Table S1: Primary and secondary resources capturing multi-omics data in the context of cancer.

Resource	Description	URL	Pro/Con	Ref
Omics Data Resources for Cancer Research				
GEO	High-throughput gene expression and functional genomics data. It contains all types of omics data. No. of data analyses modules are available. Currently, it maintains more than 43K series datasets for cancer.	https://www.ncbi.nlm.nih.gov/geo/	**Pro:** It encompassed both raw and processed data. Processed data in the form of a well structured table is available. **Cons:** In some of the datasets, vital clinical and demographic information of samples is missing.	[12]
NCBI-SRA	It maintains raw sequencing data and alignment information from multiple high-throughput sequencing platforms. It encompasses over 445K experiments for cancer.	https://www.ncbi.nlm.nih.gov/sra/	**Pro:** It is largest repository encompassing raw sequencing reads data provides opportunity for novel research. **Con**: The complexity (raw) and size of data (which is usually in gigabytes [Gb]) in challenging for noncomputational users.	[14]
GDC	Multi-omics (WGS, RNA-Seq, WXS, miRNA-Seq,	https://portal.gdc.cancer.gov/	**Pro:** Complete clinical and demographic information is	[25]

	methylation, ATAC-Seq, etc.) and clinical data available for nearly 33 primary cancer types.		available for each sample. **Cons:** Programming skills are required to generate ready to use expression data matrix since data is available for each individual sample. Data is mainly from US population	
ICGC	Genomic aberrations data for 22 primary cancer types from 86 Cancer projects. It encompasses 81,782,588 somatic mutations data for analyses and visualization from nearly 24K individuals.	https://icgc.org/	**Pro:** It contains data from various regions of the globe, i.e. US, Europe, UK, Asia, Canada, etc. **Con:** Data is taken from limited resources.	[26]
ArrayExpress	MIAME supportive microarray data and annotated gene-expression experimental data from microarray and NGS-high throughput genomics experiments. It holds 18,834 experiments data for cancer samples.	https://www.ebi.ac.uk/arrayexpress/	**Pro:** Both raw and processed data available. Data available in ready to use matrix **Con:** Maintain large amount of duplicated data from the GEO.	[27]

UCSC Cancer Genomics Browser	Integrative analysis and visualization of cancer genomic and clinical data from multiple resources, including TCGA, ICGC, GDC, Pan-Cancer Atlas, UCSC RNA-seq compendium, TCGA ATAC-seq and literature. It encompass data for more than 82k cancer samples obtained from different resources	https://xena.ucsc.edu/	**Pro:** Large number of datasets is available**.** It contains ready-to-use dataset matrices **Con:** Primarily contain data from other resources	[28]
T3CA	The 3' UTR cancer atlas is comprehensive resource of alternative polyadenylation (APA) usage for 10,537 tumors across 32 cancer types.	http://tc3a.org	**Pro:** It allows researchers to explore and visualization of APA events in human cancers. **Con:** It encompass only TCGA data.	[29]
ChimerDBv3.0	An enhanced database for fusion genes from cancer transcriptome and literature data mining. It holds a total 33,316 fusion gene pairs, out of which 10% are known in	http://ercsb.ewha.ac.kr/fusiongene/	**Pro:** Information is available in tabulated form **Con:** 90% of the fusion pairs are predicted	[30]

	literature and 90% are predicted from 4569 patients in 23 cancer types of TCGA RNA-seq data.			
LinkedOmics	A resource for exploring the multi-omics data for 32 cancer types and a total of 11,158 patients from the TCGA project. It is also integrates mass spectrometry (MS)-based global proteomics data for selected tumor samples obtained from CPTAC.	http://www.linkedomics.org	**Pro:** It is an unique tool for disseminating data from large-scale cancer omics projects. **Con:** It contains only data from TCGA and CPTAC.	[31]
TCGASpliceSeq	Compendium of alternative mRNA splicing in cancer. It holds Percent Spliced In (PSI) values for 80,000+ splice events on 19,036 genes for 6238 tumor samples and 496 adjacent normal samples for 15 tumor types.	http://bioinformatics.mdanderson.org/TCGASpliceSeq	**Pro:** It is graphical web-based application that can be queried and explored for investigation of splicing sites in cancer samples. **Con:** Mainly data was taken from the TCGA and not updated after its development.	[32]

COSMIC	Resource for somatic and germline mutations in cancer. The latest version, COSMIC v91 (April 2020), includes a total of 9,917,630 gene expression variants, 34,657,730 coding mutations (Mutation Id), 11,453,569 coding mutations (Legacy COSM), 15,156,086 non-coding variants, 21,901,140 genomic mutations, 19,396 fusions, 1,207,390 CNV, obtained from 1,443,198 tumour samples, curated from 27,496 publications.	https://cancer.sanger.ac.uk/cosmic/	**Pro:** The COSMIC is a comprehensive resource that provides overall the aetiology and landscape of mutations in human carcinoma to facilitate in designing the highly sensitive and specific cancer signatures and drivers. It is continuously updated with the new data.	[15]
canSAR	Multi-omics data - genomics, proteomics, protein network & interactions, pharmacological and chemical data with structural biology. It contains the entire human proteome	http://cansar.icr.ac.uk/	**Pro:** It is most comprehensive repository that holds information from multiple disciplines in the field of cancer research.	[33]

	(20,375 sequences) and >542,000 non-human sequences and holds >518,000 protein chains from >171,000 PDB structures. It encompasses multi-omics profiling data on >25 000 cancer patients of 26 cancer types.			
SNP500Cancer	A public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. It maintains >13 400 SNPs, 9124 of which have been sequenced in the SNP500Cancer population.	http://snp500cancer.nci.nih.gov	**Pro:** It can provide vital information on SNPs in cancer with high confidence. **Con:** It has limited data.	[34]
canEvolve	Multi-omics data from 90 cancer studies. It stores data for genomics, transcriptomics, protein expression profiles, miRNA, copy number variations, protein-protein	http://www.canevolve.org/	**Pro:** It presents data in graphical format for researchers.	[35]

	interaction data for nearly 10,000 patients from 90 cancer genomics studies.			
DriverDBv3	It maintains multi-omics data, i.e. somatic mutation, RNA expression, miRNA expression, methylation, copy number variation and clinical data in addition to annotation for cancer driver genes.	http://ngs.ym.edu.tw/driverdb	**Pro:** Information is presented in a graphical user friendly interface. It is also updated within couple of years. Con: Repository is mainly based on the TCGA Data	[36]
CR2Cancer	A resource for exploring the chromatin regulator (CR) in human cancers. It maintains genomic, transcriptomic, proteomic, clinical and functional information for over 400 CRs across multiple cancer types	http://cis.hku.hk/CR2Cancer	**Pro:** It provides a user-friendly web interface to conveniently browse, search and download data of interest.	[37]
GDSC	It manages extensive information on drug sensitivity and drug response among	https://www.cancerrxgene.org/	**Pro:** In the GDSC, user can query drug sensitivity with different keywords like cancer	[38]

	cancer cells. It manages drug sensitivity information for 200 anticancer drugs across 1000 cancer cell lines.		gene, compound and cancer cell line in graphical formats. **Con**: It maintains only cell line data.	
CCLE	The Cancer Cell Line Encyclopedia (CCLE) encompasses gene expression, mutation, RNAseq, methylation, metabolomics data and pharmacologic characterization of human cancer cell lines. Currently, it maintains 136,488 unique datasets for more than 1450 cancer cell lines.	https://portals.broadinstitute.org/ccle	**Pro:** It holds a comprehensive genetic and pharmacological data for large panel of human cancer cell lines in a user friendly interface. **Con:** It maintains only cell line data.	[39]
Omicseq	Search engine for exploring the omics datasets. It contains 50,484 unique, high quality genome-wide profiling datasets collected from major international resources including: 36,694 datasets from	http://www.omicseq.org	**Pro:** It maintains ready to use processed datasets for cancer patients. **Con**: It is not updated after it's development and website is unstable.	[40]

	TCGA, 3,935 from ENCODE and 2,331 from Roadmap Epigenome, 2,079 from the CCLE, 661 from ICGC, 660 from GEUVADIS.			
CircRiC	An interactive web portal, provides comprehensive association of circRNAs with multi-omic data, including mRNA, proteomic, mutation, and drug sensitivity, across nearly1000 cancer cell lines. It maintains 2649 significant circRNA–protein associations, 9604 circRNA–mRNA associations, and 117,258 circRNA–mutation association data.	https://hanlab.uth.edu/cRic	**Pro:** User-friendly web interface with 4 interactive modules: expression landscape, biogenesis, drug response, and integrative analysis for analyzing relation of cirRNA with multi-dimensional data. **Con:** It maintains only cell line data.	[22]
CircNet	It maintains tissue-specific circRNA expression profiles and circRNA–miRNA-gene regulatory networks based on	http://circnet.mbc.nctu.edu.tw/	**Pro:** Interactive platform to analyze circRNAs in human tissue samples. **Con:** It maintains only limited	[21]

	464 RNA-seq samples from 26 human tissues for 104 disease conditions, including cancer. It provides information on novel circRNAs, integrated miRNA-target networks, expression profiles of circRNA isoforms, genomic annotations and sequences of circRNA isoforms (282,948 exon positions).		amount of data.	
Manually Curated Resources for Cancer Research				
Liverome	Manually curated database of liver cancer-related gene signatures with self-contained context information. It holds 143 signatures involve 6,927 genes obtained from 98 HCC studies.	http://liverome.kobic.re.kr	**Pro:** It maintains comprehensive information of liver cancer associated signatures. **Con:** It is not updated after it's development in 2010. Besides, data taken from the Pubmed only	[41]
CancerPDF	A repository of cancer-	https://webs.ii	**Pro:** It encompasses	[42]

	associated peptidome found in human biofluids. It holds 14,367 entries with 9,692 unique peptide sequences for nearly 27 cancer conditions.	itd.edu.in/raghava/cancerpdf/	comprehensive information for more than 14,000 peptides in a user friendly tabular fashion. **Con:** Data taken from Pubmed only and it's not updated after it's development in 2017.	
CancerPPD	Resource for anticancer peptide (ACP) and proteins. It manages 3,491 and 121 entries for ACP and anticancer proteins, respectively.	https://webs.iitd.edu.in/raghava/cancerppd/	**Pro:** It maintains comprehensive information on 3491 entries of anti-cancer peptides. **Con:** Tertiary structure of some peptides not maintained and not updated after the development.	[43]
CBD	Manually curated biomarker database for colorectal cancer. It encompasses 870 identified CRC biomarkers	http://sysbio.suda.edu.cn/CBD/	**Pro:** User friendly platform to query and download data. **Con:** It maintains only selected data, i.e. taken from Pubmed. Other relevant databases, i.e. Scopus, EBSCO, OVID, Web of Science, and EMBASE.	[44]
cBioPortal	An open platform for exploring multidimensional	http://cbioportal.org	**Pro:** Graphical and tabular representation of processed	[45]

	cancer genomics data. It manages data from more than 13,000 tumor samples from nearly 40 datasets from TCGA and other large-scale genomic studies.		multi-omics data.	
HCMDB	Repository maintains cross-platform transcriptomics dataset of human metastatic cancer. It manages gene expression profiles of 11,425 samples from 455 experiments of 29 primary tumor types.	http://hcmdb.i-sanger.com/	**Pro:** Uniformly processed data is available, which can be easily explored and analyzed by non-computational users. **Con:** It encompass only transcriptomics data and not updated after it's development.	[46]
CancerDR	A comprehensive database that manages information for drug resistance in cancer. It manages comprehensive information of 148 anti-cancer drugs, and their pharmacological profiling across 952 cancer cell lines.	https://webs.iitd.edu.in/raghava/cancerdr/	**Pro:** It allows user to identify promiscuous drug molecules that can kill wide range of cancer cells. **Con:** It maintains only cell line data and not updated after 2013.	[47]

SomamiR	A Knowldegebase of cancer somatic mutations altering microRNA-ceRNA interactions. It holds 512,047 somatic mutations in experimentally identified miRNA target sites and 2,868 677 somatic mutations in predicted miRNA target sites.	http://compbio.uthsc.edu/SomamiR	**Pro:** It maintains vast amount of information on somatic mutations in miRNAs target sites in user friendly manner. **Con**: It is not updated since 2015.	[48]
OncoMX	A resource for exploring cancer biomarkers in the context of related cancer and healthy data. It provides integrated data encompassing more than 1,000 unique biomarker entries (939 from the Early Detection Research Network [EDRN] and 96 from the US Food and Drug Administration) mapped to 20,576 genes that have either	http://data.oncomx.org/	**Pro:** It maintains comprehensive information for cancer biomarkers. **Con:** It is based on text-mining tools.	[49]

	mutation or differential expression in cancer			
ResMarkerDB	A freely accessible database that manages biomarkers of drug response to antibody therapy in colorectal and breast cancer. It holds more than 500 biomarker-drug-tumour associations, covering more than 100 genes.	http://www.resmarkerdb.org	**Pro:** It maintains comprehensive information on 500 biomarker-drug-tumour associations, as well non-coding DNA data in response to drug treatment. **Con:** It is text-mining based database. There is a narrow coverage of cancer types and therapies.	[50]
CancerEnD	Resource that maintains 8524 unique expressed enhancers, associated genes, somatic mutations and copy number variations of 8063 cancer samples from 18 cancer types of TCGA.	https://webs.iitd.edu.in/raghava/cancerend/	**Pro:** Comprehensive information for cancer associated enhancers in a user friendly interface. **Con:** Data is taken from only TCGA samples.	[4]
CancerLivER	A repository of gene expression and biomarker data of liver cancer. It	https://webs.iitd.edu.in/raghava/cancerlive	**Pro:** Ready to use matrix available for liver cancer samples. Biomarker information	[51]

	encompass nearly 115 dataset with more 9,600 samples	r	available in the form of table **Con:** Only RNA and miRNA datasets available. Biomarker data is taken from limited number of resources.	
HCCDB	A free access database of hepatocellular carcinoma expression atlas. It maintains 15 public HCC gene expression datasets containing a total 3,917 samples	http://lifeome.net/database/hccdb	**Pro:** It maintains comprehensive information on HCC genes and gene expression data in a user friendly manner. **Con:** It maintains very limited number of datasets.	[52]
dBMHCC	An HCC biomarker database with a reliable prediction system for novel HCC phosphorylated biomarkers. It maintains 5,068 biomarkers, out of which 611 are experimentally validated and rest are predicted, while it holds a total 1,280 HCC phosphorylated markers, out	http://predictor.nchu.edu.tw/dBMHCC	**Pro:** It maintains comprehensive information on HCC biomarker which is presented in simple graphical and tabulated manner. It is developed in June 2020, thus it maintains updated information.	[53]

	of which 203 experimental validated. Besides, it holds information for 234 HCC-related pathways, 17 phosphorylation-related motifs and their 255 corresponding protein kinases.			
ApoCanD	A freely available database of apoptotic proteins related to cancer. It maintains comprehensive information, i.e. mutation status, copy number variation and gene expression levels of 82 apoptosis proteins in context of cancer in tumour samples and cell lines.	https://webs.iitd.edu.in/raghava/apocand/	**Pro:** It maintains comprehensive information regarding Apoptotic cancer proteins from different resources, like CCLE, COLT-cancer database, PDB, PFAM and ANNOVAR, etc. **Con:** It is not updated after it's development in 2016.	[54]
PCMDB	An extensive database that maintains methylation information of genes involved in pancreatic	https://webs.iitd.edu.in/raghava/pcmdb/	**Pro:** Comprehensive information is maintained in a user friendly graphical and tabulated manner.	[55]

	cancer. It maintains 65907 entries for methylation status of 4342 unique genes. Out of which, 53,565 entries for 88 cell lines) and 12,342 entries for 3078 tissue samples.		**Con:** It is not updated after it's development in 2014.	
CCDB	A curated resource on genes involved in cervical cancer. It hold comprehensive information of 537 genes that are linked with cervical cancer causation processes such as methylation, gene amplification, mutation, polymorphism and change in expression level.	https://webs.iitd.edu.in/raghava/ccdb/	**Pro:** It maintains comprehensive information in a user friendly graphical and tabulated manner. **Con:** It is not updated after it's development in 2010.	[56]
Circ2Disease	It holds a total 273 manually curated associations between 237 circRNAs and 54 human diseases including cancer from 120 studies. It provides a user-friendly interface to		**Pro:** It is first database that maintains experimentally validated disease-related circRNAs. **Con:** It encompasses only a limited amount of data.	[16]

	browse, search, analyze regulatory network and download data			
Circ2Traits	A database maintaining 1,951 circular RNA potentially associated with 105 different diseases including cancer and traits. It encompasses the complete putative miRNA-circRNA-mRNA-lncRNA interaction network for each disease.	http://gyanxet-beta.com/circdb/	**Pro:** Comprehensive information is available in a tabulated manner. **Con:** It maintains information for predicted human circular RNA, which is yet to experimentally validate.	[17]
MiOncoCirc	A resource maintains circRNAs that are detected in more than 2,000 tumor samples from tissues, cell line and normal tissue. It capture exome a poly(A)-independent RNA sequencing method data.	https://miococirc.github.io/	**Pro:** It maintains circRNA from cancer cell lines as well as tumor samples	[18]
exoRBase	A resource of circRNA,	http://www.ex	**Pro:** Webserver will assist the	[19]

	lncRNA and mRNA derived from RNA-seq data analyses of human blood exosomes from coronary heart disease (6 samples), colorectal cancer (12 samples), hepatocellular carcinoma (21 samples), pancreatic adenocarcinoma (14 samples) or breast cancer (2 samples). It maintains 58,330 circRNAs, 15,501 lncRNAs and 18,333 mRNAs from 87 blood exosomal RNA-seq samples from 6 datasers.	oRBase.org	community in identifying molecular signatures in blood. **Con:** It maintains only limited amount of data from a total 87 samples across 4 primary cancer types and cardiovascular diseases.	
LncRNADisease 2.0	A database that maintains 10,564 experimentally supported lncRNA-disease associations, 1004 circRNA-disease association and 195,395 computationally identified lncRNA-disease	http://www.rnanut.net/lncrnadisease/	**Pro:** User-friendly webserver maintains extensive information, i.e. gene symbol, gene category, disease information, regulatory relationship for vast amount of lncRNA-disease association. **Con:** Large portion (94%) of	[20]

	associations across 4 species (Human, mice, rat, red junglefowl) for various diseases including cancer (nearly 44% of data).		database is computationally supported lncRNA-disease associations.	