doi: 10.1093/nar/gkv1264

BreCAN-DB: a repository cum browser of personalized DNA breakpoint profiles of cancer genomes

Pankaj Narang[1], Parashar Dhapola[2] and Shantanu Chowdhury[2,3,*]

[1]School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, [2]GNR Center for Genome Informatics, Institute of Genomics and Integrative Biology, CSIR, Delhi, India and [3]Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR, Delhi, India

Received August 15, 2015; Revised October 29, 2015; Accepted November 03, 2015

ABSTRACT

BreCAN-DB (http://brecandb.igib.res.in) is a repository cum browser of whole genome somatic DNA breakpoint profiles of cancer genomes, mapped at single nucleotide resolution using deep sequencing data. These breakpoints are associated with deletions, insertions, inversions, tandem duplications, translocations and a combination of these structural genomic alterations. The current release of BreCAN-DB features breakpoint profiles from 99 cancer-normal pairs, comprising five cancer types. We identified DNA breakpoints across genomes using high-coverage next-generation sequencing data obtained from TCGA and dbGaP. Further, in these cancer genomes, we methodically identified breakpoint hotspots which were significantly enriched with somatic structural alterations. To visualize the breakpoint profiles, a next-generation genome browser was integrated with BreCAN-DB. Moreover, we also included previously reported breakpoint profiles from 138 cancer-normal pairs, spanning 10 cancer types into the browser. Additionally, BreCAN-DB allows one to identify breakpoint hotspots in user uploaded data set. We have also included a functionality to query overlap of any breakpoint profile with regions of user's interest. Users can download breakpoint profiles from the database or may submit their data to be integrated in BreCAN-DB. We believe that BreCAN-DB will be useful resource for genomics scientific community and is a step towards personalized cancer genomics.

INTRODUCTION

It is widely understood that incidence of structural genomic alterations (GAs) is frequently associated with cancer and indeed has been posed as a prominent cause for acquisition of hallmarks of cancer initiation/progression (1,2). These GAs could be large-sized deletions, insertions, inversions, translocations, fusions and copy number variations (3–5). Studies indicate that GAs not only account for larger genomic heterogeneity in populations but may also have a much larger than anticipated impact on oncogenesis and cancer progression (6,7). Most studies so far have focused mainly upon association of mutations, especially non-synonymous point mutations present in the coding regions of the genome. Results from such studies have shown that though certain point mutation events in cancer are more associated with specific genes like *BRCA1* in breast cancer (8), *TP53* in glioblastoma multiforme (9), such mutations are present in only a small percentage of disease cases (10). This has led to an increased interest in large-sized GAs (that may span from tens of bases to few megabases of genome) to better understand cancer etiology.

Profiling GAs have been more challenging than profiling point mutations, mainly due to requirement of deep coverage genome sequence data and algorithms that can map sequence reads to identify breakpoints at single base resolution. With fulfilment of these requirements, various efforts to elucidate the role of GAs in cancer genome have been made. However, they were unable to include either large sample size, single base resolution or/and accessible data for individual breakpoint profile. With these features in mind, BreCAN-DB was developed to allow users navigate genomes for mining individualized single base resolution breakpoint events in a readily scalable setting.

Recent studies involving meta-analysis of large-scale data sets have indicated the utility of integrating and methodologically making statistical analysis to gain biological insights; these span from analysis of structural variants at low resolution, gene expression and small RNA expression (11–13). Moreover, in last few years, various prominent resources have been developed such as COSMIC (14), The Cancer Genome Atlas (TCGA) data portal (https://tcga-data.nci.nih.gov/tcga/), UCSC Cancer Genomics Browser

[*]To whom correspondence should be addressed. Tel: +91 11 2766 6157; Fax: +91 11 2766 7471; Email: shantanuc@igib.res.in

(15) and International Cancer Genome consortium (ICGC) data portal (16), serving data related to genomic variations in cancer both as pooled datasets as well as at personalized level. Of these, COSMIC uses a gene-focused approach where most mutation data are available for genes nominated from Cancer Gene Census (17) while TCGA and ICGC provide genomic SNP and copy number profiles at personalized level. BreCAN-DB complements these databases by adopting an individual profile-based approach for large-sized GAs over gene-based approach, so that information on low recurrence but potentially relevant events is retained. It is being increasingly appreciated to understand genome in a personalized context. Also, many of the genomic variations have already been associated with response to therapy while many others still remain elusive (18). Hence, it is crucial that each cancer's structural variation is studied individually so that viable targets for precision therapy may be developed.

To facilitate studies involving DNA breakpoints at personalized levels, we have made BreCAN-DB: '**BRE**akpoint profiles of **CAN**cer genomes **D**ata**B**ase', a browsable repository of personalized cancer DNA breakpoint profiles mapped over the entire genomes of 99 cancer and matched control pairs from five cancer types *viz*. glioblastoma multiforme (GBM), breast invasive carcinoma (BRC), lung adenocarcinoma (LUAD), ovarian serous cystadenocarcinoma (OV) and head and neck squamous cell carcinoma (HN). We also provide functionality of visualizing and comparing DNA breakpoint profiles between samples of same cancer type or across different cancer types using BreCAN-DB.

MATERIALS AND METHODS

Implementation

BreCAN-DB has been implemented in three phases: Phase 1- collection of whole genome sequencing data of cancer and matched normal tissues, Phase 2- mapping of somatic DNA breakpoint profiles and breakpoint hotspots and Phase 3- user interface development for interactive visualization through web browser. These three phases are described below.

Data collection

To generate DNA breakpoint profiles, we retrieved the Illumina whole genome paired end sequencing reads from 99 cancer-normal pairs comprising five cancer types *viz*. 15 pairs of BRC from Banerji et. al (19), 27 pairs of GBM (9), 17 pairs of LUAD (20), 23 pairs of OV (21) and 17 pairs of HN (22) from The Cancer Genome Atlas (phase 1 of Figure 1A). Sequencing data for GBM, LUAD, OV and HN were downloaded in the BAM format (23), while data for BRC were downloaded in SRA (sequence read archive) format from the database of Genotypes and Phenotypes (24) (dbGaP Study Accession: phs000369 and phs000178). The paired-end sequencing reads of BRC were aligned to the human reference genome (UCSC build hg19) using BWA aligner with default parameters to create BAM files (25). Further, we also incorporated data of 138 cancer genomes from Yang et al. where breakpoints were called on cancer genomes using Meerkat (26).

Mapping of breakpoint profiles

Breakpoints of GAs were identified in all normal and cancer genomes independently using Meerkat with default parameters (26). The algorithm of Meerkat calls GAs at positions where cluster of read pairs are discordant and subsequently, the breakpoints of these GAs are precisely determined at single nucleotide resolution using soft-clipped reads. These GAs include insertions, deletions, intra and inter-chromosomal translocations, inversions, tandem duplications and combination of events. To reduce the number of false positives, we filtered GAs, supported by at least two discordant pairs and two soft-clipped reads (phase 2 of Figure 1A). GAs with size more than 10 Mb were also filtered out to get a final list of high confidence GAs. GAs specifically present in only cancer genomes (somatic GAs) were identified by subtracting GAs of all normal genomes from each cancer genome using Meerkat. Further, these GAs were filtered using Database of Genomic Variants (27) to get list of only cancer-specific GAs.

Mapping of breakpoint hotspots

To identify DNA breakpoint hotspots, we extended breakpoints of each cancer-specific GA by 10 Kb upstream and downstream (phase 2 of Figure 1A). Further, using boundaries of extended breakpoint regions from all cancer genomes, we segmented the reference genome into non-overlapping bins of unequal sizes. The bins that overlapped with extended regions in more than one cancer genome were identified and their enrichment score was calculated as number of samples harbouring extended breakpoint regions overlapping with these bins. Next, we compared these scores with expected distribution of score calculated by randomizing the GAs of all samples 10000 times and calculated their significance. The bins with P-value <0.05 after Bonferroni Hochberg correction for multiple hypothesis testing were finally selected as cancer breakpoint hotspots. Similar strategy was used to find breakpoint hotspots for all five types of cancer separately.

Development of web-accessible interface

The interactive front end interface of database was designed using HTML, CSS and JavaScript and made web-accessible through Apache2 HTTP server running on CentOS7 Linux platform (phase 3 of Figure 1A). The server is hosted at Institute of Genomics and Integrative Biology and can be accessed at http://brecandb.igib.res.in. For visualization of breakpoint profiles in BreCAN-DB, these breakpoint profiles were converted to json format and integrated with JBrowse, an advanced, open source and JavaScript-based genome browser (28,29). JBrowse uses an approach where most feature rendering is done by the client machine rather than server, resulting in quick response to the user. Further, other modules to submit and download breakpoint profiles in BED format were implemented using PHP. The modules 'Compare Breakpoint profiles' to compare a new genomic profile against breakpoint profiles of database and 'Map Breakpoint hotspots' to map breakpoint enriched regions were implemented using BEDTools (30) and the former outputs sortable barplots, implemented using D3.js (31).

RESULTS

Database content and statistics

BreCAN-DB is DataBase of personalized somatic DNA BREakpoint profiles of CANcer genome. The current version of the database comprises whole genome somatic breakpoint profiles of 99 cancer-normal pairs from five types of cancer obtained from The Cancer Genome Atlas and dbGaP (19). The raw breakpoints were mapped at single nucleotide resolution in cancer genome deep sequencing data using Meerkat software (26), which were further filtered using GAs from corresponding cancer type pooled normal data and Database of Genomic variants (27) (see Methods; Figure 1A). These profiles were named as GBM_profile1 to GBM_profile27 for 27 breakpoint profiles of GBM, BRC_profile1 to BRC_profile15 for 15 BRC, LUAD_profile1 to LUAD_profile17 for 17 LUAD, OV_profile1 to OV_profile23 for 23 OV and HN_profile1 to HN_profile15 for 15 HN as shown in Figure 1B. These DNA breakpoints correspond to large-sized deletions, tandem duplications, insertions, inversions, intra and inter-chromosomal translocations and combination of such events. A total of 1 92 626 breakpoints were identified with average number of 1945 breakpoints per cancer genome. Table 1 provides the number of breakpoints in different cancer genomes. Further, we have also mapped significant breakpoint hotspots in five different types of cancer, i.e. genomic regions with more number of observed DNA breakpoints in comparison to expected (see Methods; Figure 1A). BreCAN-DB also collates 4415 breakpoint hotspots distributed over the whole genome (see Table 2). Further, we included 138 whole genome breakpoint profiles from 10 cancer types mapped by Yang et al. (26). This data set included three cancer types already present in our analysis *viz.* breast invasive carcinoma (BRCA; $n = 35$), glioblastoma multiforme (GBM; $n = 16$) and ovarian serous cystadenocarcinoma (OV; $n = 9$); beyond these it included seven other cancer types *viz.* colorectal adenocarcinoma (CRC; $n = 14$), multiple myeloma (MM; $n = 7$), prostrate adenocarcinoma (PR; $n = 7$), hepatocellular carcinoma (HCC; $n = 19$), lung squamous cell carcinoma (LUSC; $n = 18$), uterine corpus endometrioid carcinoma (UCEC; $n = 10$) and kidney renal clear cell carcinoma (KIRC; $n = 3$).

Apart from serving as repository, we have included five modules in BreCAN-DB that help users analyse their own data with breakpoints present in the database. These five modules are ‘Breakpoint Profiles and hotspots browser’, ‘Map Breakpoint hotspots’, ‘Compare Breakpoint Profiles’, ‘Submit Breakpoint Profiles’, ‘Download Breakpoint Profiles’ (Figure 1C), these have been discussed further in the following sections.

Breakpoint Profiles and hotspots browser

To visualize the distribution of DNA breakpoint profiles of cancer genomes, we provided a genome browser as shown in Figure 2A. The browser tracks have been divided into mainly three sections. Section 1 contains the tracks of human reference genome (assembly hg19), chromosome bands and Refseq genes downloaded from UCSC. Section 2 contains five tracks, each representing breakpoint hotspots (see Methods for details) for one type of cancer. Section 3 contains tracks of personalized whole genome breakpoint profiles, each representing breakpoint profile of a sample from particular cancer type. A separate tab was created in this section to include data from external sources (currently from Yang et al.).

The views are available from chromosome level to nucleotide level. For better visualization, each DNA breakpoint in breakpoint profile tracks has been shown as rectangular bar representing a 20 Kb region centralized at the breakpoint. Moreover, information on chromosome banding is also provided as separate track. Any of these tracks can be switched on or off by user for interactive visualization. The genomic sequence of hotspots and breakpoints can also be obtained by clicking on the respective feature. Also, we have annotated each breakpoint such that its associated genomic alteration type is directly visible to the user. It also allows user to upload breakpoint profiles in Browser Extensible (BED) format (a flexible format that allows features to be represented in form of genomic start and end coordinates) and visualize it along with one or more selected profiles already available in the database. In this way, user can directly compare whether the uploaded breakpoints overlap with his gene of interest or breakpoint hotspots or breakpoint profile of any of the five types of cancer.

Map breakpoint hotspots

We also provide a module wherein user can submit DNA breakpoint profiles of multiple samples and obtain breakpoint hotspots, regions with breakpoints in more than given number of samples (Figure 2B). The breakpoint profiles in BED format must be zipped with one file per sample and uploaded to the module. These breakpoints should be provided at nucleotide base resolution as this module first extends each breakpoint by 10Kb upstream and downstream. Next, it divides the human genome into non-overlapping bins of unequal sizes using boundaries of extended regions and finds bins that overlap with more than threshold number (given by user) of samples. The output file contains genomic coordinates of breakpoint hotspots along with number of samples having overlap of extended regions.

Compare breakpoint profiles

One of the important modules of BreCAN-DB allows one to compare user-uploaded new breakpoint profiles against hotspots or individual profiles already present in database (Figure 2C). The breakpoint profiles can either be uploaded as a file from local machine or paste in BED format. User can select type of cancer against which given breakpoint profile will be compared. On submitting data, module reports number of overlapping breakpoints with the hotspots and with the breakpoint profiles as sortable bar plots.

Submit breakpoint profiles

For further populating the database, we have included a module wherein user can submit whole genome breakpoint profiles of different cancer genomes which will be made

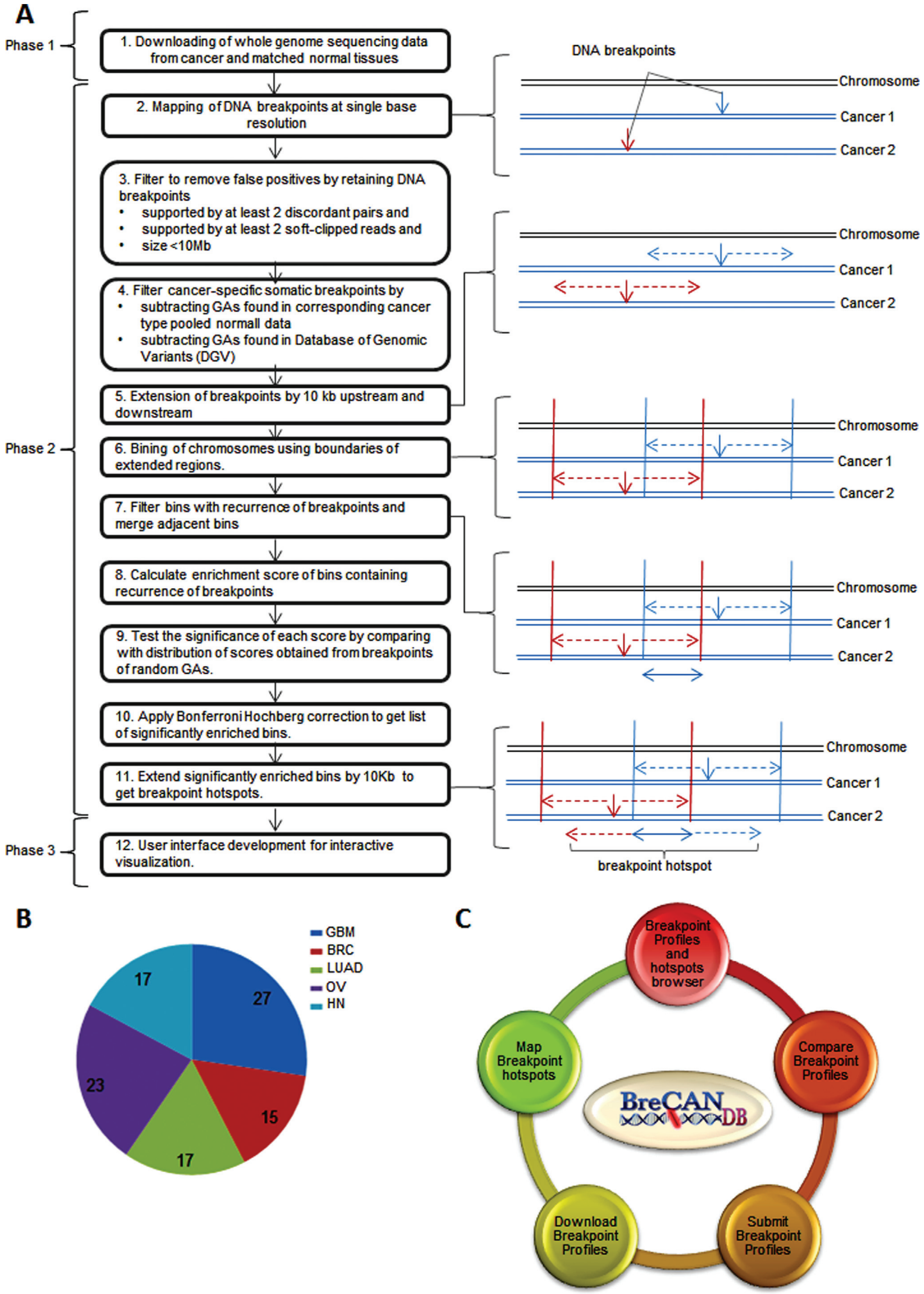
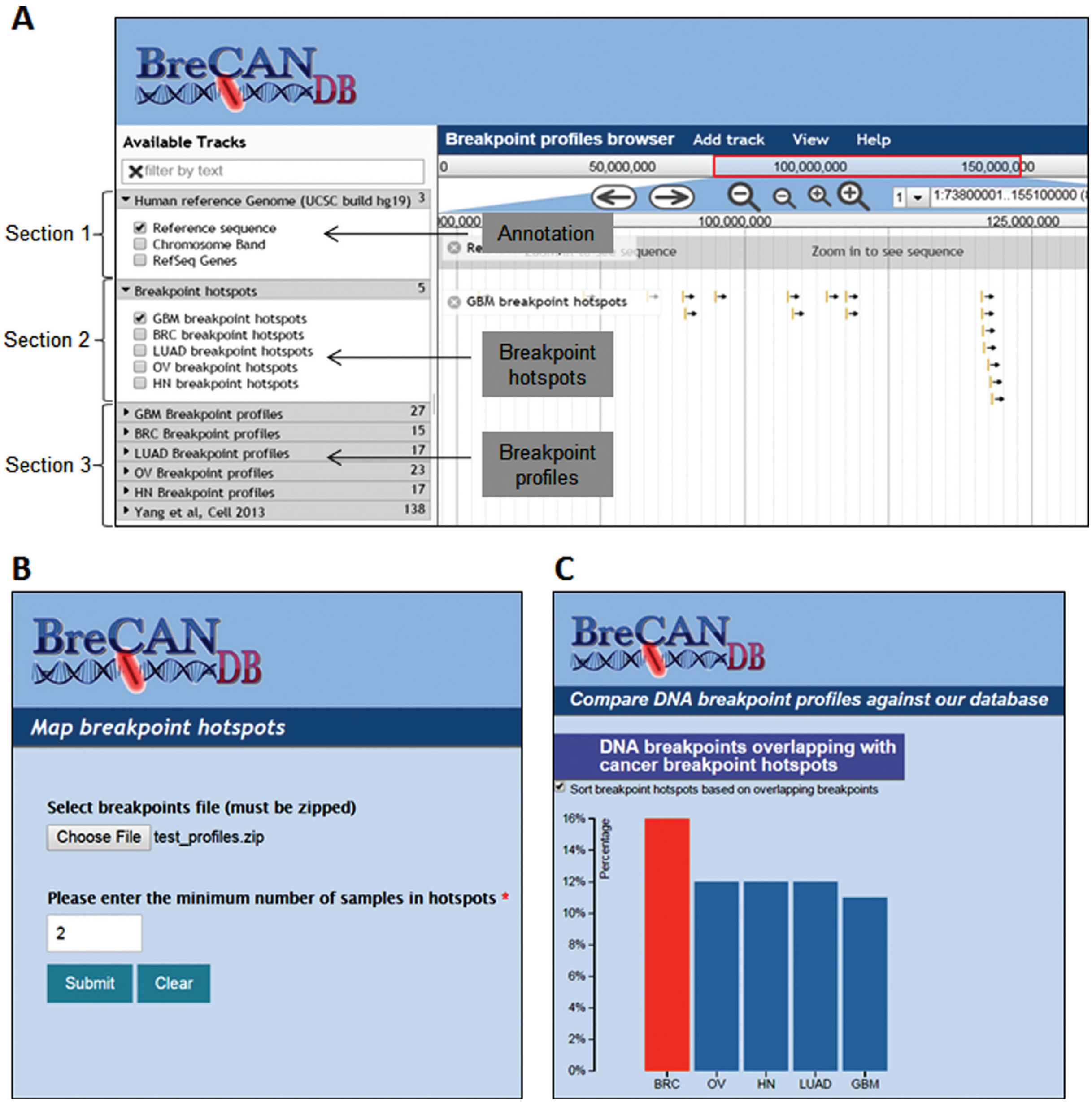

A
Phase 1
1. Downloading of whole genome sequencing data from cancer and matched normal tissues
Phase 2
2. Mapping of DNA breakpoints at single base resolution
3. Filter to remove false positives by retaining DNA breakpoints
• supported by at least 2 discordant pairs and
• supported by at least 2 soft-clipped reads and
• size <10Mb
4. Filter cancer-specific somatic breakpoints by
• subtracting GAs found in corresponding cancer type pooled normall data
• subtracting GAs found in Database of Genomic Variants (DGV)
5. Extension of breakpoints by 10 kb upstream and downstream
6. Bining of chromosomes using boundaries of extended regions.
7. Filter bins with recurrence of breakpoints and merge adjacent bins
8. Calculate enrichment score of bins containing recurrence of breakpoints
9. Test the significance of each score by comparing with distribution of scores obtained from breakpoints of random GAs.
10. Apply Bonferroni Hochberg correction to get list of significantly enriched bins.
11. Extend significantly enriched bins by 10Kb to get breakpoint hotspots.
Phase 3
12. User interface development for interactive visualization.
DNA breakpoints
Chromosome
Cancer 1
Cancer 2
breakpoint hotspot
B
GBM
BRC
LUAD
OV
HN
27
15
17
23
17
C
Breakpoint Profiles and hotspots browser
Compare Breakpoint Profiles
Submit Breakpoint Profiles
Download Breakpoint Profiles
Map Breakpoint hotspots
BreCAN DB

Figure 1. (**A**) Three phases for the development of BreCAN-DB: Phase 1- collection of whole genome sequencing data from cancer and matched normal tissues; Phase 2- mapping of breakpoint profiles and breakpoint hotspots; and Phase 3- designing a web-accessible and interactive user interface. (**B**) Pie chart representing number of cancer-normal pairs in five types of cancer. (**C**) Five modules in BreCAN-DB.



▾ Human reference Genome (UCSC build hg19) 3
☑ Reference sequence
☐ Chromosome Band
☐ RefSeq Genes

▾ Breakpoint hotspots 5
☑ GBM breakpoint hotspots
☐ BRC breakpoint hotspots
☐ LUAD breakpoint hotspots
☐ OV breakpoint hotspots
☐ HN breakpoint hotspots

▸ GBM Breakpoint profiles 27
▸ BRC Breakpoint profiles 15
▸ LUAD Breakpoint profiles 17
▸ OV Breakpoint profiles 23
▸ HN Breakpoint profiles 17
▸ Yang et al, Cell 2013 138

Select breakpoints file (must be zipped)
test_profiles.zip

Please enter the minimum number of samples in hotspots *
2

☑ Sort breakpoint hotspots based on overlapping breakpoints

Figure 2. (**A**) A view of ‘Breakpoint Profiles and hotspots Browser’ module in BreCAN-DB representing tracks of overlapping genes, breakpoint profiles and hotspots. The tracks in browser are divided into three sections: Section 1- tracks of human reference genome assembly hg19, chromosome bands and Refseq genes; Section 2- tracks of breakpoint hotspots. Each rectangular bar in these tracks represents one breakpoint hotspot; and Section 3- tracks of individual breakpoint profiles. Each rectangular bar in these tracks represents a 20 Kb region centralized at the breakpoint position. (**B**) ‘Map Breakpoint hotspots’ module takes breakpoint profiles as input and maps regions with breakpoint in at least given number of samples. (**C**) ‘Compare Breakpoint Profiles’ module compares user provided breakpoint profile against breakpoint hotspots and profiles present in the database and shows overlaps in the form of sortable bar plots.

Table 1. Summary of data sets used

Cancer type	Source	Number of cancer-normal pairs	Number of Breakpoints	Number of breakpoints per cancer genome
Glioblastoma multiforme (GBM)	TCGA (9)	27 pairs	55 874	2069
Breast invasive carcinoma (BRC)	Banerji et al. (19)	15 pairs	22 077	1471
Lung adenocarcinoma (LUAD)	TCGA (20)	17 pairs	36 086	2122
Ovarian serous cystadenocarcinoma (OV)	TCGA (21)	23 pairs	56 435	2453
Head and neck squamous cell carcinoma (HN)	TCGA (22)	17 pairs	22 154	1303
Total		99 pairs	1 92 626	1945

Table 2. Summary of breakpoint hotspots

Cancer type	Number of hotspots	Length of genomic region (Mb)
GBM	839	36
BRC	759	31
LUAD	846	37
OV	990	45
HN	981	43
Total	4415	192

open for scientific community after curation. While submitting such profiles, we request submitters to describe methods to map breakpoints clearly in order to reduce method dependent biasing of the results.

Download breakpoint profiles

We also provide a 'Download' module to download all the 99 whole genome breakpoint profiles in BED format. This will be useful for users who want to locally run analysis through integrating BreCAN-DB data sets with their own data.

CONCLUSION AND FUTURE DEVELOPMENTS

Each cancer genome has unique set of structural GAs, SNPs and other genomic variations and keeping this in mind, BreCAN-DB has been developed as a unique database that not only allows user to visualize individual DNA breakpoint profiles across cancer types but also perform comparative study of their samples against existing data. Through BreCAN-DB, we provide a platform that is readily scalable for further sample's data and allows study of somatic DNA personalized breakpoint at single nucleotide resolution. In summary, we have first developed a curated data set representing high-resolution breakpoint profiles for 99 cancer-normal pairs. Secondly, we systemically identified hotspot regions across cancer types using a novel strategy. Thirdly, we made these data accessible through an integrated next-generation genome browser. Fourthly, we created a platform, where users can perform comparative cancer genomics by checking intersection between their and BreCAN-DB data sets. These data sets can easily be extended by addition of more breakpoint profiles requiring no additional analyses or change in existing data sets as demonstrated through inclusion of data set from Yang et al. (26).

In our data set, we noted that the mean number of breakpoints were higher than previously reported (26). Since we used the same software as used in previous study (Meerkat) for profiling breakpoints, it is unlikely that this difference is due to methodology of mining breakpoints. Rather, we speculate that this might be a sample specific effect as the number of breakpoints was variable across samples in both studies. Furthermore, we found that the average number of genomic alterations in our study (excluding translocations and inversions) were similar to the ones noted earlier (this study did not report translocations and inversions) (32).

We request users to help scientific community by submitting such breakpoint profiles from their study to BreCAN-DB. We believe that this database will prove useful for understanding large-scale personal cancer genomics data.

AVAILABILITY

The database is freely available at http://brecandb.igib.res.in.

ACKNOWLEDGEMENTS

We acknowledge Dr Andrew Lynn and Dr Subhajyoti De for their valuable suggestions. We thank members of Chowdhury lab for their help and support. We also acknowledge IT team at CSIR-IGIB for their assistance in deployment of server.

FUNDING

Department of Biotechnology for BINC-SRF [to P.N.]; Council of Scientific and Industrial Research for SPM-JRF [to P.D.] and CSIR Project BSC0123 (GENCODE-A); Senior Research Fellowship from Wellcome Trust/Department of Biotechnology India Alliance [500127/Z/09/Z to S.C.]; Funding for open access charge: Wellcome Trust/DBT India Alliance [500127/Z/09/Z to S.C.].

Conflict of interest statement. None declared.

REFERENCES

1. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.
2. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
3. Shlien,A. and Malkin,D. (2009) Copy number variations and cancer. *Genome Med.*, **1**, 62.
4. Santarius,T., Shipley,J., Brewer,D., Stratton,M.R. and Cooper,C.S. (2010) A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer*, **10**, 59–64.
5. Weir,B., Zhao,X. and Meyerson,M. (2004) Somatic alterations in the human cancer genome. *Cancer Cell*, **6**, 433–438.
6. Sottoriva,A., Spiteri,I., Piccirillo,S.G.M., Touloumis,A., Collins,V.P., Marioni,J.C., Curtis,C., Watts,C. and Tavaré,S. (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 4009–4014.
7. Aubele,M., Mattis,A., Zitzelsberger,H., Walch,A., Kremer,M., Hutzler,P., Höfler,H. and Werner,M. (1999) Intratumoral heterogeneity in breast carcinoma revealed by laser-microdissection and comparative genomic hybridization. *Cancer Genet. Cytogenet.*, **110**, 94–102.
8. Comprehensive molecular portraits of human breast tumours. (2012) *Nature*, **490**, 61–70.
9. Brennan,C.W., Verhaak,R.G.W., McKenna,A., Campos,B., Noushmehr,H., Salama,S.R., Zheng,S., Chakravarty,D., Sanborn,J.Z., Berman,S.H. *et al.* (2013) The somatic genomic landscape of glioblastoma. *Cell*, **155**, 462–477.
10. Peller,S. and Rotter,V. (2003) TP53 in hematological cancer: low incidence of mutations with significant clinical relevance. *Hum. Mutat.*, **21**, 277–284.
11. Yoshihara,K., Wang,Q., Torres-Garcia,W., Zheng,S., Vegesna,R., Kim,H. and Verhaak,R.G.W. (2015) The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, **34**, 4845–4854.
12. Wang,Y., Wu,N., Liu,J., Wu,Z. and Dong,D. (2015) FusionCancer: a database of cancer fusion genes derived from RNA-seq data. *Diagn. Pathol.*, **10**, 131.
13. Fehrmann,R.S.N., Karjalainen,J.M., Krajewska,M., Westra,H.-J., Maloney,D., Simeonov,A., Pers,T.H., Hirschhorn,J.N., Jansen,R.C.,

Schultes,E.A. *et al.* (2015) Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.*, **47**, 115–125.

14. Forbes,S.A., Beare,D., Gunasekaran,P., Leung,K., Bindal,N., Boutselakis,H., Ding,M., Bamford,S., Cole,C., Ward,S. *et al.* (2014) COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.*, **43**, D805–D811.
15. Goldman,M., Craft,B., Swatloski,T., Ellrott,K., Cline,M., Diekhans,M., Ma,S., Wilks,C., Stuart,J., Haussler,D. *et al.* (2013) The UCSC Cancer Genomics Browser: update 2013. *Nucleic Acids Res.*, **41**, D949–D954.
16. Zhang,J., Baran,J., Cros,A., Guberman,J.M., Haider,S., Hsu,J., Liang,Y., Rivkin,E., Wang,J., Whitty,B. *et al.* (2011) International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database (Oxford).*, bar026.
17. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
18. Duffy,M.J. and Crown,J. (2008) A personalized approach to cancer treatment: how biomarkers can help. *Clin. Chem.*, **54**, 1770–1779.
19. Banerji,S., Cibulskis,K., Rangel-Escareno,C., Brown,K.K., Carter,S.L., Frederick,A.M., Lawrence,M.S., Sivachenko,A.Y., Sougnez,C., Zou,L. *et al.* (2012) Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, **486**, 405–409.
20. Collisson,E.A., Campbell,J.D., Brooks,A.N., Berger,A.H., Lee,W., Chmielecki,J., Beer,D.G., Cope,L., Creighton,C.J., Danilova,L. *et al.* (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, **511**, 543–550.
21. Integrated genomic analyses of ovarian carcinoma. (2011) *Nature*, **474**, 609–615.
22. Lawrence,M.S., Sougnez,C., Lichtenstein,L., Cibulskis,K., Lander,E., Gabriel,S.B., Getz,G., Ally,A., Balasundaram,M., Birol,I. *et al.* (2015) Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, **517**, 576–582.
23. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
24. Tryka,K.A., Hao,L., Sturcke,A., Jin,Y., Wang,Z.Y., Ziyabari,L., Lee,M., Popova,N., Sharopova,N., Kimura,M. *et al.* (2014) NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.*, **42**, D975–D979.
25. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
26. Yang,L., Luquette,L.J., Gehlenborg,N., Xi,R., Haseley,P.S., Hsieh,C.-H., Zhang,C., Ren,X., Protopopov,A., Chin,L. *et al.* (2013) Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell*, **153**, 919–929.
27. MacDonald,J.R., Ziman,R., Yuen,R.K.C., Feuk,L. and Scherer,S.W. (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.*, **42**, D986–D992.
28. Skinner,M.E., Uzilov,A. V, Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
29. Skinner,M.E. and Holmes,I.H. (2010) Setting up the JBrowse genome browser. *Curr. Protoc. Bioinformatics*. Chapter 9, Unit 9.13.
30. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
31. Bostock,M., Ogievetsky,V. and Heer,J. (2011) D^3: Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.*, **17**, 2301–2309.
32. De,S. and Michor,F. (2011) DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat. Struct. Mol. Biol.*, **18**, 950–955.