Can SAR Database: An Overview on System, Role and Application

Pooja Deolankar[1], SulakshanaVispute[2]

[1](Computer Applications, Navinchandra Mehta Institute of Technology & Development, India)
[2](Computer Applications, Navinchandra Mehta Institute of Technology & Development, India)

ABSTRACT: *The intention of this paper is to provide an technical overview on the largest cancer database, the canSAR database system. This overview includes the basic definitions and terminology, findings and advancements infield of cancer research through canSAR database, with basic system architecture, design, data source, processing pipelines, screening tests and structure activity relationship of system.*
Keywords: canSAR database, processing pipelines, screening test, structure activity relationship.

I. INTRODUCTION

Cancer, is when abnormal cells divide in an uncontrolled way. Some cancers may eventually spread into other tissues. There are more than 200 different types of cancer. Cancer starts when gene changes make one cell or a few cells begin to grow and multiply too much. This may cause a growth called a tumor [1].
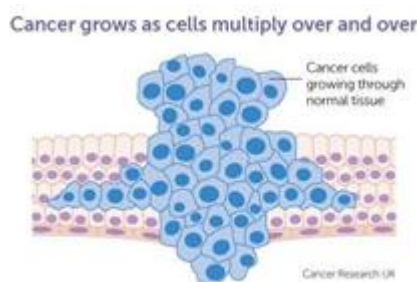
Figure 1 Cancer grows as cells multiply over and over [2]

There are 5 main categories of cancer according to the type of cell they start from, Carcinoma, Sarcoma, Leukemia, Lymphoma and myeloma, Brain and spinal cord cancers [3]. To constrain the growth of disease such as cancer which may led to painful death has compelled the researchers around the world to do the rigorous, time consuming and quality effective cancer cure research. Due to years of study on curing cancer has made many different treatments available for human, these treatments give hope to patient for longer life and with help of new age medicines patients are also relived from some of their pain. A cancer screening technique has proven to be effective in detecting cancer. Checking for cancer (or for conditions that may become cancer) in people who have no symptoms is called screening. Screening can help doctors find and treat several types of cancer early. Early detection is important because when abnormal tissue or cancer is found early, it may be easier to treat. Several screening tests have been shown to detect cancer early and to reduce the chance of dying from that cancer [5]. A few screening tests that can reduce Cancer Deaths are Colonoscopy, sigmoidoscopy, and high-sensitivity fecal occult blood tests (FOBTs) and Low-dose helical computed tomography and Mammography and Pap test and human papillomavirus (HPV) testing [6]. These test and treatments are effective but not absolute. As earlier as the growth of cancer cells is detected in the body, the measures to cure it can be started. By the time symptoms appear, cancer may have begun to spread and is harder to treat [5].

The Advances in large-scale genomics technologies have transformed the drug discovery field by providing unparalleled biological and pharmacological information about potential disease causing genes. Cancer research in particular benefits from these advances as it is often caused by complex genetic events that may not become obvious without large-scale, systematic analysis of multi-genome heterogeneous data [9, 10]. A new age canSAR database is programmed to accept, analyze and integrate cancer related data and provide to cancer cure researchers around the world. The canSAR takes its name from words 'Cancer' and 'SAR' of 'Structure-Activity-Relationship' which is study of the relationship between a drug's molecular structure and the drug's biological activity [11]. A smaller-scale prototype of canSAR attracted 26,000 users in more than 70 countries, and helped identify 46 previously overlooked drug treatment possibilities for cancer molecules [12].

Such several reasons have contributed for the development of one of a kind canSAR database. The latest employed version is canSAR v2.0 aiding to cancer research globally.

II. CANSAR DATABASE

canSAR is an integrated database that brings together biological, chemical, pharmacological (and eventually clinical) data. Its goal is to integrate this data and make it accessible to cancer research scientists from multiple disciplines, to aid hypothesis generation and decision making towards identifying causes, biomarkers and therapeutics for cancer. canSAR name reflects the fact that it integrates cancer-relevant data with compound Structure-Activity-Relationship (SAR) data [7]. The database was developed by researchers at The Institute of Cancer Research, London. Team leader Dr.Bissan Al-Lazikani said the database uses findings from laboratories, patient research, genetics and chemistry studies to perform "extraordinarily complex virtual experiments". "It can spot opportunities for future cancer treatments that no human eye could be expected to see," Dr Al-Lazikani said. The new canSAR database will cope with a huge expansion of cancer data following advances in technologies such as DNA sequencing. It contains information about nearly one million testable drugs and over a thousand types of cancer cell. Research that had previously taken months to complete could now be performed in minutes. Professor Paul Workman, deputy chief executive of The Institute of Cancer Research, said: "This is an extraordinary time for cancer research, as advances in scientific techniques open up new possibilities and generate unprecedented amounts of data." "Our aim is to make this wealth of information, coming from both the clinic and from the laboratory, freely available in a very user-friendly form to as many people as possible." [12].

The resource is being made freely available by The Institute of Cancer Research (ICR) and Cancer Research UK, and will help researchers worldwide make use of vast quantities of data, including data from patients, clinical trials and genetic, biochemical and pharmacological research. "This database speaks many different languages - a chemist and a clinician can access data from each other without having to understand each other's jargon. It is so easy to use that anyone can have a go - I fully envisage a bright A-level student using it and in the future, that might even be where we see fresh ideas coming from." [8].

A. Cancer detection and death statistics

The International Cancer Genome Consortium profiles up to 20,000 cancer patients and the world's largest single database of cancer patients has launched on 11th November 2013. It will combine near real-time cancer data on the 350,000 cancers diagnosed each year in England, along with detailed clinical information and over 11m historical cancer records [13]. Every day in the UK there are more than 400 people diagnosed with cancer that will survive the disease for more than 10 years thanks to research [14]. Statistics shows that cancer is a leading cause of death worldwide, accounting for 8.2 million deaths in 2012 [4]. The most common causes of cancer death are cancers of: lung (1.59 million deaths), liver (745 000 deaths), stomach (723 000 deaths), colorectal (694 000 deaths), breast (521 000 deaths), esophageal cancer (400 000 deaths) [4].

III. CANSAR DATA SOURCE, DATABASE DESIGN AND PROCESSING PIPELINE

A self-sufficient intelligent system like canSAR is implemented on web by running it on an Apache web server implemented in PHP, JavaScript, Perl and Java. The data reside in an Oracle 11g database. Chemical compound search and handling is supported by the Accelrys direct cartridge. The data processing pipelines are written in Perl, Python and Java and utilize Openbabel, CDK [19] and Pipeline Pilot (Accelrys Inc) [20].

canSAR data comes from a wide variety of public sources and publications. The ICR internal version of canSAR also includes internally generated experimental data. For latest data sources and statistics. The data and minor functionality update is done on a monthly basis. However, most of the data in canSAR comes from public sources which have different update cycles. The Data Sources update status report is made public on website. The updating of 3D structures happens every week, a couple of days after Protein Data Bank (PDB) release new structures in order for the researchers at Institute of Cancer Research to provide additional annotations and integrate this data with the respective chemical and biological information in canSAR. The organisms that does canSAR store data on, canSAR target dictionary focusses on human data, although we have a large number of model organism targets. Target key word searches only search human targets. Chemical screening data is from a wide variety of organisms [7].

The design and implementation of canSAR has been entirely use-case driven. canSAR was designed in a modular manner such that the modelling and inclusion of new data types in the future [e.g. clinical outcome or drug metabolism and pharmacokinetics (DMPK) data] is relatively easy. This provides the ability and support to expand the functionality and usefulness of canSAR as new requirements are found. The data model is designed around elemental and association data types. Elemental types (e.g. molecular target, compound, and cell) represent the core components of the system. These elements can then be associated via different association types, e.g. a compound and a molecular target can be associated via a 'screening' association, a 'drug action'

association, or a '3D-structural complex' association. The associations in turn are then linked to the experimental data, sources of origin and publication references where relevant. These generic key data types provide the flexibility described above.

Data sources from collaborators are mirrored to ensure that canSAR is up-to-date. In order to process these external sources into a format suitable for integration into canSAR, a suite of pipelines have been created which allows the seamless integration of new data from existing sources, and most importantly, check for uniqueness among the different data sources. A fraction of the data within canSAR is duplicated in several data sources and it is imperative that this redundancy is recorded and that the useful information for each duplicate is extracted. In some cases, this is reasonably trivial, but in others (compounds and molecular targets are good example), accurate and appropriate tests for uniqueness and data storage are challenging. Most of the data pipelines involved the following steps: acquisitions of local copy, pre-processing into suitable format, uniqueness testing of existing data, loading of data into canSAR and rebuilding of data source logs. Some of the data sources are updated monthly, however some are updated morefrequently or at variable intervals. Each of the specific pipelines is adapted to suit the needs of the specific data source. Where possible and appropriate, a source's API is utilized through the use of simple REST or SOAP services. This has the benefit of ensuring that the data are as up-to-date as possible and removes steps of mirroring and integrating.
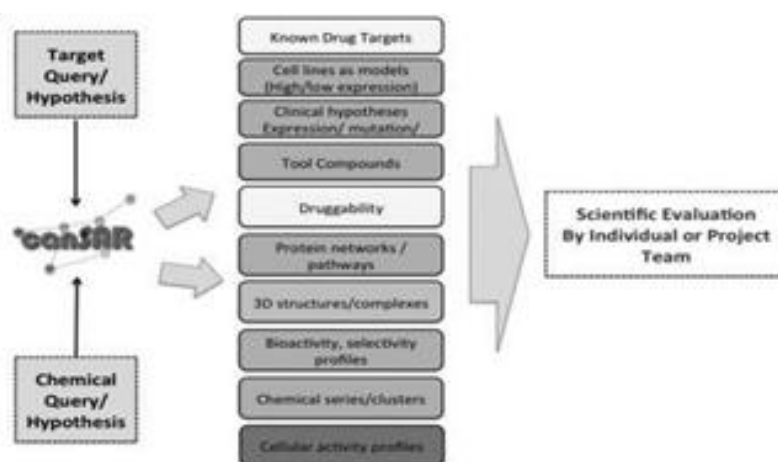

Figure 2 Process flow of Common requested information

To ensure long-term growth and utility, canSAR provides flexibility for new data and data types. Providing the centralization and ability to rapidly interrogate such disparate data is of immediate benefit to users who can now access data in seconds that would have taken days or weeks to compile; some common information requests taken into consideration in the design of canSAR.Common requested information in cancer translational research used to design the user interface. Query can be from a biological or chemical starting point. Users often require complex, interconnected but concise information. The data reports generated in canSAR are used to underpin subsequent discussions and the design of the next experiment [15].

IV. DATA FLOW

In simple working canSAR database links the raw goldmines of genetic data to a whole raft of independent chemistry, biology, and patient and disease information. It collates billions of experimental results from around the world including ones on the presence of genetic mutations, the levels of genes and their resultant proteins in a tumor, and the measured activity of a compound or drug on tested proteins.The system then "translates" these data into a common language so that they can be compared and linked. It can even explore the patterns of interaction between proteins in a cell using similar systems that are used to explore human interactions in social networks. Once these masses of data are collated and translated, canSAR then uses sophisticated machine learning and artificial intelligence to draw paths between them, predict risks and make drug-relevant suggestions that can be tested in the lab. It's a bit like predicting the likely winners of a 100m Olympic race. The computer first "learns" the important factors from past race winners such as cardiovascular fitness, muscle mass, past performance, their training schedule, and then it uses this learning to rank new athletes based on how well they fit the profile of winners. Using canSAR potential cancer targets can be spotted by bringing lots of sources of existing data togetherin one place and deciphering important properties from previous successful drug targets [13].

V. DATA CONTENT AND GROWTH

canSAR contains the full complement of the human proteome as well as 16 332 proteins from 2136 model organisms that fall into 8631 major protein families. It contains annotations and data for >11 000 cell lines, both cancer and non-transformed cell lines, and 3690 patient-derived tissue samples with >10 000 experimental result sets. The fully searchable 2D structure and annotation for nearly one million small molecule drugs and chemical probes have collectively >8 000 000 experimental bioactivities as well as 10 million calculated properties. There are >93 000 3D structures for 31 130 proteins, collectively containing 16 700 ligands determined in complex with a protein. We have annotated and classified these ligands into different classes as described below and identified ca 16 000 ligands that are functionally relevant. All ligands have 3D ligand-interaction maps in canSAR. In addition to the collated and annotated data, canSAR contains curation and additional analyses including 4544 3D-structure–based, 8197 Ligand-based and 9446 protein-network–based draggability results for human proteins. Importantly, all data from all areas of research are seamlessly integrated and are fully referenced to their original sources and specific publications where available to ensure that researchers can rapidly identify the original source of the information without complex and lengthy searching [16].

The database is modular and extensible to allow for future data growth. The public version of canSAR (v1.0) contains ~8 Million experimentally derived measurements, ~700,000 unique biologically active chemical structures and data for >1,000 cancer cell lines. These data are collated from a number of public sources, and collectively annotated to ensure seamless integration. canSAR will also contain annotated molecular target data representing the human genome and a number of model organisms. Context driven data are generated 'on the fly' from collaborator databases such as ChEMBL, ROCK and Array Express [29].

VI. DATA EXPORT AND SHARING AND WEB VISUALIZATION

Users often wish to export specific analysis results obtained from the interface. In most sections of canSAR, users are able to export the current results in various formats including Excel, tab-delimited text, SDF and MIABE (Minimum Information About a Bioactive Entity) [17] compliant XML [18]. As for web visualization, canSAR integrates an unprecedented set of translational research data, with its major unique advantage of the presentation and interrogation tools. canSAR is available via http://cansar.icr.ac.uk. Users can obtain logical and concise summaries of broad state-of-the-art knowledge easily. The interface provides biological annotation, gene expression, disease association, structural and pharmacological data, and produces graphical and tabular summary reports pertaining to any aspect of these data.

Additionally, wider ranging and more expert-style questions can be asked of the data using the expert and batch query tools. The interface development is driven by typical use-cases in cancer translational research, drug discovery and chemical biology and caters for users from different backgrounds (biology, chemistry, clinical etc.). Example entry points include target/cell line/compound keyword searches, sequence similarity searches, compound structure searches and the facility to upload lists of identifiers as a query. Expert tools allow retrieval of extensive chemogenomic annotations, polypharmacology maps, compound selectivity and bioactivity profiles, expression details and pathway enrichment analysis [18].

VII. BROWSING AND INFORMATION RETRIVEL

canSAR has a single global search capability that enables keyword searches to be performed across the system, thus speeding up the retrieval of data. Additional, object-specific search capabilities such as protein sequence and chemical structure searches remain in place.
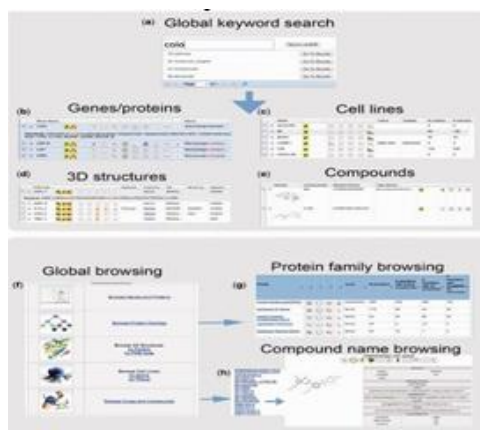

Figure 3 Global keyword searching [22]

Above figure shows (**a**) the single global search capability that enables keyword searches to be performed across (**b**) genes and proteins, (**c**) cell lines (**d**) 3D structures and (**e**) compounds. The results are displayed in tabular forms with icons representing the availability of data such as cancer mutations and bioactive chemical probes. The user can sort the results based on these data. (**f**) New browsing functionality that allows exploring database's content through browsing genes and proteins, protein families, 3D structures or drugs and compounds. For example, (**g**) protein families have a summary feature where the user can sort and select families based on the types of data available. (**h**) Molecular targets, compounds and cell lines may be browsed by name.

Browsing canSAR data can be accomplished through protein/gene name alphabetical dictionaries, protein-families, 3D structure and compound dictionaries. These new searches and browsing capabilities allow users to explore the data in canSAR in a broader way, e.g. the user can rank members of a particular protein family based on the availability of reported cancer mutations, bioactive chemical probes, draggability orcan browse through all cell lines from a particular tissue type [21].

VIII.   TARGET SYNOPSIS CAPABILITY

canSAR retrieves instant summaries about a particular target of interest. To enhance this ability, a new 'wiki'-style summary page that distils the key information about the target in a human-readable summary form and covers function, cellular localization, drugs and clinical candidates, cancer mutations, RNAi data, expression levels in cancer and non-transformed cell lines and drug or chemical probe bioactivity information among others is developed. Each section is linkable and subject to exploration with drill-down capabilities. The cell-line matrix displays all available information for the specific target in all cell lines, so in one table, the user can see the mutation state, expression level, and, where it exists, RNAi information. Then the interactive protein-interaction networks provide a powerful resource that not only identifies protein interactions of the target of interest, but also provides chemical biology annotation and genomic information in an 'at-a-glance' form so that the user can immediately identify any drug targets in the network, or targets with known bioactive chemical probes, that are RNAi screen hits or that have known mutations in cancer. Each protein in the network is colored according to its chemical tractability, thus highlighting proteins within the immediate network of the target of interest that would be most amenable to drug discovery. canSAR database also provides advanced functionalities for Cell-Line synopses and Protein family synopses [22].

canSAR uses different analysis tools for different input targets. Tools used are (a) Polypharmacology Map Tool provides an interaction-style map of the inputted targets where the interactions indicate the number of targets in common that have been screened. (b) Cancer Protein Annotation Tool is a multidisciplinary annotation tool, it generates summary annotations for up to 1000 genes or proteins using the information stored in canSAR. It is intended to be used alongside experimental data to help prioritize gene lists for further exploration. (c) Expression Details Tool provides an interactive hetmap of gene expression for a given list of targets. (d) Pathway/Go and Annotation Tool provides details about pathways, GO terms and other annotations from a set of targets. (e) Alignments and superposition is a tool for visualizing and annotating alignments and superposition's of sequences and structures [23].

IX.   CANSAR-3D

The 3D structure is a powerful tool in understanding themolecular mechanisms of disease and the design anddevelopment of new drugs. As well as having access to the large number of protein structures (>93 000) in the Protein Data Bank (PDB) [24]. canSAR-3D, the 3D structural component of canSAR, contains annotated PDB structures where each structure is linked to the protein and full genomicinformation, on the one hand, and the full chemical information for its ligand on the other. The researchers at the Institute of Cancer Research, have curated the ligands in the PDB into five categories to better distinguish those that are genuine small molecule endogenous ligands or chemical modulators of protein function from biologically unimportant small molecules such as surfactants. The categorization of small molecule ligands is computationally assigned based on the presence of >6 non-Hydrogen atoms. Exceptions to this rule and the surfactants in this list are further classified by manual curation. There is a per-protein structural summary page that graphically represents the structures available for a protein and regions they cover within criteria set by user. 3D structure superposition may also be carried out on-the-fly. The 3D-druggability scores [26, 27] for all identified pockets within each structure are automatically calculated and updated in canSAR on a weekly basis [25].

X.   WORKING BENEFITS AND APPLICATIONS

The impact benefits of this intelligent agent include (a)Single portal for multidisciplinary data, canSAR provides a single-point portal to access broad data available for a gene, protein, and drug or cell line. There new identifying and prioritizing tools for more complex queries such as (b)Identifying tools for probing a target's activity, for a gene or protein of interest, the user can identify available evidence for disease association through

altered mutation or expression in patient tissue on the target synopsis with cancer cell lines that will be useful to use as model systems to study its target, also the user can identify which drugs have shown activity against their target or their model cell lines using available data. (c) Prioritizing lists of genes for further experiments, tool provides the user with a spreadsheet summarizing the evidence for draggability, and the availability of structural and chemical probes. (d) Multidisciplinary network analysis, protein interaction network is used to identify druggable targets and to check whether network neighbors have been reported as hits in a cancer RNAi screen or whether they are mutated in cancer [28].

XI. AIMS AND ACHIEVEMENT

The aim of researchers at the Institute of cancer Research, is to empower cancer translational research and drug discovery by using computational techniques to bridge the gap between biological, chemical and clinical knowledge. By developing and applying novel computational tools and approaches to integrate biological, chemical and clinical data at a large scale. These tools are employed in the following areas: (a) Using chemogenomic data to support decision-making in the drug discovery process, from target selection to lead identification. This helps identifying opportunities and risks early on in the process, thus helping streamline novel drug discovery.(b) Using clinical, RNAi screening and high-throughput biology data to identify targets for cancer. (c) Analyzing protein interaction networks for optimal therapy intervention points. (d) Identifying patterns and rules in targets andnetworks which yield successful drugs, and applying these rules specifically in the complex field of cancer drug discovery. (e) Leveraging large-scale structural biology and Structure Activity Relationship (SAR) data in determining target tractability and in aiding drug design. (f) Understanding rules governing ligand binding and target/ligand selectivity and using them predicatively in assessing target and compound development. (g) Utilizing published clinical outcome data to enhance the assessment and understanding of drug/target activities. (h) Aiding experimental hypotheses by providing information on suitable cellular test system and chemical probes.

These aims are achieved through four major areas of research and development within the group,

(a) Integrated chemogenomic tool, e.g., canSAR, canSAR-3D.
(b) Objective target assessment and prioritization.
(c) Mapping 'druggable' protein networks for therapy and overcoming drug resistance.
(d) Three-dimensional structural analysis of cancer-associate protein families [30].

XII. CONCLUSION

The recent advances in genomics technology and Structure-Activity-Relationship (SAR) data applications have integrated data from multiple disciplines, to aid hypothesis generation and decision making towards identifying causes, biomarkers and therapeutics for cancer research. This resource has brought together all cancer researchers around world and has enabled a chemist, a clinician, a doctor to work on same platform, exchange data, analyze multidimensional data originating from different data sources applying them to advanced analytical tools and provide a human-readable result statement for user regardless of their medical disciplinary background in real-time. The latest updates, results, advancements are made publicly available by canSAR's public accessible data repository. These advances are largely transforming the online data analysis, drug testing and drug match synopses, so as researchers can work more effectively on their research in much less time with canSAR 's ability to rapidly answer complex scientific queries. Now with new 3D structuring analysis a better viewing and analyzing capability is provided to canSAR database. The more advance analyzing tool functionality is needed to identify and correlate the patient's genetic data and mutation. The advanced visualization tools are substantial for real-time analysis of causes, mutations and effects of new drugs on these genes similar to a clinical trial.

REFERENCES

[1]. Cancer Research UK, on what is cancer http://www.cancerresearchuk.org/about-cancer/what-is- cancer
[2]. Cancer Research UK, cancer cells multiply over and over http://www.cancerresearchuk.org/about-cancer/what-is- cancer
[3]. Cancer Research UK, on Types of cancer, 5 main categories http://www.cancerresearchuk.org/about-cancer/what-is- cancer/how-cancer-starts/types-of-cancer
[4]. World Cancer Report 2014, IARC Nonserial Publication, Stewart, B.W., Wild, C.P. IARC http://apps.who.int/bookorders/anglais/detart1.jsp?codlan=1&codcol=76&codcch=31
[5]. National Cancer Institute, Cancer Screening http://www.cancer.gov/about-cancer/screening
[6]. National Cancer Institute, Screening Tests http://www.cancer.gov/about-cancer/screening/screening-tests
[7]. Institute of Cancer Research, Frequently Asked Questions https://cansar.icr.ac.uk/cansar/frequently-asked-questions/#faq1
[8]. Institute of Cancer Research, World's largest disease database will use artificial intelligence to find new cancer treatments http://medicalxpress.com/news/2013-11-world-largest- disease-database-artificial.html
[9]. Schlabach M, Luo J, Solimini N, Hu G, Xu Q, Li M, Zhao Z, Smogorzewska A, Sowa M, Ang X, et al. Cancer proliferation gene discovery through functional genomics. Science. 2008;319:620–624.[PMC free article] [PubMed]
[10]. Zheng-Bradley X, Rung J, Parkinson H, Brazma A. Large scale comparison of global gene expression patterns in human and mouse. Genome Biol. 2010.11:R124. [PMC free article] [PubMed]

[11].   Illustrated Glossary of Organic Chemistry http://www.chem.ucla.edu/harding/IGOC/S/structure_activity_relationship.html
[12].   Institute of Cancer Research, World's biggest cancer database to aid new treatment development, Copyright Press Association 2013 http://www.cancerresearchuk.org/about-us/cancer-news/news-   report/2013-11-11-worlds-biggest-cancer-database-to-aid-   new-treatment-development
[13].   Article originally published at ' The Conversation'. Institute of Cancer Research, Artificial intelligence uses biggest disease database to fight cancer http://www.icr.ac.uk/blogs/science-talk-the-icr-blog/page-details/artificial-intelligence-uses-biggest-disease-database-to- fight-cancer
[14].   Institute of Cancer Research, Our research by cancer type http://www.cancerresearchuk.org/our-research/our-research-by-cancer-type
[15].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: an integrated cancer public translational research and drug discovery resource http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245005/#gkr 881-B1
[16].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: updated cancer research and drug discovery knowledgebase http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964944/
[17].   Orchard S, Al-Lazikani B, Bryant S, Clark D, Calder E, Dix I, Engkvist O, Forster M, Gaulton A, Gilson M, et al.
[18].   Minimum information about a bioactive entity (MIABE) Nature Biotech. 2011;10:661–669.[PubMed]
[19].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: an integrated cancer public translational research and drug discovery resource http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245005/
[20].   Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): an open-source java library for chemo- and bioinformatics. J. Chem. Informat. Comput. Sci.2003;43:493–500. [PubMed]
[21].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: an integrated cancer public translational research and drug discovery resource http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245005/#gkr 881-B41
[22].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: updated cancer research and drug discovery knowledgebase http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964944/
[23].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: updated cancer research and drug discovery knowledgebase,   Nucleic   Acids   Res.   2014 Jan; 42(Database   issue):   D1040–D1047. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964944/
[24].   The   Cancer   Research   UK   Cancer   Therapeutics   Unit,   at   The   Institute   of   Cancer   Research,   Tools https://cansar.icr.ac.uk/cansar/tools/#main_tab_holder:tab_too ls_main:tab_tools_1
[25].   Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlic A, Quesada M, et al. The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res. 2013;41:D475–D482. [PMC free article] [PubMed]
[26].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: updated cancer research and drug discovery knowledgebase http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964944/#gkt 1182-B12
[27].   Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, et al. ChEMBL: a large-scale bioactivity database for drug discovery.Nucleic Acids Res. 2012;40:D1100– D1107. [PMC free article] [PubMed]
[28].   Patel MN, Halling-Brown MD, Tym JE, Workman P, Al-Lazikani B. Objective assessment of cancer genes for drug discovery. Nat. Rev. Drug Discov. 2012;12:35–50. [PubMed]
[29].   Articles from Nucleic Acids Research are provided by courtesy of Oxford University Press, canSAR: updated cancer research and drug discovery knowledgebase, Using canSAR http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3964944/#gkt 1182-B12
[30].   The Institute of Cancer Research, canSAR: Integrated Cancer Drug Discovery Platform, by DrBissan Al-Lazikani, Computational Biology &Chemogenomics team, http://www.icr.ac.uk/our-research/researchers-and-teams/dr-bissan-al-lazikani/resources
[31].   The Institute of Cancer Research, Research overview, DrBissan Al-Lazikani, Computational Biology &Chemogenomics team, http://www.icr.ac.uk/our-research/research-divisions/division-of-cancer-therapeutics/computational-iologychemogenomics/research -overview.