



Title	CR2Cancer: a database for chromatin regulators in human cancer
Author(s)	RU, B; SUN, J; TONG, Y; Wong, CN; Chandra, A; Tang, ATS; Chow, LKY; Wun, WL; Levitskaya, Z; Zhang, J
Citation	Nucleic Acids Research, 2018, v. 46 n. D1, p. D918-D924
Issued Date	2018
URL	http://hdl.handle.net/10722/249561
Rights	This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

CR2Cancer: a database for chromatin regulators in human cancer

Beibei Ru, Jianlong Sun, Yin Tong, Ching Ngar Wong, Aditi Chandra, Acacia Tsz So Tang, Larry Ka Yue Chow, Wai Lam Wun, Zarina Levitskaya and Jiangwen Zhang*

School of Biological Sciences, The University of Hong Kong, Hong Kong 999077, China

Received August 14, 2017; Revised September 16, 2017; Editorial Decision September 19, 2017; Accepted September 28, 2017

ABSTRACT

Chromatin regulators (CRs) can dynamically modulate chromatin architecture to epigenetically regulate gene expression in response to intrinsic and extrinsic signalling cues. Somatic alterations or mis-expression of CRs might reprogram the epigenomic landscape of chromatin, which in turn lead to a wide range of common diseases, notably cancer. Here, we present CR2Cancer, a comprehensive annotation and visualization database for CRs in human cancer constructed by high throughput data analysis and literature mining. We collected and integrated genomic, transcriptomic, proteomic, clinical and functional information for over 400 CRs across multiple cancer types. We also built diverse types of CR-associated relations, including cancer type dependent (CR-target and miRNA-CR) and independent (protein-protein interaction and drug-target) ones. Furthermore, we manually curated around 6000 items of aberrant molecular alterations and interactions of CRs in cancer development from 5007 publications. CR2Cancer provides a user-friendly web interface to conveniently browse, search and download data of interest. We believe that this database would become a valuable resource for cancer epigenetics investigation and potential clinical application. CR2Cancer is freely available at <http://cis.hku.hk/CR2Cancer>.

INTRODUCTION

Chromatin biology has been the focus of intense research due to its significant implication in development and pathogenesis (1). Chromatin is a macromolecular complex of DNA and histone proteins forming repeating units of nucleosome, which consists of 147 bp of DNA wrapped around an octamer of histones (two copies of H2A, H2B, H3 and H4) (2). The interplay of epigenetic mechanisms regulates the architecture of chromatin to control DNA-templated processes such as transcription, replication and repair (3,4).

These epigenetic marks mainly include DNA methylation, covalent histone modifications (e.g. acetylation and methylation) and nucleosome placement, which constitute a dynamically changing and complex epigenomic landscape. Chromatin regulators (CRs), a class of enzymes with specialized function domains, can recognize, shape and maintain the epigenetic state in a cell context-dependent fashion (5). For instance, DNA methyltransferases (DNMTs) transfer methyl groups to 5' position of cytosines in CpG dinucleotides while ten-eleven translocation (TET) enzymes catalyze demethylation (6). Histone acetyltransferases (HATs) and deacetylases (HDACs) can deposit and remove acetyl groups to lysine amino acids on histone N-terminal tails, respectively (7). Chromatin remodeling complexes are capable of either moving, evicting or restructuring nucleosomes in an ATP-dependent manner (8).

Genomic alterations or dysregulation of CRs have now been identified in a wide range of cancer types and are increasingly regarded as novel therapeutic targets (9–11). Since CRs can regulate local or global epigenetic patterns to influence multiple target genes simultaneously, aberrant CRs offer cancer cells a highly efficient mechanism to rewire transcriptional regulatory circuits. Recent cancer genome sequencing studies showed that CR genes are among the most recurrently mutated gene sets (e.g. DNMT3A, EP300, CREBBP, ARID1A and SMARCB1), and 25%–30% of the identified driver mutations have an effect on genes encoding CRs (12,13). For example, the histone methyltransferase EZH2 is mutated in 22% of diffuse large B cell lymphoma cases, which directly promotes the trimethylation of histone H3 at K27 mark (H3K27me3) to repress the expression of genes required for cell cycle exit and cell differentiation (14,15). In addition, the reversible nature of epigenetic modifications provides plausible prevention and treatment options for cancer therapy through targeting CRs (16,17). As the first epigenetic therapies approved in 2005, two DNMT1 inhibitors (5-azacytidine and 5-aza-2'-deoxycytidine) can induce global hypomethylation for the treatment of myelodysplastic syndrome (18,19). Subsequently, more than 6 molecules inhibiting the activity of HDACs have also been granted for clinical use in sev-

*To whom correspondence should be addressed. Tel: +852 2299 0341; Fax: +852 2559 9114; Email: jzhang1@hku.hk

eral cancers (20). Apart from these approved drugs, a large number of novel molecules targeting various CRs, such as EZH2, EP300, SIRT1 and LSD1, are currently under development or in clinical trials, which provide an exciting outlook for cancer epigenetic therapies (16,17).

It is therefore clear that comprehensive collection of CRs together with related annotation information would contribute greatly to the study of cancer development and treatment. Currently, there are three major databases developed for CRs and histone modifications. H1stone collects 55 histone proteins, 106 distinct sites of their modifications and 152 histone-modifying enzymes in human (21). WERAM stores experimentally identified and computationally predicted histone regulators as well as site-specific regulator-histone relations in eukaryotes (22). CR Cistrome integrates ChIP-Seq data of CRs and histone modifications to specifically explore their relationships in human and mouse (23). These data resources mainly focus on the collection of basic information and ChIP-Seq data of CRs, and include less knowledge about DNA methylation and chromatin remodeling proteins. Although they are crucial for deciphering the mechanism and function of CRs, a systematic and specialized source for CRs in human cancer is needed urgently.

Over the past decade, large-scale cancer genomic projects, such as The Cancer Genome Atlas (TCGA) (24) and the Cancer Cell Line Encyclopedia (CCLE) (25), have produced multi-dimensional omics data for a number of cancer types. This definitely provides biologists an opportunity to uncover the molecular mechanisms of regulation and interaction in pan-cancer level (26–28). In addition, low-throughput experiments of tumor-associated CRs are rapidly emerging, but hidden in thousands of scattered publications. In this study, we curated, assembled and integrated these data as well as other public resources to build a database CR2Cancer with user-friendly interface for easy data access. Through this database, users can explore the functional, genomic, transcriptomic, proteomic, clinical, biological network and literature-reported information for CRs across multiple cancer types. We believe that CR2Cancer represents a valuable resource that provides unique and useful knowledge of CRs for investigators in cancer epigenetics research community.

DATABASE OVERVIEW

CR2Cancer contains over 400 CRs that are annotated with eleven categories of knowledge (Figure 1). (i) The function category provides information on protein domains, function description and classification as well as substrates and products of chemical reactions that CRs catalyze. In addition, this category shows the Gene Ontology terms and pathways including KEGG (29) and Reactome (30). (ii) The mutation category offers mutation landscape of primary tumor tissues and cancer cell lines in single and pan-cancer level. For each primary tumor type, effect of semantic mutation on protein activity is examined and presented in this category. In addition, we also collected various kinds of mutations from COSMIC (31), and curated mutated CRs from publications. (iii) The post-translational modification (PTM) category allows users to scan the positions, modifications and upstream enzymes of PTMs extracted from

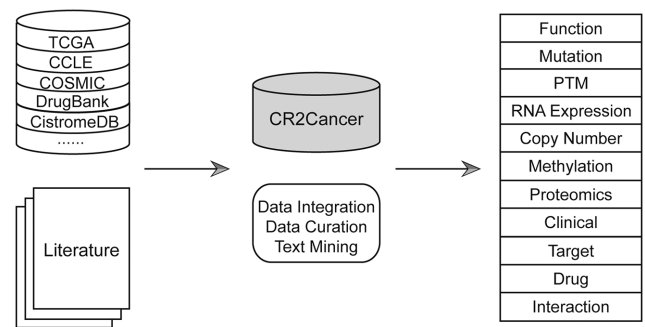


Figure 1. Overview of CR2Cancer. CR2Cancer is composed of eleven categorized annotations for 429 CRs in cancer by high throughput data analysis and literature mining.

UniProt, and explore whether PTMs are affected by mutations. (iv) The RNA expression category shows expression pattern in primary tumor tissues, cancer cell lines, and human normal tissues. For primary tumor types with > 10 corresponding normal samples, the results of differential expression analysis are also shown here. Moreover, this category includes dysregulated patterns of CRs in cancer retrieved by literature mining. (v) The copy number category offers correlation analysis between RNA expression and semantic copy number alteration (SCNA) across various cancer types. In addition, users can examine the percentage of samples for three classes of copy number change, i.e., gain, loss and neutral. (vi) Similar to copy number category, the methylation category offers correlation analysis between RNA expression and promoter methylation level as well as differential methylation status between tumor and normal samples. (vii) The proteomics category is based on reverse phase protein array (RPPA) data for primary tumor tissues and antibody staining level for human normal tissues. (viii) The clinical category shows association tests between RNA expression and clinical features, including subtype, overall survival, stage and grade. (ix) The target category presents transcriptional targets of CRs inferred by reverse engineering method and ChIP-Seq data analysis. (x) The drug category offers drug- and gene-centric networks derived from DrugBank (32) to highlight the druggable nature of CRs. (xi) The interaction category includes three parts: protein-protein interaction (PPI) network with co-occurrence and mutual exclusive mutation analysis, miRNA-target regulatory connections with expression correlation, and multiple types of interactions reported in literatures.

DATABASE CONSTRUCTION

Data collection

A comprehensive list of 429 CRs functioning as DNA modifiers, histone-modifying enzymes or chromatin remodelers, was compiled from three recent papers (33–35). Among them, 167 (38.9%) CRs have been gathered in three cancer gene databases, i.e. 50 in ONGene (36), 72 in TSGene (37) and 100 in NCG (38). We annotated all CRs with full name, aliases, chromosome location and external IDs. In addition, the site-specific substrates and products of CRs were acquired from EpiFactors (39). The somatic muta-

tion, RNA-Seq, DNA copy number, methylation, RPPA and clinical data of primary tumor tissues were downloaded from TCGA; the mutation profiles and gene expression microarray data of cancer cell lines from CCLE; the RNA-Seq and antibody staining data of human normal tissues from The Human Protein Atlas (HPA) (40).

Literature mining

A text mining pipeline was developed to extract cancer-associated CRs from public peer review papers. First, we used ‘cancer’, ‘tumor’ and ‘carcinoma’ as keywords to search PubMed database and downloaded all retrieved abstracts (update to 1 August 2017). Second, these abstracts were split into single sentences using GENIA Sentence Splitter for the input of next step (41). Third, we located gene names in the sentences by using BANNER (42) and normalized the identified gene names based on a dictionary comprised of official names and aliases of CRs. Finally, the CR-mentioned sentences were manually curated to extract aberrant molecular changes (e.g. mutation and dysregulation) of CRs and their interactions with other biological molecules in specific cancer type. The cancer names in our results were normalized to the ‘Preferred Name’ registered in NCI Thesaurus. In current release, CR2Cancer recruits around 4700 aberrant molecular changes and 1300 interactions for 358 CRs from 5007 publications.

Mutation landscape creation

We collected semantic mutation data of 8,788 primary tumor samples in 26 cancer types and 967 cancer cell lines in 24 cancer types. Lollipop mutation diagrams of single nucleotide variation (SNV, e.g. missense, nonsense and silent) were produced in single and pan-cancer level using Lollipops software (43). The mutational distribution of CRs across cancer types were calculated and visualized in pie charts for primary tumor tissues and cancer cell lines, respectively. Moreover, we also retrieved over 120 000 semantic mutations for 412 CR genes from COSMIC v81 data sets.

RNA expression data preparation

For RNA-Seq data of primary cancer types, we normalized raw counts matrix using Voom (44) after filtering out genes that have over one count per million mapped reads in less than 5% samples. Differential expression analysis was carried out by using limma package (45) for 15 cancer types with >10 corresponding normal samples (Fold Change > 1.5 and adjusted *P*-value < 0.05). In addition, the expression level of CRs was visualized across diverse cancer/tissue types, and a total of 1228 images about gene expression were included in the CR2Cancer database.

Correlation between RNA expression and copy number, methylation or clinical features

In each cohort, the percentage of patients with copy number loss, neutral and gain is subject to $-1/2$, 0 and $+1/2$ copy number threshold based on GISTIC output (46). Pearson correlations were used to assess relationships between RNA

expression and copy number value. The methylation level of CRs was represented by the probe in the promoter region with the most negative correlation with expression, and Student's *t*-test was used to examine the differential methylation status between tumor and normal samples. In addition, Spearman correlations were used to assess relationships between expression and methylation, stage or grade. The log-rank test was used to examine whether RNA expression significantly correlated with the overall patient survival times. Kruskal-Wallis test was used to determine if there are significant differences among multiple tumor subtypes on expression.

Identification of transcriptional targets

We identified transcriptional targets of CRs based on two strategies, i.e., reverse engineering and ChIP-Seq data analysis. Using gene expression profiles in cancer, ARACNe (47) was run with default parameters to prioritize cancer-specific CR-target relations based on mutual information and data processing inequality, which can reduce the number of indirect connections. In addition, there now exist several resources for high quality processed ChIP-Seq data of human cell lines and tissues (23,48,49). CistromeDB is by far the most comprehensive one for ChIP-Seq and chromatin accessibility data analyzed through the standard analysis pipeline (49). We downloaded 1359 ChIP-Seq BED files of 139 CRs from CistromeDB, and used BETA (50) to infer the putative target genes.

Effect of semantic mutations on protein activity

The protein activity of CRs was calculated by using VIPER (51) from gene expression profile data and ARACNe-inferred targets. Considering those CRs with at least 20 nonsilent mutated samples in single cancer type (291 pairs, e.g. ARID1A in gastric cancer), we compared the protein activity between samples with and without mutations to evaluate the functional effect of mutations (Wilcoxon test).

Interaction network

Around 280,000 protein-protein interactions involved in CRs and their first neighbours were extracted from Pathway Commons v8 (52). In addition, we carried out co-occurrence and mutual exclusive mutation analysis (Fisher's exact test) for each edge based on cancer genome sequencing data. The miRNA-CR regulatory relations were integrated from several experiment validation or sequence-based prediction databases such as RAID (53), mir2Disease (54), miRTarBase (55), miRDB (56), miRWalk (57) and TargetSCAN (58). We then calculated Pearson correlations between the expression levels of miRNA and CR, and sorted the 12,000 miRNA-CR pairs by the frequency of cancer types in which they are negatively correlated (*P*-value < 0.05).

USER INTERFACE

CR2Cancer provides a friendly interface for users to browse, search and download data. In the ‘Browse’ page,

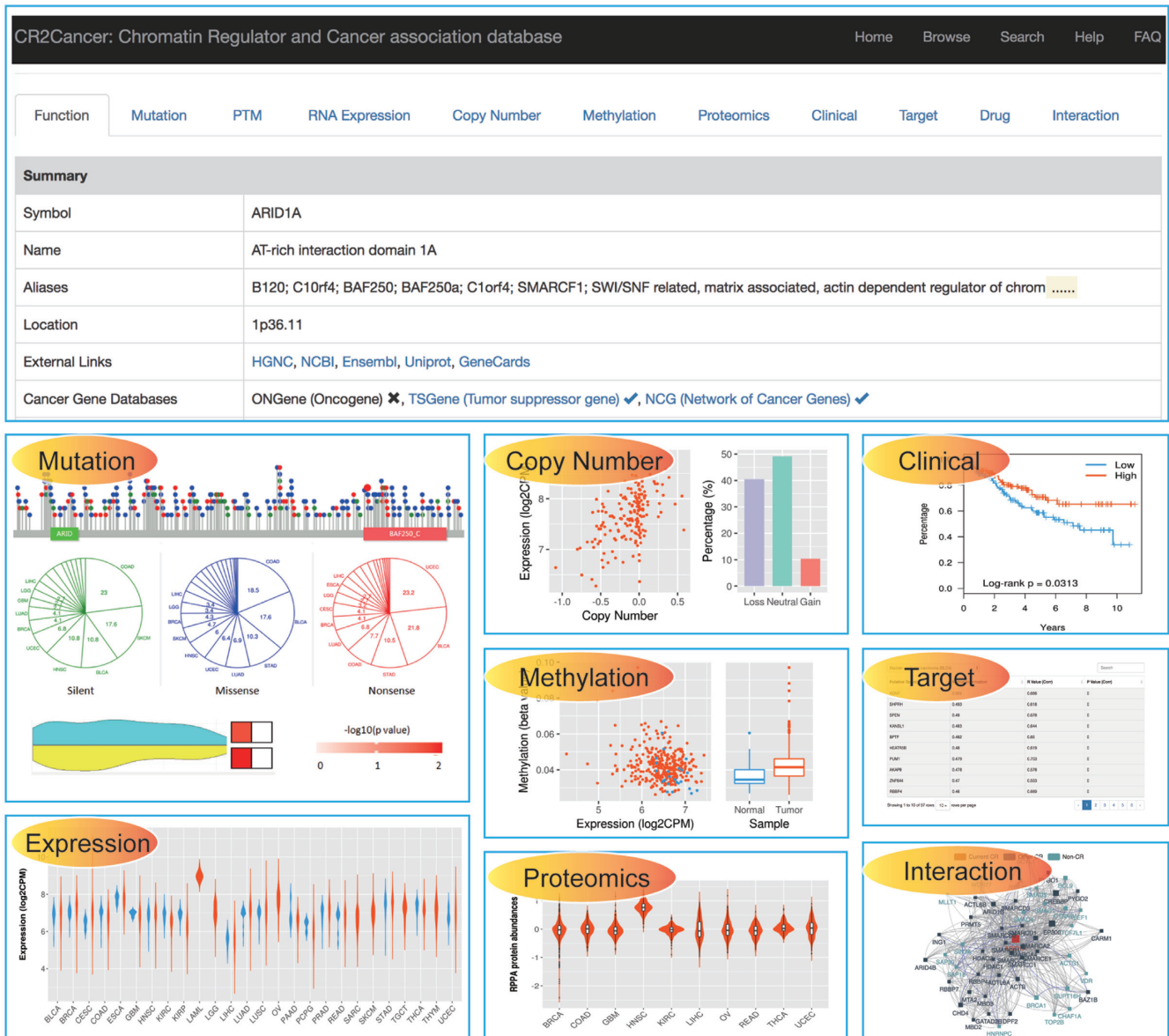


Figure 2. Browse page for ARID1A in CR2Cancer. A user can access different categorized annotations by clicking the tabs at the top of the page.

the user can browse CRs by six different panels: ‘Cancer Type’, ‘Function’, ‘Mutation Rate’, ‘Differential Expression’, ‘ChIP-Seq Data’ and ‘Targeted Drug’. The ‘Cancer Type’ shows 26 and 24 cancer types from primary tumor tissues and cancer cell lines, respectively. By clicking one cancer type, a summary table of all CRs in this cancer can be presented. The ‘Function’ panel offers detailed function annotations of CRs, and users can click the buttons at the top of the panel to filter them by DNA modification, histone modification and chromatin remodelling. In the ‘Mutation Rate’ page, users can obtain the mutation rate of CRs in 8,788 primary tumor samples and in 967 cancer cell lines. We observed that five CRs, KMT2C, ARID1A, CREBBP, TRRAP and TAF1L, are common in the top 10 mutated CR genes of tumor tissues and cell lines. The ‘Differential Expression’ page summaries differential expression analy-

sis of CRs in 15 primary cancer types with more than ten normal samples. We found that 327 (76.2%) CRs are differentially expressed in at least one cancer type, and most of them show consistent (up- or down-) dysregulated patterns across diverse cancers. Finally, users can browse 139 CRs with ChIP-Seq profiles and 39 CRs regarded as drug targets in ‘ChIP-Seq Data’ and ‘Targeted Drug’ panels, respectively. In above pages, to click on the hyperlinks of gene symbols can pinpoint to the entry page of CR, which includes 11 categories of knowledge in different tabs (Figure 2). For each tab, the summary table contains basic information of CR and relevant contents in this tab, which is linked to responding parts for more details. In the ‘Search’ page, CR2Cancer can be queried in an easy-to-use manner with a number of search tools. Users can retrieve CRs of interest in simple and batch format by gene symbol, name, En-

trez ID, UniProt ID, ensembl ID and substrates. Moreover, users can search cancer name of primary tumors, cell lines and text mining results, and obtain a summary page of CRs involved in this cancer. The 'Advanced search' module allows users to query gene and cancer at the same time, and returns the related information of queried CR in the tumor type of interest. In order to conduct advanced data analysis, we have prepared all data in plain text files, which can be accessed from the 'Download' page.

DISCUSSION AND FUTURE DIRECTIONS

Cancer is a large group of more than 200 diseases characterized by uncontrolled cell growth. Although it is typically viewed as a genetic disorder, increasing evidences show that the deregulation of epigenetic mechanisms represents a common feature of tumorigenesis and metastasis (1). Comprehensive analyses of cancer genomes have also revealed that CR genes involved in epigenetic modification are often mutated across diverse cancer types (12,13). These findings firmly establish that epigenetic alterations are not just a result of the cancerous state but play causal roles in disease pathogenesis. Furthermore, it is now clear that cross talk between genetic and epigenetic mechanisms enable the acquisition of cancer hallmarks (59). In addition, the reversible nature of epigenetic modifications makes CRs appealing drug targets for cancer therapy, and a number of small molecules that inhibit the activity of aberrant CRs have been in clinical use or trial (16,17). With the exponential growth of cancer biomedical data, a comprehensive resource for CRs in cancer has become especially imperative for cancer epigenetics research.

In this study, we developed a database CR2Cancer that collects annotation data and abnormal patterns of CRs in human cancer based on high throughput data analysis and literature mining. The information is integrated into 11 categories for each CR: function, mutation, PTM, RNA expression, copy number, methylation, proteomics, clinical, target, drug and interaction. Users can browse all CRs by six different paths or use the advanced search tools to access the CR of interest. Compared with the previous similar databases (21–23), CR2Cancer mainly has two distinctive characteristics. On one hand, in addition to basic knowledge of CRs, our database provides their multi-dimensional molecular scheme (e.g. genome, transcriptome and proteome) and integrative analyses (e.g. association tests between expression and copy number, methylation or clinical features) across various human cancers. On the other hand, CR2Cancer combines both high-throughput and literature supported experiment data, which enable users to cross check the role of CRs during cancer development. We expect that CR2Cancer can serve investigators focusing on cancer epigenetics for not only the elucidation of underpinnings of pathogenesis but also the design of novel targeted therapy.

In the future, we will collect more large-scale cancer genomics data sets from several separated studies (60) and special consortiums (61), which can offer in-depth knowledge of the commonalities and differences among different cancer types or same ones in different platforms. We also will continually curate the literatures on cancer-associated

CRs in a regular basis. Furthermore, genome-wide maps of DNA methylation, histone modifications and chromatin accessibility will be integrated in our database to provide insights of gene regulation through complex processes that involve functional elements, transcription factors and chromatin regulators (62). Finally, we plan to add drug pharmacological data from high throughput screening studies for drug discovery and repurposing.

ACKNOWLEDGEMENTS

The authors are grateful to the editor and the anonymous reviewers for their valuable suggestions and comments, which has led to the improvement of this paper.

FUNDING

Research Grants Council of Hong Kong, China [HKU 17127014 General Research Fund (GRF), HKU T12-710/16R Theme-based Research Scheme (TBRs)]. Funding for open access charge: University of Hong Kong Seed Funding Programme for Basic Research.

Conflict of interest statement. None declared.

REFERENCES

- Dawson, M.A. (2017) The cancer epigenome: concepts, challenges, and therapeutic opportunities. *Science*, **355**, 1147–1152.
- Luger, K., Dechassa, M.L. and Tremethick, D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.*, **13**, 436–447.
- Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. *et al.* (2010) The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.*, **28**, 1045–1048.
- Tessarz, P. and Kouzarides, T. (2014) Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.*, **15**, 703–708.
- Chen, T. and Dent, S.Y. (2014) Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat. Rev. Genet.*, **15**, 93–106.
- Li, E. and Zhang, Y. (2014) DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.*, **6**, a019133.
- Yang, X.J. and Seto, E. (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, **26**, 5310–5318.
- Clapier, C.R. and Cairns, B.R. (2009) The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.*, **78**, 273–304.
- Plass, C., Pfister, S.M., Lindroth, A.M., Bogatyrova, O., Claus, R. and Lichter, P. (2013) Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nat. Rev. Genet.*, **14**, 765–780.
- Yiu, T.T. and Li, W. (2015) Pediatric cancer epigenome and the influence of folate. *Epigenomics*, **7**, 961–973.
- Brien, G.L., Valerio, D.G. and Armstrong, S.A. (2016) Exploiting the epigenome to control cancer-promoting gene-expression programs. *Cancer Cell*, **29**, 464–476.
- Dawson, M.A. and Kouzarides, T. (2012) Cancer epigenetics: from mechanism to therapy. *Cell*, **150**, 12–27.
- Garraway, L.A. and Lander, E.S. (2013) Lessons from the cancer genome. *Cell*, **153**, 17–37.
- McCabe, M.T., Graves, A.P., Ganji, G., Diaz, E., Halsey, W.S., Jiang, Y., Smitheman, K.N., Ott, H.M., Pappalardi, M.B., Allen, K.E. *et al.* (2012) Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 2989–2994.
- Beguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K.L., Rosen, M., Shen, H., Yang, S.N., Wang, L., Ezponda, T. *et al.* (2013) EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell*, **23**, 677–692.

16. Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K. and Schapira, M. (2012) Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discov.*, **11**, 384–400.
17. Pfister, S.X. and Ashworth, A. (2017) Marked for death: targeting epigenetic changes in cancer. *Nat. Rev. Drug Discov.*, **16**, 241–263.
18. Cataldo, V.D., Cortes, J. and Quintas-Cardama, A. (2009) Azacitidine for the treatment of myelodysplastic syndrome. *Expert Rev. Anticancer Ther.*, **9**, 875–884.
19. Saba, H.I. (2007) Decitabine in the treatment of myelodysplastic syndromes. *Ther. Clin. Risk Manag.*, **3**, 807–817.
20. Thurn, K.T., Thomas, S., Moore, A. and Munster, P.N. (2011) Rational therapeutic combinations with histone deacetylase inhibitors for the treatment of cancer. *Future Oncol.*, **7**, 263–283.
21. Khare, S.P., Habib, F., Sharma, R., Gadwal, N., Gupta, S. and Galande, S. (2012) Histome—a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res.*, **40**, D337–D342.
22. Xu, Y., Zhang, S., Lin, S., Guo, Y., Deng, W., Zhang, Y. and Xue, Y. (2017) WERAM: a database of writers, erasers and readers of histone acetylation and methylation in eukaryotes. *Nucleic Acids Res.*, **45**, D264–D270.
23. Wang, Q., Huang, J., Sun, H., Liu, J., Wang, J., Wang, Q., Qin, Q., Mei, S., Zhao, C., Yang, X. *et al.* (2014) CR Cistrome: a ChIP-Seq database for chromatin regulators and histone modification linkages in human and mouse. *Nucleic Acids Res.*, **42**, D450–D458.
24. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
25. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
26. Jiang, P., Freedman, M.L., Liu, J.S. and Liu, X.S. (2015) Inference of transcriptional regulation in cancers. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 7731–7736.
27. Yao, L., Shen, H., Laird, P.W., Farnham, P.J. and Berman, B.P. (2015) Inferring regulatory element landscapes and transcription factor networks from cancer methylomes. *Genome Biol.*, **16**, 105.
28. Lee, E., Ito, K., Zhao, Y., Schadt, E.E., Irie, H.Y. and Zhu, J. (2016) Inferred miRNA activity identifies miRNA-mediated regulatory networks underlying multiple cancers. *Bioinformatics*, **32**, 96–105.
29. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
30. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S. *et al.* (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
31. Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L. *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, **45**, D777–D783.
32. Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A.C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.
33. Shah, M.A., Denton, E.L., Arrowsmith, C.H., Lupien, M. and Schapira, M. (2014) A global assessment of cancer genomic alterations in epigenetic mechanisms. *Epigenet. Chromatin*, **7**, 29.
34. Yang, Z., Jones, A., Widschwendter, M. and Teschendorff, A.E. (2015) An integrative pan-cancer-wide analysis of epigenetic enzymes reveals universal patterns of epigenomic deregulation in cancer. *Genome Biol.*, **16**, 140.
35. Gonzalez-Perez, A., Jene-Sanz, A. and Lopez-Bigas, N. (2013) The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.*, **14**, r106.
36. Liu, Y., Sun, J. and Zhao, M. (2017) ONGene: A literature-based database for human oncogenes. *J. Genet. Genomics*, **44**, 119–121.
37. Zhao, M., Kim, P., Mitra, R., Zhao, J. and Zhao, Z. (2016) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.*, **44**, D1023–D1031.
38. An, O., Dall’Olio, G.M., Mourikis, T.P. and Ciccarelli, F.D. (2016) NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. *Nucleic Acids Res.*, **44**, D992–D999.
39. Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T. and Drablos, F. (2015) EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database*, **2015**, bav067.
40. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S. *et al.* (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.*, **28**, 1248–1250.
41. Kim, J.D., Ohta, T., Tateisi, Y. and Tsujii, J. (2003) GENIA corpus—semantically annotated corpus for bio-textmining. *Bioinformatics*, **19**(Suppl 1), i180–i182.
42. Leaman, R. and Gonzalez, G. (2008) BANNER: an executable survey of advances in biomedical named entity recognition. *Pac. Symp. Biocomput.*, 652–663.
43. Jay, J.J. and Brouwer, C. (2016) Lollipop in the clinic: information dense mutation plots for precision medicine. *PLoS One*, **11**, e0160519.
44. Law, C.W., Chen, Y., Shi, W. and Smyth, G.K. (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.*, **15**, R29.
45. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.*, **43**, e47.
46. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R. and Getz, G. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.*, **12**, R41.
47. Lachmann, A., Giorgi, F.M., Lopez, G. and Califano, A. (2016) ARACNE-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*, **32**, 2233–2235.
48. Chen, L., Wu, G. and Ji, H. (2011) hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics*, **27**, 1447–1448.
49. Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L. *et al.* (2017) Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res.*, **45**, D658–D662.
50. Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y. and Liu, X.S. (2013) Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nat. Protoc.*, **8**, 2502–2515.
51. Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H. and Califano, A. (2016) Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.*, **48**, 838–847.
52. Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D. and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
53. Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., Liu, L., Hou, P., Cui, T., Tan, P. *et al.* (2017) RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res.*, **45**, D115–D118.
54. Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G. and Liu, Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
55. Chou, C.H., Chang, N.W., Shrestha, S., Hsu, S.D., Lin, Y.L., Lee, W.H., Yang, C.D., Hong, H.C., Wei, T.Y., Tu, S.J. *et al.* (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.*, **44**, D239–D247.
56. Wong, N. and Wang, X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.*, **43**, D146–D152.
57. Dweep, H. and Gretz, N. (2015) miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods*, **12**, 697.
58. Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.

59. Shen,H. and Laird,P.W. (2013) Interplay between the cancer genome and epigenome. *Cell*, **153**, 38–55.
60. Gao,J., Aksoy,B.A., Dogrusoz,U., Dresdner,G., Gross,B., Sumer,S.O., Sun,Y., Jacobsen,A., Sinha,R., Larsson,E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, p11.
61. Hudson,T.J., Anderson,W., Artez,A., Barker,A.D., Bell,C., Bernabe,R.R., Bhan,M.K., Calvo,F., Eerola,I., Gerhard,D.S. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
62. Duren,Z., Chen,X., Jiang,R., Wang,Y. and Wong,W.H. (2017) Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E4914–E4923.