Genomics 112 (2020) 3696–3702

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

CancerEnD: A database of cancer associated enhancers

Rajesh Kumar[a], Anjali Lathwal[b], Vinod Kumar[a], Sumeet Patiyal[b], Pawan Kumar Raghav[b], Gajendra P.S. Raghava[b,*]

[a] Bioinformatics Centre, CSIR – Institute of Microbial Technology, Chandigarh, India
[b] Department of Computational Biology, Indraprastha Institute of Information Technology, New-Delhi, India

ARTICLE INFO

Keywords:
Enhancer
Cancer
Copy number variation
Survival
Expression

ABSTRACT

CancerEnD is an integrated resource developed for annotating 8524 unique expressed enhancers, associated genes, somatic mutations and copy number variations of 8063 cancer samples from 18 cancer types of TCGA. Somatic mutation data was taken from the COSMIC repository. To delineate the relationship of change in copy number of enhancer elements with the prognosis of cancer patients, survival analysis was done using the survival package in R. We identified 1762 overall survival associated enhancers, which can be used for prognostic purposes of cancer patients in a tissue-specific manner.

CancerEnD (https://webs.iiitd.edu.in/raghava/cancerend/) is developed on a user-friendly responsive template, that enables searching, browsing and downloading of the annotated enhancer elements in terms of gene expression, copy number variation and survival association.

We hope it provides a promising avenue for researchers to facilitate the understanding of enhancer deregulation in tumorigenesis, and to identify new biomarkers for therapy and disease-diagnosis.

1. Introduction

Cancer is a disease caused by genomic and epigenomic changes within the cell genome [1]. Research into cancer-causing aberrations has focused predominantly on the protein-coding part of the genome [2]. A major component of the cell genome is non-coding in nature which includes enhancers, promoters, insulators, and others. This non-coding part of the genome is involved in a wide variety of essential biological functions of a cell [3]. Particularly, the enhancers are one of the most important non-coding parts of the genome that functions as a master regulator of gene expression regulation in a tissue-specific manner [4].

Cancer studies present in literature predominantly focused on the mutation of the protein-coding part of the genome as compared to the non-coding part of the genome [5,6]. With the emergence of next-generation sequence data, it is a very commonly known fact that aberrations within the non-coding part of the genome are the primary driver of the vast majority of human diseases including cancer [7,8]. The non-coding genome is substantially larger than the protein-coding counterpart, thus pose problems in identifying and interpreting causality associated with the non-coding genome. However, if the full extent of the opportunity to use genomic information in healthcare is to be realized, it is necessary that consequences of changes within the non-coding genome are to be integrated along with the protein-coding genome [9]. To this extent, in recent years several projects like ENCODE [10], FANTOM [11], HaploReg [12], rSNPBase [13] have generated datasets that contributed to our understanding of the non-coding part of the human genome. More recently, some computational tools like Funseq [14], GAWVA [15] have been developed that can be used to prioritize and interrogate the cis-regulatory potential of non-coding variants in the human genome.

While the above-mentioned databases and tools can be used for the

Abbreviations: CNV, Copy Number Variation; TCGA, The Cancer Genome Atlas; COSMIC, Catalog of Somatic Mutation in Cancer; QTL, Quantitative Trait Loci; GISTIC, Genomic Identification of Significant Targets in Cancer; SQL, Structured Query Language; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast Invasive Carcinoma; CESC, Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma; ESCA, Esophageal Carcinoma; GBM, GlioblastomaMultiforme; HNSC, Head and Neck Squamous Cell Carcinoma; KIRC, Kidney Renal Clear Cell Carcinoma; KIRP, Kidney Renal Papillary Cell Carcinoma; LIHC, Liver Hepatocellular Carcinoma; LUAD, Lung Adenocarcinoma; LUSC, Lung Squamous Cell Carcinoma; PAAD, Pancreatic Adenocarcinoma; PRAD, Prostate Adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach Adenocarcinoma; THCA, Thyroid Carcinoma

* Corresponding author at: Department of Computational Biology, Indraprastha Institute of Information Technology, Delhi, Okhla Industrial Estate, Phase III, (Near GovindPuri Metro Station), 110020, Office: A-302 (R&D Block), New Delhi, India.

E-mail addresses: b.rajesh@imtech.res.in (R. Kumar), anjalil@iiitd.ac.in (A. Lathwal), vinodporiya@imtech.res.in (V. Kumar), sumeetp@iiitd.ac.in (S. Patiyal), pwnrghv@gmail.com (P.K. Raghav), raghava@iiitd.ac.in (G.P.S. Raghava).

https://doi.org/10.1016/j.ygeno.2020.04.028
Received 11 March 2020; Received in revised form 16 April 2020; Accepted 27 April 2020
Available online 01 May 2020
0888-7543/

annotation of somatic mutations, they are not well suited for annotation of germline variants as they lack tissue and cell-specific context of data. Mutations in enhancers cause disease through a diverse range of molecular mechanisms. Firstly, an enhancer can be deleted or amplified which leads to the change in gene expression of the target gene [16]. Secondly, rearrangements such as juxtaposition, translocation, and inversion can place the enhancer in some other region, leading to activation of a new target gene such as the *MYC* gene in Burkitt lymphoma [17].

Although several databases such as VISTA enhancer [18] and EnhancerAtlas [19] are there, they only catalog the enhancers present in different cells and tissues. Other manually curated databases such as DiseaseEnhancer [20] and EnDisease [21], catalog the mutations within enhancers that are associated with disease onset and progression. Some other databases such as OncoBase [22] have been developed by taking data from various resources such as The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC), Clinvar, etc. for cataloging the somatic mutations within the enhancer element. But to date, databases on expressed enhancer in particular cancer type, its associated genes, copy number variation (CNV) within enhancer elements is not there. Thus, aggregating the information regarding genomic and epigenomic changes within the enhancers is necessary and is the need of the hour. Considering the above diverse functional implications of genetic variants within enhancers and their involvement in disease onset and tumorigenesis, we aim to systematically annotate the enhancers in a cancer tissue-specific manner.

To fill the gap in literature studies, we have developed CancerEnD (https://webs.iiitd.edu.in/raghava/cancerend/), a user-friendly web server. CancerEnD aims to annotate and store cis-regulatory enhancer element in terms of somatic mutation, CNV, associated genes, and their association with overall survival of cancer patients at a single platform for 18 cancer types. In order to build this platform, RNA-seq gene expression and CNV data have been taken from TCGA [23], somatic mutation data from the Catalog of Somatic Mutation in Cancer (COSMIC) [5]. The clinical information file of patients was also downloaded from TCGA. This web resource provides valuable data for further integrative analysis across different cancer types.

2. Results

2.1. Database architecture and statistics

CancerEnD stores data on cancer associated enhancers, associated transcriptionally active genes, genomic changes such as CNV and somatic mutations within enhancer, and their association with overall patients' survival for 18 TCGA cancer types. In total, 8063 cancer samples were downloaded from TCGA and processed to build this platform. Anyone can access the data stored in CancerEnD webserver with the interactive user interface that is freely available at https://webs.iiitd.edu.in/raghava/cancerend/. Table 1 provides the detailed description of the data present in the CancerEnD database.

The complete data is stored in one big MySQL table that can be searched and browsed by various categories defined in the database such as cancer type, CNV, mutation and survival associated enhancer type in a time-efficient manner. User can browse the additional details such as CNVs, somatic mutations, and prognostic potential of enhancers corresponding to each cancer type provided in the browsing table. Fig. 1 illustrates the schematic representation of data analysis, integration, and CancerEnD database features.

2.2. Database analysis, results and interpretation

The data downloaded from different resources was analyzed for a variety of parameters that supports the equal importance of non-coding variants as protein-coding variants for disease development. Data analysis reveals that Esophageal Carcinoma (ESCA) followed by Skin Cutaneous Melanoma (SKCM) shows the maximum number of somatic mutations in the enhancer elements. Analysis of change in copy number among enhancers of each cancer type shows that Stomach Adenocarcinoma (STAD), Pancreatic Adenocarcinoma (PAAD) followed by Kidney Renal Cell Clear Carcinoma (KIRC) are more prone to genomic instability events (Table 1). The presence of high genomic instability in coding as well as non-coding regions may explain the high incidence of death rate among these cancer types [24,25].

Moreover, analysis of enhancer elements within each cancer type also unravels some hidden pictures of cancer molecular pathogenesis. Overall, 169 enhancers shows overlapping among 18 cancer types and the results regarding the same can be found in the supplementary Table 1. This could signify that the same molecular pathway may operate in these cancer types. The cell development is a dynamic process which is entirely regulated by factors that control gene expression [26]. We also identified 555 enhancers unique to a particular cancer type (Supplementary Table 1). These overlapping and exclusive enhancers among each cancer type are the ones that can serve the basis for explaining the molecular heterogeneity and similarity among the different cancer types along with providing little insights into disease progression. The top 30 intersections of enhancers among each cancer type is shown in the UpSet plot (Fig. 2).

Recent literature highlights that enhancer activation is a cell type specific process that orchestrates expression of specific genes, thus guiding cell determination and cell identity [26]. Disruption of enhancer activity by genetic or epigenetic means can impact a particular cell type function and could be the cause of tissue specific tumor development [27]. Our data analysis also identifies 1762 enhancer elements, where copy number changes in enhancer elements are associated with overall survival of cancer patients. The identified enhancers could serve as a molecular diagnostic biomarker for cancer specific therapeutic strategy.

2.3. Comparison with existing resources

Several databases were developed in the past to cater to the need for understanding the non-coding elements in cancer biology. These databases based on non-coding element are customized for a specific need and are ranged in size, scope, and purpose. Several groups developed databases that store information on somatic mutations of the non-coding element, examples include OncoBase, OncoCis [28]. Other manually curated databases were also developed by researchers to exemplify the role of enhancers in human disease that includes ENdb [29], DiseaseEnhancer etc. Several other databases such as SEdb [30], SEanalysis [31], were developed by integrating high-throughput data from various resources to support the fact that variants and sequence motifs within the region of enhancers are the prime driver of tumorigenesis as well other human disorders. Genomic alterations within non-coding parts of the genome such as somatic mutation, CNV can also change the phenotype of the organism and can be inherited by future generations [22]. Currently, there is no available database that provides information on the influence of CNV in the enhancer element and their subsequent impact on gene expression change. CancerEnD is a unique database that provides comprehensive collection of information on cancer-associated enhancers. It stores information on enhancer-gene linking, somatic mutations and CNV in the enhancer element, and association of enhancer CNV with overall patient survival for each cancer type. We have provided the list of enhancers associated CNV that have some prognostic potential in Supplementary Table 2. Table 2 provides the end to end comparison of CancerEnD with the other related databases. From the table, it is clearly seen that CancerEnD not only complements the existing databases but provides additional features in terms of expressed enhancer, mapping copy number variation to enhancer element and association of amplified/deleted enhancers to the patient overall survival. Thus, in this regard, CancerEnD is more clinically relevant and useful as compared to other existing resources.

Table 1
The summary of data present in CancerEnD.

Cancer type	Total enhancers	Unique enhancers	Genes	CNV	Somatic mutations	EnCNV survival	EnCNV survival genes
BLCA	11,996	4281	6225	237	7003	19	9
BRCA	7233	4280	6563	93	826	93	21
CESC	10,216	4482	7084	40	3618	40	10
ESCA	13,276	3182	4452	108	11,431	108	11
GBM	11,374	5868	9477	38	2557	38	19
HNSC	6098	3847	5562	62	728	62	13
KIRC	12,140	5211	8329	250	5019	250	64
KIRP	8694	4554	6972	8	2158	8	5
LIHC	7952	4385	6776	45	1496	45	16
LUAD	8454	4829	7525	125	1066	125	44
LUSC	10,897	5821	9723	95	1334	95	33
PAAD	7100	2243	2862	253	5881	253	22
PRAD	7239	3082	4032	56	4396	56	3
SARC	6907	3562	5104	59	2371	59	14
SKCM	11,438	2741	3698	49	9939	49	8
STAD	10,205	3984	6150	311	5243	311	9
THCA	4410	3285	4254	68	200	68	13
UCEC	12,835	4986	7840	83	5928	83	26
Total	168,464	74,623	112,628	1980	71,194	1762	340

2.4. Utility of the database and case study

CancerEnD can be interactively browsed and searched in a variety of different ways to satisfy the query of the user. The homepage of CancerEnD website provides a quick search page, where users can search query in the database for specific cancer type, genomic coordinates of enhancers, somatic mutation, CNV, survival association, gene and its related information that is stored in the database. The advanced search page in the database provides customized search facility for user defined query using AND/OR boolean expressions. In addition to simple and advanced search, CancerEnD is also equipped with a highly interactive and simple representation of browsing facility.

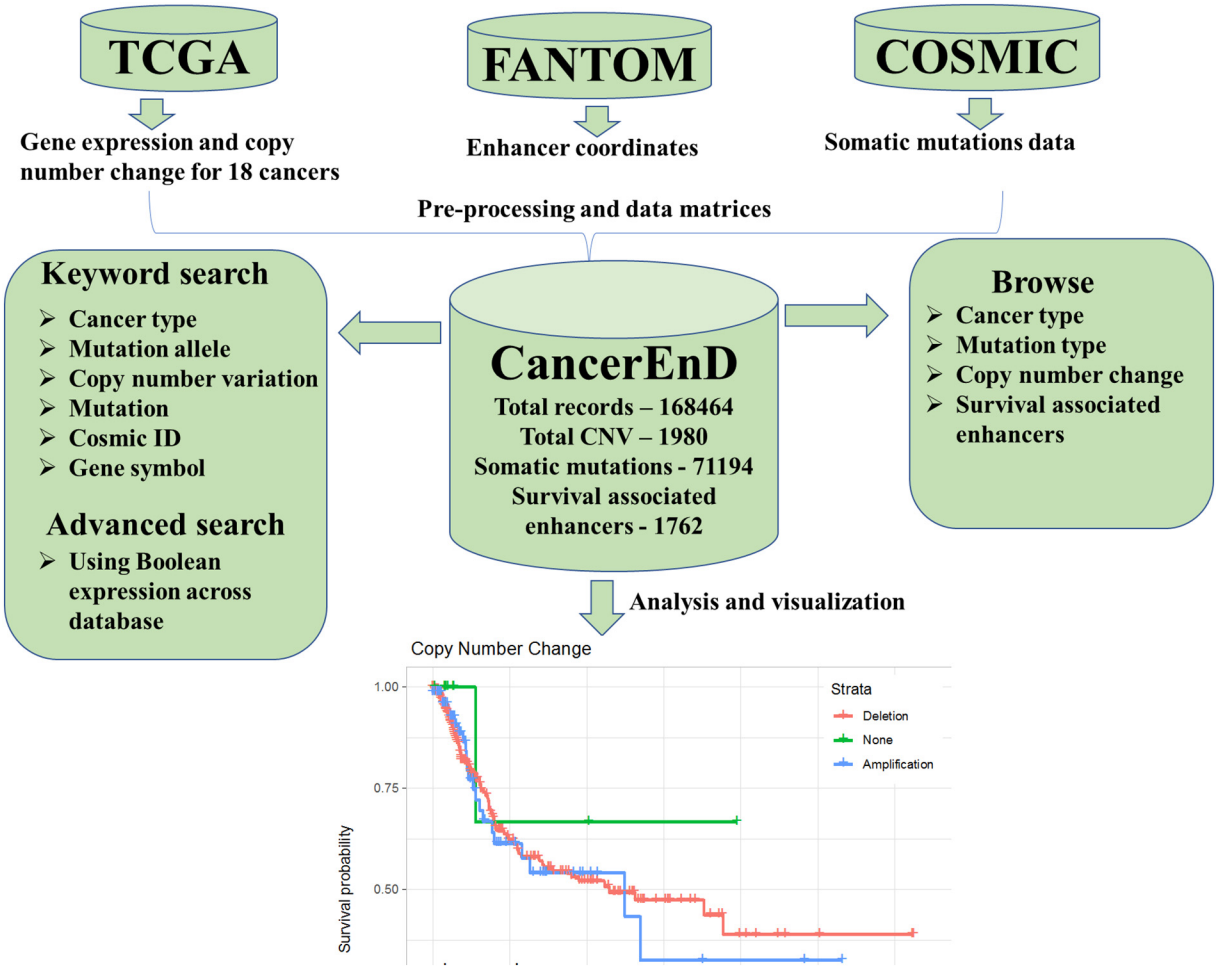

Fig. 1. Schematic representation of data analysis, integration, and CancerEnD database features.
All the data is referenced against human reference genome hg38.

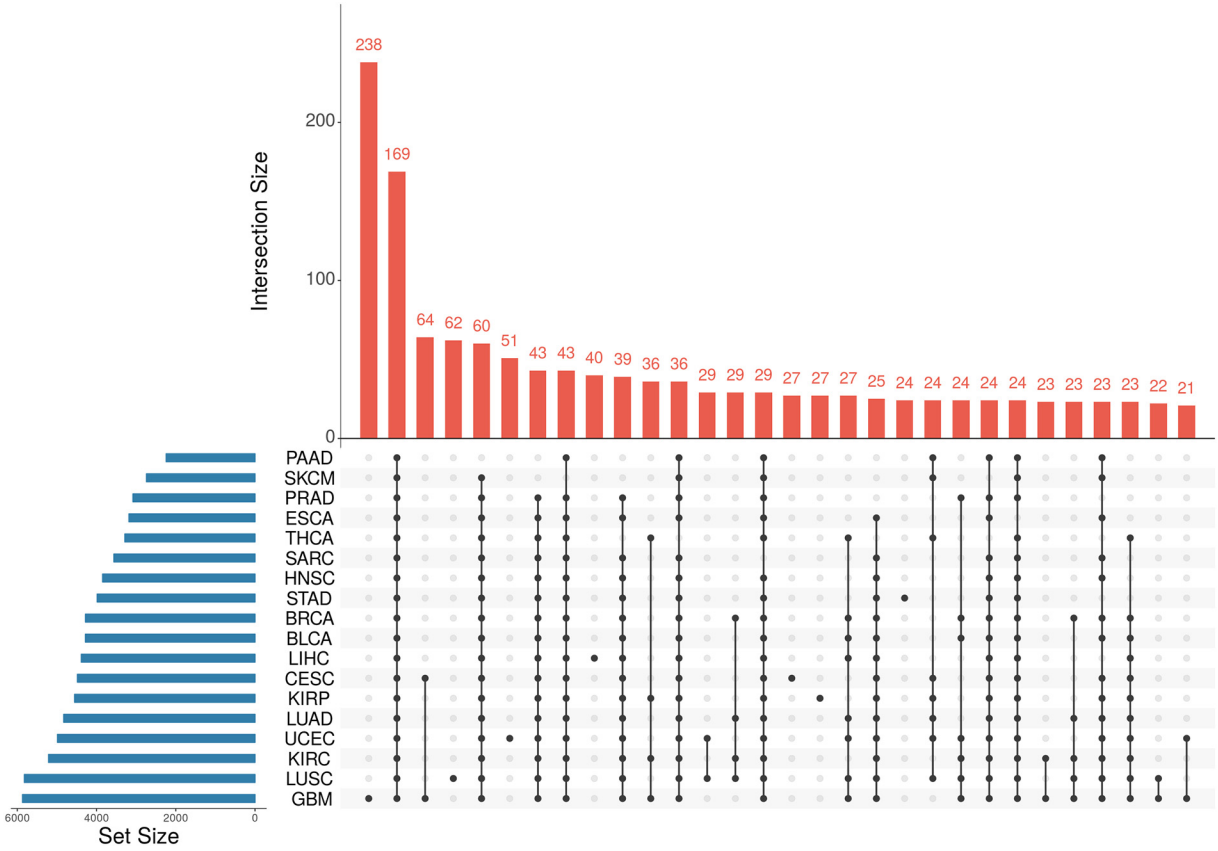

Fig. 2. UpSet plot of top 30 interactions including both common and unique enhancer among each cancer type.

We compiled the data in a tabular form corresponding to each cancer type for easy and efficient access. Users can browse for different enhancers, associated CNVs, somatic mutations and survival information for each cancer type. The advantage of the browsing facility is that users can quickly obtain all the results by clicking onto specific entry under concerned category. Moreover, 'Help' page on the website provides detailed visualization of the usage of the CancerEnD database. The full dataset is available as a tab-delimited, comma separated and XLSX file on the 'Download' page of the cancerEnD website.

2.4.1. Case study of the database

Here, we have provided step-by-step procedure to effectively search within the database. For instance, users search a "GATA3" gene name in the search page of a website. The user defined search will result in a responsive table containing information on the query such as CancerEnD_ID, cancer type, enhancer genomic coordinates, somatic mutation, reference and mutated alleles, COSMIC ID, CNV, survival association, genomic information of gene and source of data. After clicking on CancerEnD_ID, a detailed result page will open where users can get more information regarding the particular ID. Users can also click onto the survival association field, where association of a particular CancerEnD_ID with patient's overall survival is provided in downloadable PDF image format. Fig. 3 illustrates the interactive searching in the CancerEnD database.

2.4.2. Data submission limitations and update

We aim to make CancerEnD a single free to use public platform for the information regarding cancer-associated enhancers. The online data submission form in the dropdown menu of the source page, allows users of the database to submit information regarding enhancers that are associated with tumorigenesis. In the near future, we would like to expand the CancerEnD database with information on enhancer's domain, interacting residues between enhancers and transcription factors, and mutations associated with such interaction that may be the cause of tumorigenesis.

3. Discussion and conclusion

Published literature on cancer initiation, progression and survival has focused on driver mutation/variation residing in the coding region.

Table 2
Comparison of data types present in CancerEnD with other related resources.

	Cancer type	Mutation	Copy number variation	Survival linkage	Expressed enhancer
CancerEnD	18	Yes	Yes	Yes	Yes
OncoBase	68	Yes	No	No	No
Cistrome cancer	30	Yes	No	No	No
OncoCis	15 Cancer cell lines	Yes	No	No	No
SEdb	542 samples	Yes	No	No	No
SEanalysis	540 cells/tissues	Yes	No	No	No
ENdb	737 records	Yes	No	No	No
Disease enhancer	847 records	Yes	No	No	No

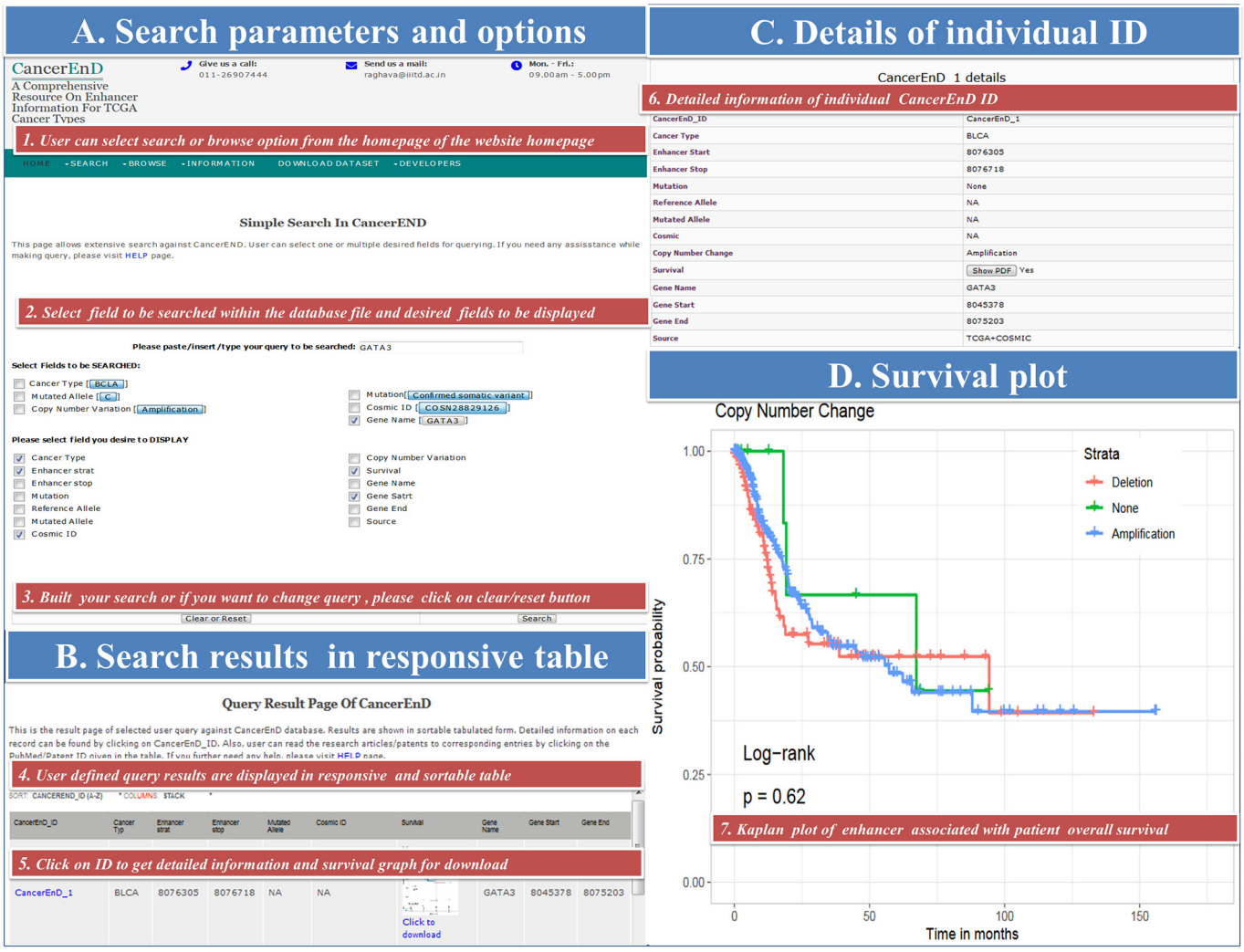

Fig. 3. Schematic illustration of the interactive searching within the CancerEnD database. (A). Homepage of the CancerEnD database. Users can select search, advance search and browse options for querying the database. We have shown the results for the gene GATA3 in the quick search box including some additional parameters such as cancer type, associated enhancer and its linkage with patient's overall survival. (B). Brief results of user defined query are displayed in a sort-enabled responsive table. (C). Users can click onto individual ID to get detailed information. (D). Users can also click on the survival plot for better visualization and to download it in the PDF format.

But, as we know that gene expression and cell phenotype are governed by a complex set of the regulatory circuit that includes promoters, enhancers, insulators, etc. For a better understanding of the disease phenotype and to provide better therapeutics, detailed knowledge of the mechanism of regulation is necessary for every gene whose dysregulated expression is known to cause disease or involved in disease progression. Enhancers have become key elements to decipher the dysregulated mechanisms of disease initiation and progression. Thus, aggregating the information regarding genomic and epigenomic changes within the enhancers are necessary. Several databases on the enhancers are there but these databases lack tissue and cell-specific context of enhancers.

Currently, there is no resource that provides information on expressed enhancers in a cancer tissue specific manner along with its associated genes, CNV and prognostic potential for particular cancer types. In this regard, we have developed a web resource CancerEnD that provides information on enhancer expression, its associated genes, somatic mutations, and CNVs within them along with their prognostic potential at a single platform for 18 cancer types. By taking data from TCGA and COSMIC, we have developed an integrative resource to fine map the casual regulatory gene pairs and annotate the enhancer mutations at the genome level. This resource provides valuable data regarding mutation and CNV. Overlapping and non-overlapping enhancer elements may provide insights into the mechanism of activation of diverse regulatory pathways in different cancer types. These types of analyses highlight the important characteristic of tumor heterogeneity. Since, enhancers are known to regulate gene expression in tissue-specific manner. The identified survival associated enhancers may guide the development of future therapeutics for cancer. Modulating the enhancer expression in a tissue-specific manner may also serve the basis for enhancer-based cancer treatment strategies.

We suggest that mutations identified within the enhancer elements must be systematically and functionally tested to investigate their effects in tumorigenesis. The experimentally tested enhancer mutations may help in the advancement of cancer therapeutics. Thus, we believe that integrating the genomic data along with identifying and prioritizing functional non-coding mutations will provide a new way in cancer research. In this regard, CancerEnD is a valuable resource that can be exploited to boost further research in the field of oncology.

4. Materials and methods

4.1. Data type and source

The present study utilized data on RNA-seq, somatic mutations, CNV and gene regulatory element i.e. enhancers to carry out an integrative analysis for the development of a bioinformatics resource. Level 3 data of RNA-seq gene expression and CNV for 18 cancer types was downloaded from TCGA with the help of TCGA assembler [32]. CNV data was downloaded with the help of FireBrowseR [33]. For gene expression, RSEM normalized count gene expression data was used. For the curation of somatic mutation, data from COSMIC was downloaded. Enhancer coordinates and their expression data were taken from the literature [34]. The literature retrieved information on enhancer coordinates and their tissue type was further confirmed by mapping them on to FANTOM database. The FANTOM database project annotated the enhancers based on integration of data from various techniques such as chromatin modification, cap analysis, histone modification, transcription factor binding sites etc.

4.2. Systematic identification of enhancer - gene interaction for each cancer type

Gene expression in eukaryotic organisms is controlled by multiple molecular systems, particularly involving transcription factor binding and enhancer elements. Specific expression of genes, that leads to carcinogenesis may be driven by the combination of enhancer and transcription factor binding. To elucidate the molecular functioning of each enhancer in the progression of cancer and their clinical utility for the advancement of cancer therapeutics, it is necessary to identify the target gene regulated by them. Byintegrating enhancers and mRNA expression data, we can accurately pinpoint the target gene regulated by an enhancer element via co-expression analysis and by utilizing the statistical approach. Transcriptional misregulation, that often leads to higher gene expression, is considered as a hallmark of disease development [35]. Normally, transcriptionally over active genes are considered as the chief players in tumorigenesis [36]. In order to identify the target genes regulated by identified enhancers, firstly, we adopted a statistical approach that utilizes TCGA RNA-seq data. We identified the transcriptionally more active genes, that showed higher expression by comparing gene expression data of tumor and matched-normal samples obtained from TCGA for 18 cancer types [25]. Enhancers are known as cis-acting element and cause higher gene expression and can regulate nearby gene expression either upstream or downstream from their start site [37,38]. We eliminated those genes from our study, whose expression change did not correspond to the nature of the enhancerelement as done in previous studies [39]. Now, based on the rationale of the nearest neighbor connection, a distance-based method for an enhancer-gene association was utilized to connect enhancer to its immediate neighbor gene but not farther than 0.1 MB on each side, as used in the previous studies [40,41]. Secondly, we further validated our identified cis enhancer-gene pairs; we used expression quantitative trait loci (eQTL) data and considered only those enhancer gene pairs that showed a positive correlation with eQTL as done in a previous study [42]. This two-step verification of results ensures that the obtained enhancer-gene pairs are of good quality and we can further use available Hi-C data to assess whether the imputed results are mean of direct regulation at chromatin level.

4.3. Somatic mutation and copy number analysis

Mutation/variant within enhancer is of high importance as these may regulate the cell-specific behavior of enhancer. To identify the somatic mutation in enhancers, somatic mutation data from COSMIC is utilized. We mapped the COSMIC somatic mutation data onto enhancer coordinates using in-house script. Copy number data was downloaded using FireBrowseR. The CNV data then processed with Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm using default parameters [43]. Significant CNV region and somatic mutation then mapped onto enhancer coordinates. The detailed description of the methodology adopted for this section is explained in our previous study [25]. Data related to enhancer expression, somatic mutations, CNV, target gene are annotated on to enhancer coordinates for 18 cancer types and are presented in the form of a web server.

4.4. Identification of survival associated enhancers

Many genes are associated with cancer prognosis and CNV may influence the target gene expression by changing the copies of the enhancer element [44,45]. We examine the association between CNV in enhancers with overall patient survival. The log-rank test was used to examine the significant difference in survival time and the Kaplan-Meier curve was used to plot the same.

4.5. Database construction, implementation, and features

The user-friendly web interface of CancerEnD (webs.iiitd.edu.in/raghava/cancerend) was developed using MySQL and run on a Linux based apache server. We used PHP for server-side scripting. The interactive user interface was designed and built using bootstrap, a popular responsive development framework including HTML, CSS, and JavaScript. The user interface is responsive, meaning that the web interface detects the user device and changes its structure and shape according to the device resolution in order to optimize the data view. This feature makes the interface compatible across a variety of devices and browsers with different screen resolutions. The database can be browsed and searched using a variety of devices including smartphones or tablets. Google Chrome, Mozilla Firefox, and Safari web browsers are preferred for better user experience. We aim to improve the accessibility and user interactivity of “CancerEnD” by asking for user feedback through the contact page on our website. We considered each object stored in the database table as a separate entry that has at least one different value across the database fields.

Authors' contribution

Conception and design: Rajesh Kumar and GPS Raghava.

Development of methodology: Rajesh Kumar, Anjali Lathwal, Pawan Raghav and GPS Raghava.

Acquisition of data: Rajesh Kumar, Anjali Lathwal, Sumeet Patiyal.

Analysis and interpretation of data: Rajesh Kumar, Vinod Kumar, Anjali Lathwal, Sumeet Patiyal, and GPS Raghava.

Web Server development: Rajesh Kumar, Anjali Lathwal and Vinod Kumar.

Writing, review, and/or revision of the manuscript: Rajesh Kumar, Anjali Lathwal and GPS Raghava.

Funding

There is no funding available for publishing this paper as an open access article.

Data availability

All data is freely available on TCGA and COSMIC websites. Moreover the data stored in CancerEnD will be freely available to download at https://webs.iiitd.edu.in/raghava/cancerend/download.html.

Declaration of Competing Interest

No potential conflict of interest was disclosed.

Acknowledgment

Authors are thankful for the funding agencies Council of scientific and industrial research (CSIR), University Grants Commission (UGC), Department of Biotechnology (DBT) andIndraprastha Institute of Information Technology (IIIT-D) Delhi for providing necessary facility and infrastructure to carry out this research work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.04.028.

References

[1] S. Diederichs, L. Bartsch, J.C. Berkmann, K. Fröse, J. Heitmann, C. Hoppe, et al., The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations, EMBO Mol Med. EMBO 8 (2016) 442–457.
[2] M. Liu, Y.C. Yang, G. Xu, C. Tan, Z.J. Lu, CCG: an integrative resource of cancer protein-coding genes and long noncoding RNAs, Discov. Med. 22 (2016) 351–359 Solariz, Inc.
[3] S.A. Shabalina, N.A. Spiridonov, The mammalian transcriptome and the function of non-coding DNA sequences, Genome Biol. 5 (2004) 105 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/15059247.
[4] J.Y. Ko, S. Oh, K.H. Yoo, Functional enhancers as master regulators of Tissue-Specific gene regulation and cancer development, Mol. Cells. Korean Soc. Mol. Cell. Biol. (2017) 169–177.
[5] S.A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, et al., COSMIC: exploring the world's knowledge of somatic mutations in human cancer, Nucleic Acids Res. 43 (2015) [cited 2020 Apr 16]. D805-11. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25355519.
[6] L. Luzzatto, Somatic mutations in cancer development, Environ Health 10 (Suppl. 1) (2011) S12 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/21489208.
[7] F. Zhang, J.R. Lupski, Non-coding genetic variants in human disease, Hum Mol Genet 24 (2015) [cited 2020 Apr 16]. R102-10. Available from. http://www.ncbi.nlm.nih.gov/pubmed/26152199.
[8] C.A. Scacheri, P.C. Scacheri, Mutations in the noncoding genome, Curr. Opin. Pediatr. (2015) 659–664 Lippincott Williams and Wilkins.
[9] B.S. Gloss, M.E. Dinger, Realizing the significance of noncoding functionality in clinical genomics, Exp Mol Med. 50 (2018) 97 Nature Publishing Group. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/30089779.
[10] I. Dunham, A. Kundaje, S.F. Aldred, P.J. Collins, C.A. Davis, F. Doyle, et al., An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74 Nature Publishing Group.
[11] Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals [Internet], (2017) [cited 2020 Apr 16]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210666/.
[12] L.D. Ward, M. Kellis, HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants, Nucleic Acids Res. 40 (2012) [cited 2020 Apr 16]. D930-4. Available from. http://www.ncbi.nlm.nih.gov/pubmed/22064851.
[13] L. Guo, J. Wang, rSNPBase 3.0: an updated database of SNP-related regulatory elements, element-gene pairs and SNP-based gene regulatory networks, Nucleic Acids Res. 46 (2018) [cited 2020 Apr 16]. D1111–6. Available from. http://www.ncbi.nlm.nih.gov/pubmed/29140525.
[14] Y. Fu, Z. Liu, S. Lou, J. Bedford, X.J. Mu, K.Y. Yip, et al., FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer, Genome Biol. 15 (2014) 480 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25273974.
[15] G.R.S. Ritchie, I. Dunham, E. Zeggini, P. Flicek, Functional annotation of noncoding sequence variants, Nat Methods 11 (2014) 294–296 Nature Publishing Group. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/24487584.
[16] D.-J. Kleinjan, P. Coutinho, Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease, Brief Funct Genomic Proteomic 8 (2009) 317–332 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/19596743.
[17] R. Dalla-Favera, M. Bregni, J. Erikson, D. Patterson, R.C. Gallo, C.M. Croce, Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 7824–7827.
[18] A. Visel, S. Minovitsky, I. Dubchak, L.A. Pennacchio, VISTA Enhancer Browser–a database of tissue-specific human enhancers, Nucleic Acids Res. 35 (2007) D88–D92 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/17130149.
[19] T. Gao, B. He, S. Liu, H. Zhu, K. Tan, J. Qian, EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types, Bioinformatics 32 (2016) 3543–3551 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/27515742.
[20] G. Zhang, J. Shi, S. Zhu, Y. Lan, L. Xu, H. Yuan, et al., DiseaseEnhancer: a resource of human disease-associated enhancer catalog, Nucleic Acids Res. 46 (2018) D78–D84 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/29059320.
[21] W. Zeng, X. Min, R. Jiang, EnDisease: a manually curated database for enhancer-disease associations, Database (Oxford) 2019 (2019) [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/30788500.
[22] X. Li, L. Shi, Y. Wang, J. Zhong, X. Zhao, H. Teng, et al., OncoBase: a platform for decoding regulatory somatic mutations in human cancers, Nucleic Acids Res. 47 (2019) [cited 2020 Apr 16]. D1044–55. Available from. http://www.ncbi.nlm.nih.gov/pubmed/30445567.
[23] J.N. Weinstein, E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, et al., The cancer genome atlas pan-cancer analysis project, Nat. Genet. (2013) 1113–1120 Nature Publishing Group.
[24] P. Hudler, Genetic aspects of gastric cancer instability, Sci. World J. 2012 (2012) Hindawi Limited.
[25] R. Kumar, S. Patiyal, V. Kumar, G. Nagpal, G.P.S. Raghava, In silico analysis of gene expression change associated with copy number of enhancers in pancreatic adenocarcinoma, Int J Mol Sci 20 (2019) MDPI AG. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/31336658.
[26] K.J. Kron, S.D. Bailey, M. Lupien, Enhancer alterations in cancer: a source for a cell identity crisis, Genome Med. 6 (2014) 77 BioMed Central Ltd. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25473436.
[27] H.-M. Herz, D. Hu, A. Shilatifard, Enhancer malfunction in cancer, Mol Cell 53 (2014) 859–866 Cell Press. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/24656127.
[28] D. Perera, D. Chacon, J.A.I. Thoms, R.C. Poulos, A. Shlien, D. Beck, et al., OncoCis: annotation of cis-regulatory mutations in cancer, Genome Biol. 15 (2014) 485 BioMed Central Ltd. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25298093.
[29] X. Bai, S. Shi, B. Ai, Y. Jiang, Y. Liu, X. Han, et al., ENdb: a manually curated database of experimentally supported enhancers for human and mouse, Nucleic Acids Res. 48 (2020) D51–D57 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/31665430.
[30] Y. Jiang, F. Qian, X. Bai, Y. Liu, Q. Wang, B. Ai, et al., SEdb: a comprehensive human super-enhancer database, Nucleic Acids Res. 47 (2019) D235–D243 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/30371817.
[31] F.-C. Qian, X.-C. Li, J.-C. Guo, J.-M. Zhao, Y.-Y. Li, Z.-D. Tang, et al., SEanalysis: a web tool for super-enhancer associated regulatory analysis, Nucleic Acids Res. 47 (2019) W248–W255 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/31028388.
[32] Y. Zhu, P. Qiu, Y. Ji, TCGA-assembler: Open-source software for retrieving and processing TCGA data, Nat. Methods (2014) 599–600 Nature Publishing Group.
[33] M. Deng, J. Brägelmann, I. Kryukov, N. Saraiva-Agostinho, S. Perner, FirebrowseR: an R client to the Broad Institute's Firehose Pipeline, Database (Oxford) 2017 (2017) [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/28062517.
[34] H. Chen, C. Li, X. Peng, Z. Zhou, J.N. Weinstein, Cancer Genome Atlas Research Network H, et al., A pan-cancer analysis of enhancer expression in nearly 9000 patient samples, Cell 173 (2018) 386–399 [cited 2019 May 2]. e12. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867418303076.
[35] T.I. Lee, R.A. Young, Transcriptional regulation and its misregulation in disease, Cell (2013) 1237–1251.
[36] E. Furuta, H. Okuda, A. Kobayashi, K. Watabe, Metabolic genes in cancer: their roles in tumor progression and clinical implications, Biochim Biophys Acta 1805 (2010) 141–152 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/20122995.
[37] L.A. Pennacchio, W. Bickmore, A. Dean, M.A. Nobrega, G. Bejerano, Enhancers: five essential questions, Nat. Rev. Genet. (2013) 288–295.
[38] G.A. Maston, S.K. Evans, M.R. Green, Transcriptional regulatory elements in the human genome, Annu. Rev. Genomics Hum. Genet. Annual Rev. 7 (2006) 29–59.
[39] L. Yao, Y.G. Tak, B.P. Berman, P.J. Farnham, Functional annotation of colon cancer risk SNPs, Nat. Commun. 5 (2014) 5114 Nature Publishing Group. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25268989.
[40] S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, et al., GeneHancer: genome-wide integration of enhancers and target genes in GeneCards, Database (Oxford) 2017 (2017).
[41] S. Mei, C.A. Meyer, R. Zheng, Q. Qin, Q. Wu, P. Jiang, et al., Cistrome cancer: a web resource for integrative gene regulation modeling in cancer, Cancer Res. 77 (2017) e19–e22. American Association for Cancer Research Inc. [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/29092931.
[42] J. Gong, S. Mei, C. Liu, Y. Xiang, Y. Ye, Z. Zhang, et al., PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types, Nucleic Acids Res. 46 (2018) D971–D976 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/29036324.
[43] C.H. Mermel, S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, Genome Biol. 12 (2011) R41 BioMed Central.
[44] E.R. Gamazon, B.E. Stranger, The impact of human copy number variation on gene expression, Brief Funct Genomics [Internet] 14 (2015) 352–357 [cited 2020 Apr 16]. Available from. http://www.ncbi.nlm.nih.gov/pubmed/25922366.
[45] X. Li, Y. Liu, J. Lu, M. Zhao, Integrative analysis to identify oncogenic gene expression changes associated with copy number variations of enhancer in ovarian cancer, Oncotarget. 8 (2017) Impact Journals LLC. [cited 2020 Apr 16]. 91558–67. Available from. http://www.ncbi.nlm.nih.gov/pubmed/29207666.