyses, violin plots and heatmaps. Data are uploaded anonymously and privately, utilizing cookies to store the necessary information for linking the user's browser with their uploaded data. [Fig 1](#) summarizes the data flow and architecture of QuIN.

Network Construction from Interaction Data

QuIN implements two approaches for constructing networks from chromatin interactions obtained via pre-processing tools such as ChIA-PET tool [22] or HOMER [9]. The first approach is based on using the interaction data alone to define network nodes by merging overlapping interaction anchors, whereas the second approach incorporates additional data, such as open chromatin sites from DNASE-Seq [23] or ATAC-Seq [24] technologies, to define or refine the nodes in the network, which is useful to filter out potential false positive interactions. Once nodes are defined, pairs of nodes are connected by identifying chromatin interactions whose anchors overlap the boundaries of these nodes; these interactions thus define the edges of the network. QuIN keeps track of the number of interactions that connect node pairs as edge weights. Finally, a breadth first search method is carried out to isolate the connected components of the network. A more detailed description of the algorithms implemented for network construction can be found in [S1 Text](#).

Annotating and Analyzing Networks

Users can annotate interaction networks by uploading various types of datasets including genomic regions, genes, SNPs, and diseases/traits listed in the GWAS catalog [19]. Once annotated, QuIN offers a diverse set of measures and algorithms for further analyses of the network. Previous studies have shown that network characteristics of chromatin interactions can be linked to regulatory or three-dimensional structural properties of genes and gene products [8,25]. For example, we have shown that hubs (nodes with high degree) and spokes (nodes with low degree) exhibit distinct functional and etiological properties [8]. Among these, we noted that the hubs lacked disease-associated SNPs, suggesting an evolutionary selection that favors the nonrandom spatial clustering of the genomic domains. Currently, implemented network measures within QuIN include: (1) connectivity degree which measures the number of edges connected to a node, quantifying the number of interactions to one particular loci; (2) betweenness, measuring the number of times a node exists on the shortest path between all other pairs of nodes, suggesting loci which may be essential for others sites to interact; and (3) closeness & harmonic centrality, which measure the average shortest paths to all other nodes in the component or network, suggesting loci which are more central and are interacting with different locus. To further understand the connectivity between annotations and to utilize network topology for biological discovery, QuIN also provides a target discovery function, which allows users to export all of the shortest paths from source nodes to target nodes, collecting all direct and indirect interactions between two sets of annotated nodes. Additionally, QuIN provides methods for analyzing the frequency of interactions between pairs of annotations to reveal whether interactions between nodes annotated with certain features (e.g. nodes harboring SNPs vs. nodes overlapping with enhancers) are enriched relative to the random expectation. QuIN provides two options for calculating expected frequencies between two annotations: (1) by permutation to randomize the network annotations, and (2) by calculating a theoretical expected frequency. Significance of an observed frequency is calculated using (1) one-tailed binomial test p-values and (2) p-values with respect to the null distribution derived

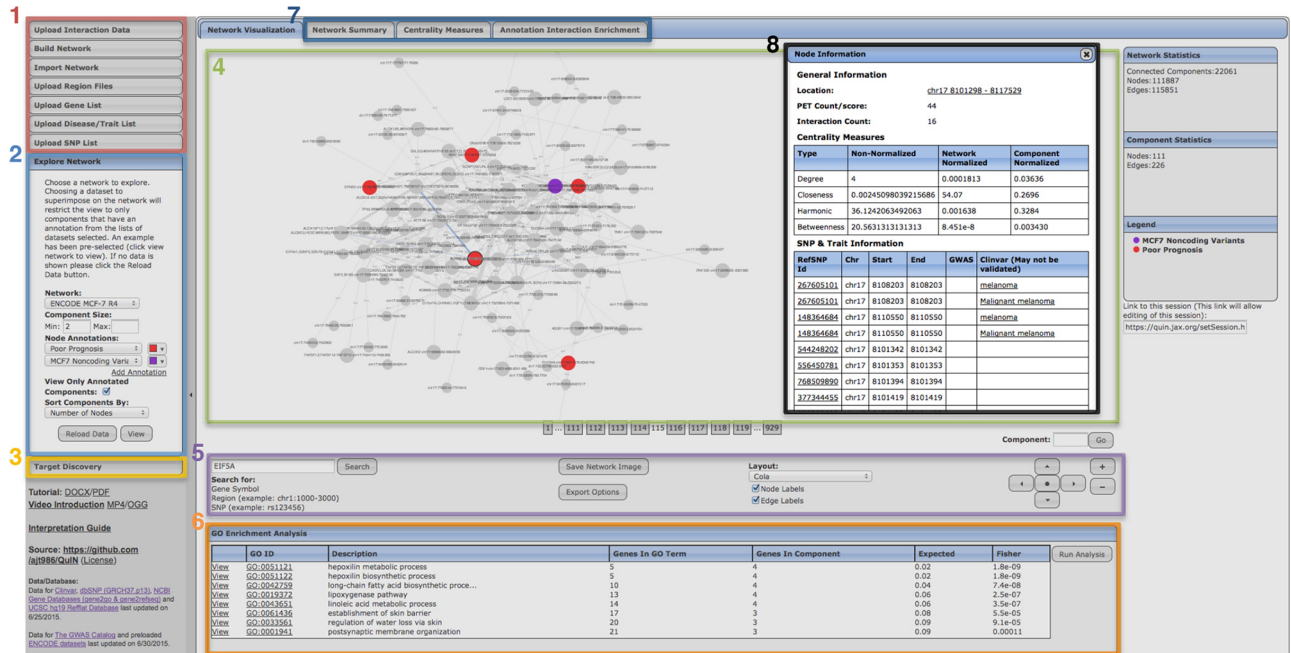


Fig 2. A screenshot of QuIN's web interface highlighting its features. (1) menus for uploading data and building networks, (2) options for visualizing and annotating a network, (3) target discovery menu for visualizing and exporting direct and indirect targets from source annotations to target annotations (4) network visualization panel, (5) options for searching, querying, or exporting the network, (6) the menu for performing GO Enrichment Analysis on the current subnetwork, (7) tools for summarizing network construction statistics, centrality measures and enrichment of interactions between annotations, (8) dialog box showing additional information about a selected node, including centrality measures, SNPs, and associated diseases.

doi:10.1371/journal.pcbi.1004809.g002

from permutations. Further details on statistical tests are provided in [S1 Text](#). Finally, to enable using other existing methods for network analyses, the chromatin interaction network can be exported in standard Graph Modelling Language (.gml) format, allowing the network generated by QuIN to be imported into existing tools that support this format, such as Cytoscape [16]. A screenshot of the web based graphical user interface summarizing QuIN's features is shown in [Fig 2](#).

Results

A case study to discover gene targets of non-coding variants in breast cancer cell line MCF-7

We demonstrate the core functionalities of QuIN by conducting a case study using RNA Pol2 ChIA-PET data in the breast cancer cell line MCF-7. For this analysis, we constructed an interaction network in MCF-7 using ChIA-PET (GSM970209) and DNASE-Seq (GSM816627) data generated by the ENCODE consortium [13]. This resulted in an MCF-7 network comprising 59,082 nodes, 65,308 edges, and 8,133 connected components. In the absence of chromatin interaction datasets, variants are associated to the closest gene transcription start site (TSS). However, increasing evidence points out to the importance of noncoding variants in disease biology [26] and the frequency of distal interactions between genes and their non-coding regulatory elements. Hence, we sought to use the MCF-7 interaction network to discover gene targets of non-coding variants that are likely to be associated with breast cancer. QuIN can be employed to conduct similar analyses using coding variants. However, coding variants have a clearer phenotypic outcome: amino-acid change in the corresponding gene sequence and a

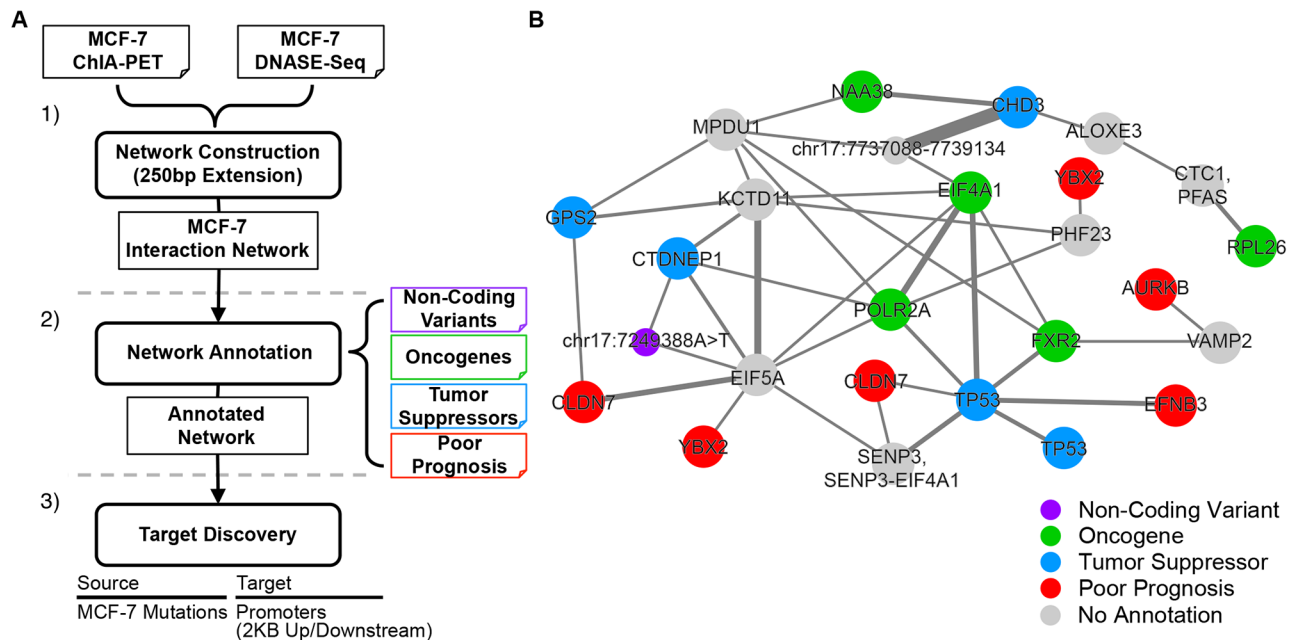


Fig 3. A breast cancer case study with QuIN. (A) Workflow of the case study analysis. (1) Upload the DNASE-Seq and Interaction data into QuIN, constructing an MCF-7 interaction network where each node represents an open chromatin site. (2) Annotate the network with Non-Coding Variants (NCVs) in MCF-7 and cancer associated gene lists. (3) Perform target discovery between NCVs (source) and promoters and cancer gene lists (targets) and find all direct and indirect associations between NCVs and their gene targets. (B) A simplified network example showing the interactions between a node harboring an NCV (shown in purple) and known oncogenes (green), genes associated with poor prognosis in breast cancer (red), and tumor suppressor genes (blue). Nodes shown were selected based on their overlap with an annotation or if the node is necessary to connect the NCV to the annotated node. Width of the edges correspond to the relative number of paired end tags supporting the edge.

doi:10.1371/journal.pcbi.1004809.g003

possible effect on that protein’s function or stability. Therefore, the primary impact of a coding variant is expected to be local and just restricted to the gene (and the protein) harboring the variant. Workflow of our case study is explained in Fig 3A with an example annotated network shown in Fig 3B.

For this purpose, we first annotated the MCF-7 interaction network with non-coding variants (NCV) found in the MCF-7 cell line obtained from the COSMIC database [27] (cancer.sanger.ac.uk), which catalogues somatic mutations in cancer, and identified 36 NCVs that overlap with nodes in the network. To identify distal and proximal gene targets of these potentially disease-causing variants in the MCF-7 network, we used the target discovery feature of QuIN and identified all shortest paths from nodes harboring NCVs to nodes associated with promoters (defined as 2kb up/downstream of TSS using RefSeq gene definitions). We identified the ‘direct targets’ of these NCVs by finding the genes that are either within the same node as the variant or one edge away in the network. We also captured ‘indirect targets’ by further expanding the search and finding genes that are 2–4 edges away from the NCV containing node, which we determined by studying the enrichment of cancer-related genes among targets identified via different edge distances (S3 Fig) (see S1 Text for details and Fig 4A for definitions). Recent studies have shown that indirect interactions are useful for predicting gene co-expression patterns [28] and identifying complex regulatory interactions such as enhancer-enhancer-promoter interactions [29], pointing out to the significance of identifying and studying indirect chromatin interactions through a network-based approach. In summary, our analyses revealed 90 genes that are ‘direct targets’ and 638 genes that are ‘indirect targets’ of the NCVs in the MCF-7 network. (S1 Table).

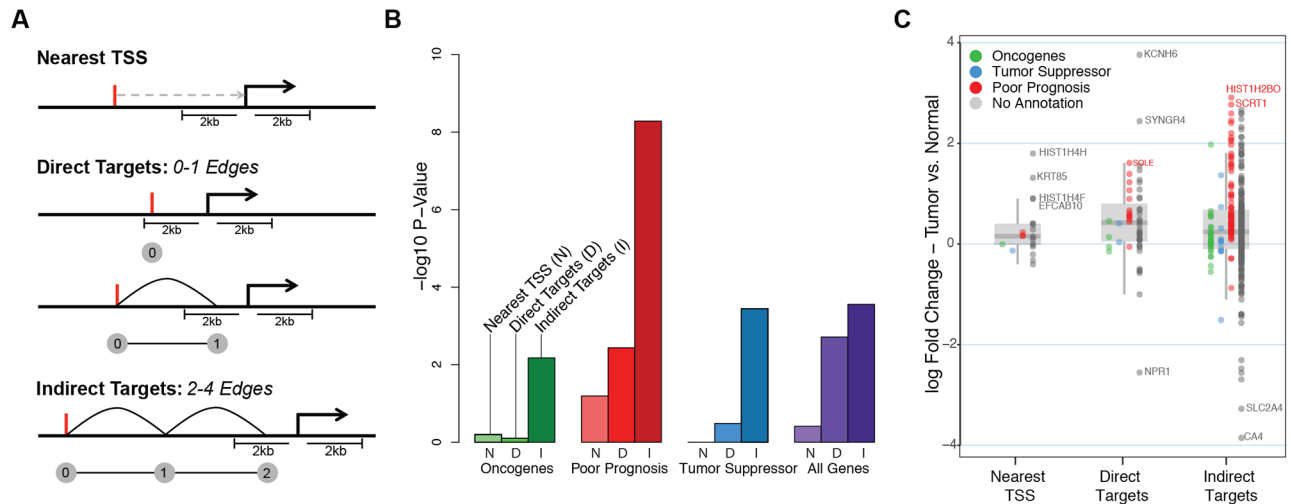


Fig 4. Comparison of ChIA-PET gene targets with nearest gene targets. (A) A cartoon describing different approaches to associate NCVs with gene targets (Nearest TSS, Direct Targets, & Indirect Targets). (B) Enrichment p-values (based on Fisher’s exact test) of cancer related genes (known oncogenes (green), tumor suppressor genes (blue), poor prognosis genes (red), and the combined gene list (purple)) among NCV gene targets obtained via nearest TSS, direct target, indirect target associations. (C) Boxplot showing the differential expression (between cancer and normal tissues) for NCV target genes obtained via nearest TSS, direct target, indirect target associations.

doi:10.1371/journal.pcbi.1004809.g004

To understand whether gene targets obtained from ChIA-PET interaction networks can be useful in discovering genes with potential implications in breast cancer, we compared the gene lists obtained from MCF-7 interaction networks (both direct and indirect targets) with the gene list obtained via ‘nearest TSS’ annotation (Fig 4A). Using the nearest TSS approach, we identified 30 genes, which we compared against direct (90) and indirect targets (638) obtained from the ChIA-PET network analyses. First, we annotated our MCF-7 network with cancer associated genes including known oncogenes [30,31,32], tumor suppressor genes identified by [33], genes previously associated with poor prognosis in breast cancer patients [34], and a unified list of cancer related genes including all previous lists and other gene lists [27,35,36] (further details of these lists are provided in S1 Text).

Next we calculated the enrichment of known breast cancer-related genes in targets discovered by i) nearest TSS; ii) direct ChIA-PET targets; and iii) indirect ChIA-PET targets using Fisher’s exact test. Our analyses showed that ChIA-PET based targets are significantly more enriched for cancer-related genes than nearest gene targets (Fig 4B). For example, for genes associated with poor prognosis in breast cancer, we have shown that the enrichment Fisher’s p-value for direct targets is 3.65×10^{-3} , indirect targets is 5.18×10^{-9} , whereas nearest gene targets is 6.39×10^{-2} (Fig 4B and S4 Fig, S2 Table). This analysis also revealed that with the help of chromatin interaction networks and our tool QuIN, we can capture gene targets of NCVs that are more likely to be relevant for the disease than nearest TSS assignments, even if they are separated via multiple edges in the network. To further assess their relevance to breast cancer, we studied these three lists of genes by comparing their expression levels among TCGA samples [37]. For this, we first calculated the differential expression of genes between breast cancer samples and normal breast epithelium samples using TCGA RNA-seq data. We found that targets discovered by ChIA-PET include genes that are more differentially expressed between cancer and normal tissues in comparison to nearest gene targets (Fig 4C). Moreover, we have also observed that even indirect targets of the NCVs could be disease relevant, which highlights the system-level impact of disease-causing variants and the importance of studying these interactions at the network level. Similarly, we observed that gene targets

discovered using ChIA-PET tend to show more differential expression between the MCF-7 line and MCF-10A, a non-cancer mammary epithelial cell line, using both single-cell and bulk RNA-seq datasets [38] (GSE52712) (S5 Fig). As an example, Fig 3B and S6 Fig show a simplified subnetwork revealing the indirect relationships of a region harboring an NCV to genes associated with cancer, including the well-known tumor repressor *TP53* and multiple breast cancer-associated genes, such as *EIF4A* [39], *EIF5A* [40,41], *AURKB* [42] and *CLDN7* [43]. Furthermore, we show linear and network representations of the chromatin interactions in an example locus along with the gene expression fold changes for selected genes (S1 Fig using TCGA and S7 Fig using MCF-7 and MCF-10A gene expression datasets). Our analyses revealed the potential impact of NCVs on cancer-related gene regulatory programs through their chromatin connectivity.

To further interrogate the regulatory implications of ChIA-PET interactions, we calculated co-expression of gene pairs using 914 breast cancer samples from the TCGA data. Our results showed that direct interactions overall have the highest co-expression values. Yet, indirect interactions have significantly higher co-expression levels when compared to non-interacting gene pairs, implying the co-regulation of genes connected via direct or indirect chromatin interactions (S8 Fig). In addition, we integrated MCF-7 network with protein-protein interactions obtained from the STRING database [44]. After super-imposing the protein-protein interactions onto the MCF-7 network using QuIN, we identified genes exhibiting indirect chromatin interactions that also have direct protein-protein interactions (S9 Fig). One such example is *TP53* and *AURKB*, which show a direct protein-protein interaction while indirectly interacting in the chromatin interaction network. This further increases our evidence of the importance of indirect interactions when analyzing chromatin interaction data, which could be relevant in multiple biological contexts.

This case study demonstrates how researchers can use QuIN to easily interrogate chromatin interaction datasets in conjunction with other data sources, such as variant catalogs, to identify and prioritize candidate regulatory elements relevant for normal and abnormal cell function. It also illustrates the power of network-based approaches to facilitate data-driven selection of gene targets of disease-associated mutations and polymorphisms, in comparison to the traditional selection of gene targets based on linear genomic distance, which does not take into consideration the three-dimensional nature of regulatory interactions in the genome.

Availability and Future Directions

To the best of our knowledge, QuIN is the first web accessible and easy to use platform for analyzing 3D chromatin interaction datasets and we plan to extend its capabilities in multiple areas. First, we plan to implement algorithms for mining network motifs that might be important for biological functions, taking further advantage of network theory approaches. Second, we intend to provide more genomes beyond hg19 to support a broader range of species and experiments. Future extensions include expanding QuIN's data integration capabilities with the addition of expression data, including their visualization in the context of the chromatin interaction network. Additionally, as chromatin interaction datasets become abundant, it will be important to implement features for comparing networks in terms of interaction losses and gains, which will allow users to infer potential regulatory changes resulting from these differences.

QuIN is an open source project released under the GNU General Public License Version 3 and is available on GitHub (<https://github.com/UcarLab/QuIN/>) and in S1 software. In addition, we have an online web server (<http://quin.jax.org>) providing a publicly available working version of our software.

Supporting Information

S1 Fig. An example loci using linear and three-dimensional representation with TCGA gene expression data. (A) BASIC browser screenshot of the region corresponding to the network example in [Fig 3B](#) and [S6 Fig](#). Genes shown have been selected based on representation within the network. (B) Network representation of the same region with nodes aligned based on overlapping markers, genes, and interactions. (C) Tumor vs. Normal fold change of TCGA expression data for selected genes found within the network. The TSS nearest to the non-coding variant, *KCTD11*, has been highlighted in red.

(TIF)

S2 Fig. Table summarizing features of six computational tools developed to analyze chromatin interaction datasets. Of the tools reviewed, only QuIN, CytoHiC and GenomicInteractions have provided methods for interpreting chromatin interactions as a network, with GenomicInteractions only providing the ability to construct a network by using other R packages. Considering the accessibility of the tools, QuIN, HiBrowse, and GWAS3D have been developed as web-based applications, allowing immediate access to the tools through web browsers while eliminating the steps of installing and setting up the software before use. Though a majority of the tools support the ability for users to analyze their own interaction datasets, GWAS3D alternatively only allows users to upload SNPs of interest to analyze with chromatin interaction data available with the tool. All of the tools have shown some integration with public data/databases with varying levels of comprehensiveness, however CytoHiC requires other Cytoscape plugins to achieve this functionality and HOMER's Hi-C suite fulfills this by being part of a larger package, providing similar integration by pipelining the data with other HOMER command-line tools. For analyzing the frequency of interactions between annotations, only QuIN, HiBrowse, and HOMER's Hi-C suite provide methods for evaluating the significance of these frequencies. Finally, QuIN proves to be the only tool that offers the ability to systematically discover both direct and indirect targets, taking advantage of the network representation to determine indirect targets of a node of interest.

(TIF)

S3 Fig. P-Values of enrichment of cancer gene lists in NCV target genes (Fisher's exact test) at various indirect edge cutoffs.

(TIF)

S4 Fig. Enrichment analyses of gene targets. Enrichment p-values (based on Fisher's exact test) of cancer-related genes (known oncogenes (green), known tumor suppressor genes (yellow), genes associated with good (orange) and poor prognosis (red), oncogenes and tumor suppressor genes identified by Davoli et al (2013) (pink and blue)), and the union of all cancer related genes (purple) in NCV gene targets obtained via nearest tss, direct targets, indirect targets, direct and indirect target methods.

(TIF)

S5 Fig. Differential gene expression analyses for target genes. Boxplots showing the differential expression (between cancer and normal tissues) for NCV target genes obtained via nearest TSS, direct target, indirect target associations for (A) MCF-7 vs. MCF-10A Single Cell and (B) MCF-7 vs. MCF-10A Bulk samples.

(TIF)

S6 Fig. An example network. Network image generated and saved using QuIN corresponding to [Fig 3B](#), displaying exact position of nodes as well as all promoters overlapping the nodes.

Values of edges indicate the number of paired end tags and the total number of interactions in parenthesis.

(TIF)

S7 Fig. An example locus using linear and three-dimensional representation with MCF-7 and MCF-10A gene expression data. (A) BASIC browser screenshot of the region corresponding to the network example in [Fig 3B](#) and [S6 Fig](#). Genes shown have been selected based on representation within the network. (B) Network representation of the same region with nodes aligned based on overlapping markers, genes, and interactions. (C) MCF-7 vs. MCF-10A fold change for selected genes found within the network. The nearest TSS of the non-coding variant, KCTD11, has been highlighted in red.

(TIF)

S8 Fig. Direct and indirect interacting gene pairs exhibit elevated gene expression correlations. Gene expression correlations between gene pairs connected via direct and indirect interactions in the MCF-7 ChIA-PET network, compared to correlations between not interacting genes.

(TIF)

S9 Fig. QuIN enables superimposing other types of interactions with chromatin interactions. An example list of interactions from the STRING database (A) superimposed on a ChIA-PET chromatin interaction subnetwork (B).

(TIF)

S1 Text. Supplementary Methods.

(PDF)

S1 Table. Target discovery output for the case study.

(CSV)

S2 Table. Cancer gene hits among gene target lists.

(CSV)

S1 Software. Archive for QuIN's source code.

(7Z)

Author Contributions

Conceived and designed the experiments: AT DGS DU. Performed the experiments: AT EJM. Analyzed the data: AT EJM FM. Contributed reagents/materials/analysis tools: OL YR MLS PVL. Wrote the paper: AT EJM FM PVL DU.

References

1. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295: 1306–1311. PMID: [11847345](#)
2. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, et al. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38: 1348–1354. PMID: [17033623](#)
3. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, et al. (2006) Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16: 1299–1309. PMID: [16954542](#)
4. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–293. doi: [10.1126/science.1181369](#) PMID: [19815776](#)

5. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462: 58–64. doi: [10.1038/nature08497](https://doi.org/10.1038/nature08497) PMID: [19890323](https://pubmed.ncbi.nlm.nih.gov/19890323/)
6. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. *Genome Res* 12: 996–1006. PMID: [12045153](https://pubmed.ncbi.nlm.nih.gov/12045153/)
7. Singh Sandhu K, Li G, Sung WK, Ruan Y (2011) Chromatin interaction networks and higher order architectures of eukaryotic genomes. *J Cell Biochem* 112: 2218–2221. doi: [10.1002/jcb.23155](https://doi.org/10.1002/jcb.23155) PMID: [21520242](https://pubmed.ncbi.nlm.nih.gov/21520242/)
8. Sandhu KS, Li G, Poh HM, Quek YL, Sia YY, et al. (2012) Large-scale functional organization of long-range chromatin interaction networks. *Cell Rep* 2: 1207–1219. doi: [10.1016/j.celrep.2012.09.022](https://doi.org/10.1016/j.celrep.2012.09.022) PMID: [23103170](https://pubmed.ncbi.nlm.nih.gov/23103170/)
9. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589. doi: [10.1016/j.molcel.2010.05.004](https://doi.org/10.1016/j.molcel.2010.05.004) PMID: [20513432](https://pubmed.ncbi.nlm.nih.gov/20513432/)
10. Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, et al. (2012) Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology* 13: 1196–1204. doi: [10.1038/ni.2432](https://doi.org/10.1038/ni.2432) PMID: [23064439](https://pubmed.ncbi.nlm.nih.gov/23064439/)
11. Paulsen J, Sandve GK, Gundersen S, Lien TG, Trengereid K, et al. (2014) HiBrowse: multi-purpose statistical analysis of genome-wide chromatin 3D organization. *Bioinformatics* 30: 1620–1622. doi: [10.1093/bioinformatics/btu082](https://doi.org/10.1093/bioinformatics/btu082) PMID: [24511080](https://pubmed.ncbi.nlm.nih.gov/24511080/)
12. Li MJ, Wang LY, Xia Z, Sham PC, Wang J (2013) GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Res* 41: W150–158. doi: [10.1093/nar/gkt456](https://doi.org/10.1093/nar/gkt456) PMID: [23723249](https://pubmed.ncbi.nlm.nih.gov/23723249/)
13. Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247) PMID: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
14. Hamston N, Ing-Simmons E, Perry M, Baresic A, Lenhard B (2015) GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data. *BMC Genomics* 16: 963. doi: [10.1186/s12864-015-2140-x](https://doi.org/10.1186/s12864-015-2140-x) PMID: [26576536](https://pubmed.ncbi.nlm.nih.gov/26576536/)
15. Shavit Y, Lio P (2013) CytoHiC: a cytoscape plugin for visual comparison of Hi-C networks. *Bioinformatics* 29: 1206–1207. doi: [10.1093/bioinformatics/btt120](https://doi.org/10.1093/bioinformatics/btt120) PMID: [23508968](https://pubmed.ncbi.nlm.nih.gov/23508968/)
16. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504. PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
17. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, et al. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42: D980–985. doi: [10.1093/nar/gkt1113](https://doi.org/10.1093/nar/gkt1113) PMID: [24234437](https://pubmed.ncbi.nlm.nih.gov/24234437/)
18. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29: 308–311. PMID: [11125122](https://pubmed.ncbi.nlm.nih.gov/11125122/)
19. Welter D, MacArthur J, Morales J, Burdett T, Hall P, et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 42: D1001–1006. doi: [10.1093/nar/gkt1229](https://doi.org/10.1093/nar/gkt1229) PMID: [24316577](https://pubmed.ncbi.nlm.nih.gov/24316577/)
20. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, et al. (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res* 43: D36–42. doi: [10.1093/nar/gku1055](https://doi.org/10.1093/nar/gku1055) PMID: [25355515](https://pubmed.ncbi.nlm.nih.gov/25355515/)
21. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, et al. (2015) The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 43: D670–681. doi: [10.1093/nar/gku1177](https://doi.org/10.1093/nar/gku1177) PMID: [25428374](https://pubmed.ncbi.nlm.nih.gov/25428374/)
22. Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, et al. (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol* 11: R22. doi: [10.1186/gb-2010-11-2-r22](https://doi.org/10.1186/gb-2010-11-2-r22) PMID: [20181287](https://pubmed.ncbi.nlm.nih.gov/20181287/)
23. Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* 2010: pdb prot5384.
24. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10: 1213–1218. doi: [10.1038/nmeth.2688](https://doi.org/10.1038/nmeth.2688) PMID: [24097267](https://pubmed.ncbi.nlm.nih.gov/24097267/)
25. Boulos R, Arneodo A, Jensen P, Audit B (2013) Revealing long-range interconnected hubs in human chromatin interaction data using graph theory. *Physical review letters* 111: 118102. PMID: [24074120](https://pubmed.ncbi.nlm.nih.gov/24074120/)
26. Smith E, Shilatifard A (2014) Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* 21: 210–219. doi: [10.1038/nsmb.2784](https://doi.org/10.1038/nsmb.2784) PMID: [24599251](https://pubmed.ncbi.nlm.nih.gov/24599251/)

27. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, et al. (2015) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 43: D805–811. doi: [10.1093/nar/gku1075](https://doi.org/10.1093/nar/gku1075) PMID: [25355519](https://pubmed.ncbi.nlm.nih.gov/25355519/)
28. Babaei S, Mahfouz A, Hulsman M, Lelieveldt BP, de Ridder J, et al. (2015) Hi-C Chromatin Interaction Networks Predict Co-expression in the Mouse Cortex. *PLoS Comput Biol* 11: e1004221. doi: [10.1371/journal.pcbi.1004221](https://doi.org/10.1371/journal.pcbi.1004221) PMID: [25965262](https://pubmed.ncbi.nlm.nih.gov/25965262/)
29. Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, et al. (2013) Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155: 1507–1520. doi: [10.1016/j.cell.2013.11.039](https://doi.org/10.1016/j.cell.2013.11.039) PMID: [24360274](https://pubmed.ncbi.nlm.nih.gov/24360274/)
30. Higgins ME, Claremont M, Major JE, Sander C, Lash AE (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res* 35: D721–726. PMID: [17088289](https://pubmed.ncbi.nlm.nih.gov/17088289/)
31. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10: 59–64. doi: [10.1038/nrc2771](https://doi.org/10.1038/nrc2771) PMID: [20029424](https://pubmed.ncbi.nlm.nih.gov/20029424/)
32. Solimini NL, Xu Q, Mermel CH, Liang AC, Schlabach MR, et al. (2012) Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science* 337: 104–109. doi: [10.1126/science.1219580](https://doi.org/10.1126/science.1219580) PMID: [22628553](https://pubmed.ncbi.nlm.nih.gov/22628553/)
33. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, et al. (2013) Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 155: 948–962. doi: [10.1016/j.cell.2013.10.011](https://doi.org/10.1016/j.cell.2013.10.011) PMID: [24183448](https://pubmed.ncbi.nlm.nih.gov/24183448/)
34. Inaki K, Menghi F, Woo XY, Wagner JP, Jacques PE, et al. (2014) Systems consequences of amplicon formation in human breast cancer. *Genome Res* 24: 1559–1571. doi: [10.1101/gr.164871.113](https://doi.org/10.1101/gr.164871.113) PMID: [25186909](https://pubmed.ncbi.nlm.nih.gov/25186909/)
35. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177–183. PMID: [14993899](https://pubmed.ncbi.nlm.nih.gov/14993899/)
36. Zhao M, Sun J, Zhao Z (2013) TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res* 41: D970–976. doi: [10.1093/nar/gks937](https://doi.org/10.1093/nar/gks937) PMID: [23066107](https://pubmed.ncbi.nlm.nih.gov/23066107/)
37. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45: 1113–1120. doi: [10.1038/ng.2764](https://doi.org/10.1038/ng.2764) PMID: [24071849](https://pubmed.ncbi.nlm.nih.gov/24071849/)
38. Rothwell DG, Li Y, Ayub M, Tate C, Newton G, et al. (2014) Evaluation and validation of a robust single cell RNA-amplification protocol through transcriptional profiling of enriched lung cancer initiating cells. *BMC Genomics* 15: 1129. doi: [10.1186/1471-2164-15-1129](https://doi.org/10.1186/1471-2164-15-1129) PMID: [25519510](https://pubmed.ncbi.nlm.nih.gov/25519510/)
39. Modelska A, Turro E, Russell R, Beaton J, Sbarrato T, et al. (2015) The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. *Cell Death Dis* 6: e1603. doi: [10.1038/cddis.2014.542](https://doi.org/10.1038/cddis.2014.542) PMID: [25611378](https://pubmed.ncbi.nlm.nih.gov/25611378/)
40. Mathews MB, Hershey JW (2015) The translation factor eIF5A and human cancer. *Biochim Biophys Acta* 1849: 836–844. doi: [10.1016/j.bbagr.2015.05.002](https://doi.org/10.1016/j.bbagr.2015.05.002) PMID: [25979826](https://pubmed.ncbi.nlm.nih.gov/25979826/)
41. Muniz Lino MA, Palacios-Rodriguez Y, Rodriguez-Cuevas S, Bautista-Pina V, Marchat LA, et al. (2014) Comparative proteomic profiling of triple-negative breast cancer reveals that up-regulation of RhoGDI-2 is associated to the inhibition of caspase 3 and caspase 9. *J Proteomics* 111: 198–211. doi: [10.1016/j.jprot.2014.04.019](https://doi.org/10.1016/j.jprot.2014.04.019) PMID: [24768906](https://pubmed.ncbi.nlm.nih.gov/24768906/)
42. Gully CP, Zhang F, Chen J, Yeung JA, Velazquez-Torres G, et al. (2010) Antineoplastic effects of an Aurora B kinase inhibitor in breast cancer. *Mol Cancer* 9: 42. doi: [10.1186/1476-4598-9-42](https://doi.org/10.1186/1476-4598-9-42) PMID: [20175926](https://pubmed.ncbi.nlm.nih.gov/20175926/)
43. Hewitt KJ, Agarwal R, Morin PJ (2006) The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer* 6: 186. PMID: [16836752](https://pubmed.ncbi.nlm.nih.gov/16836752/)
44. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, et al. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43: D447–452. doi: [10.1093/nar/gku1003](https://doi.org/10.1093/nar/gku1003) PMID: [25352553](https://pubmed.ncbi.nlm.nih.gov/25352553/)