



Review

Epigenetics Analysis and Integrated Analysis of Multiomics Data, Including Epigenetic Data, Using Artificial Intelligence in the Era of Precision Medicine

Ryuji Hamamoto ^{1,2,*}, Masaaki Komatsu ^{1,2}, Ken Takasawa ^{1,2} , Ken Asada ^{1,2} and Syuzo Kaneko ¹

¹ Division of Molecular Modification and Cancer Biology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; maskomat@ncc.go.jp (M.K.); ktakazaw@ncc.go.jp (K.T.); ken.asada@riken.jp (K.A.); sykaneko@ncc.go.jp (S.K.)

² Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

* Correspondence: rhamamot@ncc.go.jp; Tel.: +81-3-3547-5271

Received: 1 December 2019; Accepted: 27 December 2019; Published: 30 December 2019



Abstract: To clarify the mechanisms of diseases, such as cancer, studies analyzing genetic mutations have been actively conducted for a long time, and a large number of achievements have already been reported. Indeed, genomic medicine is considered the core discipline of precision medicine, and currently, the clinical application of cutting-edge genomic medicine aimed at improving the prevention, diagnosis and treatment of a wide range of diseases is promoted. However, although the Human Genome Project was completed in 2003 and large-scale genetic analyses have since been accomplished worldwide with the development of next-generation sequencing (NGS), explaining the mechanism of disease onset only using genetic variation has been recognized as difficult. Meanwhile, the importance of epigenetics, which describes inheritance by mechanisms other than the genomic DNA sequence, has recently attracted attention, and, in particular, many studies have reported the involvement of epigenetic deregulation in human cancer. So far, given that genetic and epigenetic studies tend to be accomplished independently, physiological relationships between genetics and epigenetics in diseases remain almost unknown. Since this situation may be a disadvantage to developing precision medicine, the integrated understanding of genetic variation and epigenetic deregulation appears to be now critical. Importantly, the current progress of artificial intelligence (AI) technologies, such as machine learning and deep learning, is remarkable and enables multimodal analyses of big omics data. In this regard, it is important to develop a platform that can conduct multimodal analysis of medical big data using AI as this may accelerate the realization of precision medicine. In this review, we discuss the importance of genome-wide epigenetic and multiomics analyses using AI in the era of precision medicine.

Keywords: epigenetics; precision medicine; DNA methylation; histone modifications; machine learning; deep learning

1. Introduction

Barack Obama, the 44th president of the United States, stated his intention to fund an amount of \$215 million to the “Precision Medicine Initiative” in his 2015 State of the Union Address [1]. Since then, precision medicine has frequently been used as a term that contains concepts of personalized medicine worldwide. Generally, precision medicine refers to a medical model that proposes the customization of healthcare with medical decisions, treatments, practices or products tailored to individual patients.

In this model, diagnostic testing is often employed for selecting appropriate and optimal therapies based on the context of a patient's genetic content or other molecular or cellular analyses [2]. To date, most precision medicine interventions consist of genetic profiling, including the detection of predictive biomarkers [3]. It has been repeatedly reported that this may identify patients at risk for a specific disease or a severe variant of a disease and allow for preventive interventions to reduce the burden of disease and improve quality of life. However, it has also been reported that only a small number of patients benefit from current precision medicine, and it is of no benefit for most tumor patients [4,5]. In addition, it has been stated that the MD Anderson Cancer Center found that the gene sequencing of 2600 people only benefited 6.4% of them through the use of targeted drugs. According to the data about matching plans of the National Cancer Institute, only 2% people can benefit from targeted drugs [4,6]. These results indicate that we definitely need to explore the possibility that more patients can benefit from precision medicine. To extend precision medicine, not only genomic data but also other omics data, such as epigenetic and proteomics data, should be involved, and integrated analyses of different types of omics data are considered to be of paramount importance. In this review article, we highlight the current knowledge of the importance of epigenetic data in precision medicine by describing, in particular, the integrated analysis of multiomics data, including epigenetic data, using machine learning and deep learning technologies.

2. Characteristics of Epigenetics and Technologies for Epigenetics Analysis

2.1. General Characteristics of Epigenetics

In principle, epi-genetics is the study of heritable phenotype changes without altering the DNA sequence [7]. The Greek prefix epi- (ἐπι- “above”) in epi-genetics implies features that are “on top of” or “in addition to” the traditional genetic basis for inheritance [8]. Over the last decade, epigenetic regulators have been implicated as key factors in many pathways relevant to cancer development and progression, including cell cycle regulation, invasiveness, signaling pathways, chemo-resistance and immune evasion [9–38]. The three basic systems of epigenetic regulation are DNA methylation of gene regulatory regions, histone protein modifications, such as methylation, acetylation, phosphorylation and sumoylation and non-coding RNAs [15,20,21]. With regard to the technologies for epigenetics analysis, a number of methods have already been developed, and this field has made steady progress in technological innovation (Figure 1 and Table 1). Below, we highlight technologies for epigenetics analysis including historical context.

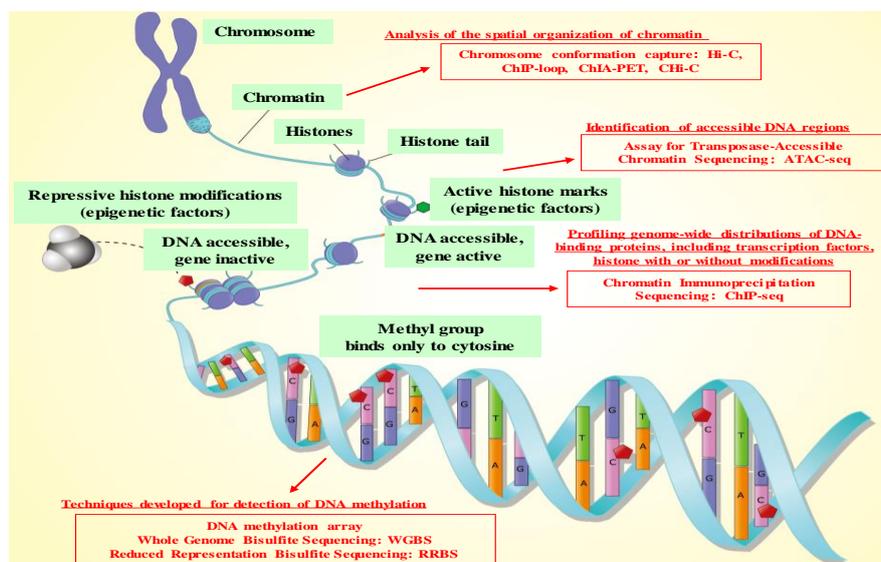


Figure 1. The summarized figure of epigenetic regulations and technologies for epigenetics analysis. Image credit: Shutterstock.com/ellepigrafica.

Table 1. List of main technologies for epigenetics and chromatin analyses.

Method Name	Purpose	Methodology	Era	Ref.
Chromatin immunoprecipitation (ChIP) assay	Analysis of histone modification and transcription factor binding status	A type of immunoprecipitation experimental technique used to investigate the interaction between proteins and DNA in the cell. It aims to determine whether specific proteins are associated with specific genomic regions, and also aims to determine the specific location in the genome that various histone modifications are associated with.	1985	[39,40]
Bisulfite sequencing (BS-Seq)	DNA methylation analysis	Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected. Hence, DNA that has been treated with bisulfite retains only methylated cytosines.	1992	[41,42]
Histone acetyltransferase (HAT) assay	Assay for histone acetyltransferase activity	Multiple biochemical HAT assays have been described; these assays measure HAT activity by detecting either the acetylated histone-based product (direct) or the free CoA product (indirect).	1995	[43,44]
DNA methylation array: differential methylation hybridization (DMH)	DNA methylation analysis	A DNA array-based method, called differential methylation hybridization (DMH), to identify hypermethylated sequences in tumor cells by simultaneously screening many CpG island loci derived from a genomic library, CGL.	1999	[45,46]
ChIP-on-chip	Genome-wide analysis of histone modification and transcription factor binding status	A technology that combines chromatin immunoprecipitation (ChIP) with DNA microarray (chip). It allows the identification of the cistrome, the sum of binding sites, for DNA-binding proteins on a genome-wide basis.	1999	[47,48]
Histone methyltransferase (HMT) assay	Assay for histone methyltransferase activity	Radiometric Assays, Mass Spectrometry, Anti-Methylation Antibody-Based Detection, Enzyme-Coupled SAH Detection, Protease-Coupled Detection, Competition Binding.	2000	[49,50]
Histone demethylase (HDMT) assay	Assay for histone demethylase activity	Measuring the release of radiolabeled formaldehyde from ³ H-labeled methylated histone substrates, by monitoring the change in methylation levels of histone substrates by immunoblotting with site-specific methyl-histone antibodies, or by using mass spectrometry to detect reductions in histone peptide masses that correspond to methyl groups.	2004	[51,52]
Reduced Representation Bisulfite Sequencing (RRBS)	Genome-wide DNA methylation analysis	An efficient and high-throughput technique for analyzing the genome-wide methylation profiles on a single nucleotide level; it combines restriction enzymes and bisulfite sequencing to enrich for areas of the genome with a high CpG content.	2005	[53,54]

Table 1. Cont.

Method Name	Purpose	Methodology	Era	Ref.
ChIP-loop	Chromosome conformation capture technique	This method combines the standard 3C protocol with a routine ChIP protocol; it allows the selective identification of long-range chromatin interactions between loci that are bound to specific proteins of interest.	2005	[55,56]
ChIP-sequencing (ChIP-seq)	Genome-wide analysis of histone modification and transcription factor binding status	By combining chromatin immunoprecipitation (ChIP) assays with next-generation sequencing (NGS), ChIP sequencing (ChIP-seq) is a powerful method for identifying genome-wide DNA binding sites for transcription factors and other proteins.	2007	[57,58]
Whole Genome Bisulfite Sequencing (WGBS)	Genome-wide DNA methylation analysis	A NGS technology used to determine the DNA methylation status of single cytosines by treating the DNA with sodium bisulfite before sequencing.	2009	[59,60]
Hi-C	Chromosome conformation capture technique	A genome-wide chromatin conformation capture protocol using proximity ligation. The technology is of special interest for three-dimensional genome organization in the nucleus and de novo genome assemblies.	2009	[61,62]
ChIA-PET	Determination of de novo long-range chromatin interactions genome-wide	The ChIA-PET method combines ChIP-based methods, and Chromosome conformation capture (3C), to extend the capabilities of both approaches.	2009	[63,64]
ATAC-seq	Identification of accessible DNA regions	This method relies on NGS library construction using the hyperactive transposase Tn5. NGS adapters are loaded onto the transposase, which allows simultaneous fragmentation of chromatin and integration of those adapters into open chromatin regions.	2013	[65–74]
Capture Hi-C (CHi-C)	Identification of higher resolution mapping of chromatin interactions	The CHi-C is a new technique for assessing genome organization based on chromosome conformation capture coupled to oligonucleotide capture of regions of interest like gene promoters.	2014	[75]

2.2. Technologies for Epigenetics Analysis before the NGS Era

In the 1980s, the basic principle of chromatin immunoprecipitation (ChIP) was established; for instance, Gilmour and Lis demonstrated that proteins were cross-linked to DNA in intact cells, and the protein-DNA adducts were isolated by immunoprecipitation with antiserum against the protein [39]. On the basis of this principle, several kinds of applied technologies were reported so far, indicating that this methodology greatly contributes to the progress of epigenetics. In the 1990s, as understandings of the physiological and biological importance of the DNA methylation were deepened, assay methods to analyze DNA methylation status were actively developed. Importantly, treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine residues unaffected. Hence, DNA that has been treated with bisulfite retains only methylated cytosines. On the basis of this molecular mechanism, Frommer et al. reported a genomic sequencing method that provides positive identification

of 5-methylcytosine residues and yields strand-specific sequences of individual molecules in genomic DNA [41], which formed the basis of subsequent development of DNA methylation assays. At the end of the 20th century, DNA microarray-based methods were also used for epigenetics analysis. In 1999, a DNA-array-based method, called differential methylation hybridization (DMH), was developed to identify hypermethylated sequences in tumor cells by simultaneously screening many CpG islands (CGIs) [45]; this technology could explore further the underlying mechanisms of DNA methylation. Likewise, a first ChIP-on-chip experiment, a technology that combines chromatin immunoprecipitation (ChIP) with DNA microarray (chip), was performed in 1999 to analyze the distribution of cohesin along budding yeast chromosome III [47]. Using tiled arrays, ChIP-on-chip allows for high resolution of genome-wide maps, which can determine the binding sites of many DNA-binding proteins like transcription factors and also chromatin modifications.

As for biochemical analysis of epigenetic regulators, such as histone acetyltransferases, histone methyltransferases and histone demethylases, various procedures were developed in the late 20th and early 21st centuries [43,49,51,76]. A series of biochemical analyses particularly unveiled the biological significance of epigenetics so far.

2.3. Technologies for Epigenetics Analysis in the NGS Era and Genome-Wide Epigenetics Analysis

In 2005, new sequencing techniques began to emerge that permitted an unbiased means to examine billions of templates of DNA and RNA. Although now almost fifteen years old, the term “next-generation sequencing (NGS)” remains the popular way to describe very-high-throughput sequencing methods that allow millions to trillions of observations to be made in parallel during a single instrument run [77]. Importantly, progress of NGS technologies produced several methods of genome-wide epigenetics analysis. Reduced representation bisulfite sequencing (RRBS) is an efficient and high-throughput technique to analyze the genome-wide methylation profiles; it combines restriction enzymes and bisulfite sequencing to enrich for areas of the genome with a high CpG content. Given the high cost and depth of sequencing to analyze methylation status in the whole genome, the RRBS technique was developed in 2005 to reduce the amount of nucleotides required for sequence to 1% of the genome [53]. Moreover, in 2009, the first human genome-wide single-base-resolution DNA methylation map was established by Whole Genome Bisulfite Sequencing (WGBS) [59], which showed the utilization of this technique to investigate the relationship between DNA methylation loci and human phenotypes in both basic and clinical research [78,79].

In terms of ChIP analysis, a new method called ChIP-sequencing (ChIP-seq), which combines chromatin immunoprecipitation with massively parallel DNA sequencing (NGS), was developed in 2007 [57]. This technique enabled genome-wide analysis of histone modifications and transcription factor binding, which could contribute to the investigation of the relationship between histone modification status or transcription factor binding status and human phenotypes in both basic and clinical research [25,26,80]. Subsequently, in 2013, a new technology called ATAC-seq (assay for transposase-accessible chromatin using sequencing) was developed [65]; ATAC-seq could identify accessible (open) chromatin regions with hyperactive mutant Tn5 Transposase that inserts sequencing adaptors into open regions of the genome [81]. This method has been applied to defining the genome-wide chromatin accessibility landscape in human cancers [82], and computational footprinting methods can be performed on ATAC-seq to identify cell specific binding sites of transcription factors and their cell specific activity [83].

Furthermore, the concept of chromatin contact mapping, or determining the three-dimensional structure conformation and interactions of chromatin domains, recently attracts a lot of attention because chromosome conformation capture methods (3C-based methods) have advanced rapidly. For example, ChIP-loop is a technique that 3C-based methods and ChIP-seq are combined, which detects interactions between two loci of interest mediated by a protein of interest [55]. In addition, the Hi-C technique, a comprehensive technique to capture the conformation of genomes, is the first of the 3C derivative technologies to be truly genome-wide [61]. Subsequently, another new technology called

ChIP-PET, which combines Hi-C with ChIP-seq, was developed to detect all interactions mediated by a protein of interest [55,63]. More recently, a new technique called Capture Hi-C (CHi-C) was developed (Table 1). The CHi-C method allows the simultaneous and higher resolution mapping of chromatin interactions for large subsets of the genome, such as all promoters or DNase hypersensitive sites.

In the NGS era, although genome-wide epigenetics analyses are enabled, the amount of data we need to analyze is rapidly increasing. Besides, given that multimodal analysis to integrate epigenetic data and other omics data like genomics data has recently been considered important, we recognize the importance of artificial intelligence (AI) utilization to analyze the epigenetic data efficiently and effectively.

3. Development of Artificial Intelligence (AI)

3.1. Machine Learning Techniques and Evolution of AI Technologies

Machine learning is a sub-set of AI technologies where computer algorithms are used to autonomously learn from data and information (Figure 2). Historically, the learning behaviors of neurons have been researched for a long time to reveal the mechanism of human cognition. One of the most famous theories is the Hebbian Learning Rule proposed by Donald Olding Hebb [84]. On the basis of the Hebbian Learning Rule in the study of artificial neural networks, we can obtain powerful models of neural computation that might be close to the function of structures found in neural systems of many diverse species [85,86]. In 1958, Frank Rosenblatt developed the perceptron, which became the first model that could learn the weights defining the categories given examples of inputs from each category [87]. In the 1980s, Kunihiko Fukushima proposed the neocognitron, which is a hierarchical, multilayered artificial neural network [88]. This neural network has been used for handwritten character recognition and other pattern recognition tasks; importantly, it served as the inspiration for convolutional neural networks [89]. In 1986, David Rumelhart, Geoff Hinton and Ronald J. Williams demonstrated the process of backpropagation, which is a method used in artificial neural networks to calculate the error contribution of each neuron after a batch of data (in image recognition, multiple images) is processed [90]. This method is a special case of an older and more general technique called automatic differentiation. With regard to the learning, it is generally used by the gradient descent optimization algorithm to tune the weight of neurons by calculating the gradient of the loss of function. Then, in 1992, Christopher Watkins developed Q-learning [91], exceedingly improving the practicality and feasibility of reinforcement learning, which is a paradigm that aims to model the trial-and-error learning process that is needed in many problem situations where explicit instructive signals are not available [92]. Additionally, Corinna Cortes and Vladimir Vapnik developed the support vector machine (SVM) machine learning algorithm, which is a model with associated learning algorithms that analyzes data used for classification and regression analysis [93–95]. The classifier that the SVM initializes is useful for predicting between two possible outcomes that depend on continuous or categorical predictor variables [96]. In 1995, Tin Kam Ho described the random forest algorithm, which is an ensemble learning method for classification, regression and other tasks, operated by constructing a large number of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [97]. This method can correct for decision trees' habit of overfitting to their training set [98].

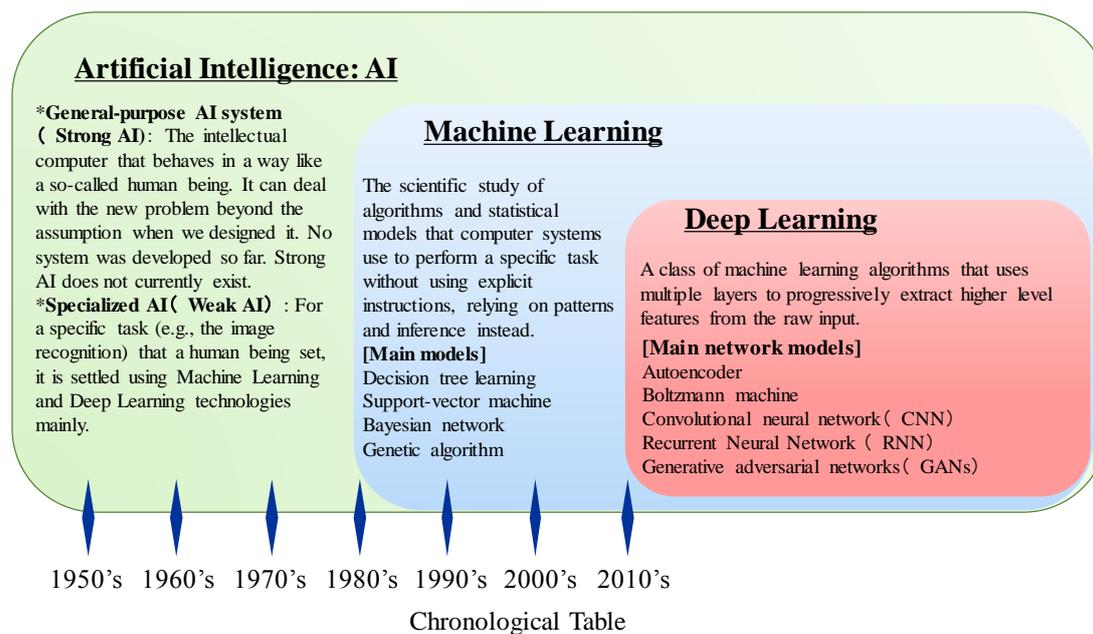


Figure 2. The summarized figure of artificial intelligence development.

3.2. AI Revolution Using Deep Learning in the Big Data Era

In the 21st century, we have access to large amounts of data, known as “Big Data”, and faster computer and advanced machine learning techniques were successfully applied to many problems throughout society, which accelerates social implementation of AI technologies. Indeed, by 2016, the market for AI-related products reached more than 8 billion dollars, and the New York Times reported that interest in AI had reached a “frenzy” [99]. In particular, advances in deep learning, a branch of machine learning that models high level abstractions in data by using a deep graph with many processing layers, drove progress and research in image and video processing, text analysis and even speech recognition (Figure 2) [100]. Artificial neural networks (ANNs) were inspired by information processing and distributed communication nodes in biological systems; meanwhile, ANNs have various differences from biological brains. Specifically, neural networks tend to be static and symbolic, while the biological brain of most living organisms is dynamic (plastic) and analog [101].

Importantly, the current progress of deep learning technologies has been truly astonishing. In 2012, AlexNet, which is the name of a convolutional neural network (CNN) designed by Alex Krizhevsky, won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC); this network achieved a top-5 error of 15.3%, more than 10.8 percentage points lower than that of the runner up [102]. AlexNet achieved state-of-the-art recognition accuracy against all the traditional machine learning and computer vision approaches, which was a significant breakthrough in the field of machine learning and computer vision for visual recognition and classification tasks and the point in history where interests in deep learning rapidly increased [103]. With regard to the accuracy for ILSVRC, the error rate of the deep learning model designed by the winners’ group in each year has significantly been improved year by year. Particularly, ResNet-152, the 152-layer Residual Neural Network (ResNet), developed by Microsoft group achieved 3.57% error rate in 2015, which won the 1st place in the ILSVRC2015 and outperformed human accuracy (5% error rate) [103,104]. In addition to the superhuman performance of AlphaGo, an AI-powered system based on the deep reinforcement learning (DRL) technology that beat the world No.1 ranked Go player [105], now AI technologies using deep learning attract a lot of attention in the various kinds of fields, including the medical field [106–108].

4. Epigenetics Analysis and Integrated Analysis of Multiomics Data, Including Epigenetic Data, Using AI Technologies in the Medical Field

4.1. Advantages of Machine Learning and Deep Learning Technologies for Analysis of Medical Big Data

In order to realize precision medicine, integrated analysis of medical big data is essential; we summarized the advantages of machine learning and deep learning technologies for analysis of medical big data (Figure 3). So far, it has been difficult to have all such characteristics by the conventional analytical techniques, but a number of machine learning and deep learning technologies possess all four features, which shows advantages of these technologies in medical research.

4.1.1. Multimodal Learning

Data in the real world usually comes as different modalities. For instance, images are associated with captions and tags, videos contain visual and audio signals, sensory perception includes simultaneous inputs from visual, auditory, motor and haptic pathways [109]. Different modalities are characterized by very different statistical properties. For example, images are usually represented as pixel intensities or outputs of feature extractors, while texts are represented as discrete word count vectors. Given the distinct statistical properties of different information resources, to discover the relationship between different modalities is very important. Multimodal Learning is a good model to represent the joint representations of different modalities, such as genomic mutation data, epigenetic data and transcriptome data in medical research (Figure 3A). In fact, it was reported that predications of cancer prognosis or anti-cancer drug sensitivities were enabled based on multimodal learning using various different types of medical data [110–112]. Since molecular mechanisms of diseases like cancer are pretty complicated and a variety of factors are intricately involved, characteristics of multimodal learning must be critical for elucidation of the mechanism of diseases.

Multimodal Learning

* We can integrate different types of omics data (genomic data, epigenetic data, medical imaging data, etc.) as input.

Multitask Learning

* We can increase learning efficiency by learning several different tasks at the same time together with sharing some components of the learning process.

Representation Learning and Semi-supervised Learning

* It is known that unlabeled data, when used in conjunction with limited amount of labeled data, can produce considerable improvement in learning accuracy.

Automatic acquisition of hierarchical features

* We can analyze high-dimensional correlations of input data.



Although it was difficult to have all such characteristics by the conventional analytical techniques, a number of ML and DL technologies possess all 4 aforementioned features, which shows advantages of these technologies in medical research

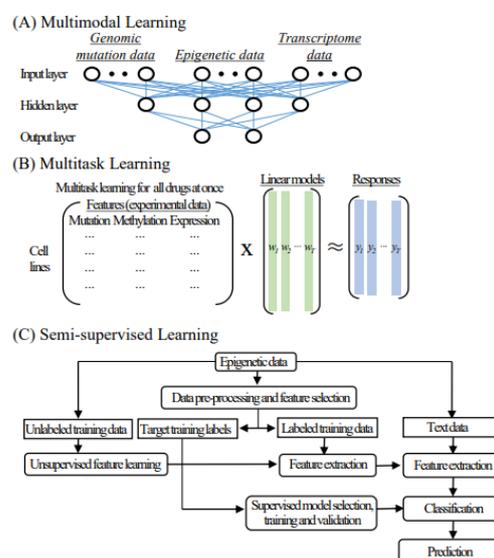


Figure 3. Advantages of machine learning and deep learning technologies in medical research. (A) An example of multimodal learning analysis using multiomics data including epigenetic data. (B) An example of multitask learning analysis using gene mutation data, DNA methylation data and gene expression data. This is a modified figure from reference [113]. (C) An example of semi-supervised learning using epigenetic data. This is a modified figure from reference [114].

4.1.2. Multitask Learning

Multitask Learning is a subfield of machine learning in which multiple learning tasks are solved at the same time, while exploiting commonalities and differences across tasks [115]. Using this approach, we can improve learning efficiency and prediction accuracy for the task-specific models, when compared

to training the models separately [116,117]. Although multitask learning algorithms have a long history in machine learning, their common theme is that by sharing information between tasks, and often by encoding that the learned models for different tasks should have some similarity to each other [113,118]. It is possible to improve over independent training of individual tasks, in particular when training data for each task may be limited [113]. Intriguingly, several multitask learning approaches have recently been proposed to predict drug sensitivity; two kernel-based methods demonstrated improved performance over elastic net regression [113,119–123]. In this regard, a kernel-based multitask approach was the winner of a DREAM competition to predict drug sensitivity in a small breast cancer cell line data set [123], and another work encoded features of drugs in a neural network based multitask strategy [119]. For example, a schematic figure of multitask models is shown in Figure 3B (modified figure from reference [113]). In this case, trace norm regularization with a highly efficient ADMM (alternating direction method of multipliers) optimization algorithm that readily scales to large data sets was used. In the precision medicine era, because to predict drug sensitivity for each patient is a fundamental task, the concept of multitask learning to analyze omics data including epigenetic data is useful.

4.1.3. Representation Learning and Semi-Supervised Learning

In machine learning, representation learning is a set of techniques that allows a system to automatically discover the representations needed for feature detection or classification from raw data [124]. This replaces manual feature engineering and allows a machine to both learn the features and use them to perform a specific task. Semi-supervised learning is a class of machine learning tasks and techniques that also make use of unlabeled data for training—typically a small amount of labeled data with a large amount of unlabeled data [125]. On the basis of these characteristics, it is known that unlabeled data, when used in conjunction with limited amount of labeled data, can produce considerable improvement in learning accuracy [126,127]. A flowchart of the training and testing processes of a semi-supervised deep learning method for cancer prediction is shown in Figure 3C (modified figure from reference [114]). The semi-supervised classification model consists of the unsupervised feature extraction stage and the supervised classification stage, which is possible to address both unlabeled and labeled data to extract more valuable information and make better predictions [114]. As the number of labeled data is often limited in the medical data, particularly for analysis of rare diseases, this characteristic is useful in such case.

4.1.4. Automatic Acquisition of Hierarchical Characteristics

Deep learning is a type of machine learning technique that aims at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. Automatically learning features at multiple levels of abstraction enable construction of a system to learn complex functions mapping the input to the output directly from data, without relying completely on human-crafted features [128].

4.2. Analysis of Epigenetic Data and Integrated Analysis of Epigenetic Data and Other Omics Data Using AI Technologies

Although the screening of genetic mutations is considered common practice for testing an individual's predisposition to cancer, it cannot reflect the current status or activity of disease [129,130]. In contrast, promoter DNA methylation is a more systematic method for evaluation due to its defined location within the promoter regions of specific genes. In general, locating gene mutations is more complex as they can occur at unsuspected sites within the gene that may be challenging to pinpoint. In this regard, several epigenetic markers have value in the early detection of cancers based on their involvement in the initiation of carcinogenic pathways [129,131,132]. Hence, epigenetic biomarkers are likely to have great potential and wide scope to be implemented as diagnostic biomarkers.

Consequently, we can expect that the strategy of combining epigenetic biomarkers and AI technologies (machine learning and deep learning technologies) might be useful for the diagnosis of diseases.

Brain tumors are clinically and biologically highly diverse, which encompasses a wide spectrum from benign tumors, which can frequently be cured by surgery alone (e.g., pilocytic astrocytoma), to highly malignant tumors that respond poorly to any therapy (e.g., glioblastoma) [133,134]. So far, a number of studies reported substantial inter-observer variability in the histopathological diagnosis of a lot of central nervous system (CNS) tumors, for instance, in diffuse gliomas, ependymomas and supratentorial primitive neuroectodermal tumors [133,135–137]. Since DNA methylation profiling is robust and reproducible even from small samples and poor quality material [138], DNA methylation profiles have been widely used to subclassify CNS tumors that were previously considered homogenous diseases [133,137,139–144]. On the basis of this previous work within single entities, Capper et al. recently presented a comprehensive machine learning approach for the DNA methylation-based classification of central nervous system tumors across all entities and age groups, and demonstrated its application in a routine diagnostic setting [133]. This study showed that the availability of the DNA methylation-based classification method using machine learning might have a substantial impact on diagnostic precision compared to standard methods, which results in a change of diagnosis in up to 12% of prospective cases [133]. In essence, this study provides new strategy for the generation of epigenetics-based tumor classifiers using AI across other cancer entities, with the potential to fundamentally transform tumor pathology.

Integrated analysis of epigenetic data with other omics data using AI technologies has also been advanced [110–112]. For example, Chaudhary et al. presented a deep learning-based model on hepatocellular carcinoma (HCC) robustly differentiated survival subpopulations of patients in six cohorts; they built the deep learning-based survival-sensitive model on 360 HCC patients' data using epigenetics (DNA methylation) data with RNA sequencing (RNA-seq) data and microRNA-sequencing (miRNA-seq) data from The Cancer Genome Atlas (TCGA), which predicts prognosis as good as an alternative model where genomics and clinical data are both considered [111]. In this case, the autoencoder method, which is an unsupervised deep learning technique, was used in the model, and it could capture sufficient variations due to potential clinical risk factors, such that it performs as accurately or even better than, having additional clinical features in the model [111]. Importantly, the autoencoder framework showed much more efficiency to identify features linked to survival compared with the principal component analysis (PCA) or individual Cox proportional-hazards-based models [111].

A fundamental integrated analysis of epigenetic data with other omics data using AI technologies is that we can clarify the significance of genetic mutations in the noncoding regions of the human genome. Although genome-wide association studies (GWAS) have already identified a large number of inherited risk loci for cancer susceptibility, many of these single-nucleotide polymorphisms (SNPs) reside in a noncoding genome within known DNA regulatory elements [82]. However, the majority of annotation tools only annotate SNPs in the coding region of a genome [145,146]. This is in part because noncoding SNPs are more challenging to annotate than SNPs in coding regions where the consequences of variation are better understood [145]. In order to predict functional SNPs in a noncoding genome, Corces et al. recently presented the genome-wide chromatin accessibility of 410 tumor samples spanning 23 cancer types from TCGA; they identified 562,709 transposase-accessible DNA elements that substantially extend the compendium of known cis-regulatory elements [82]. The integrated analysis of ATAC-seq with TCGA data identified numerous putative distal enhancers that can distinguish molecular subtypes of tumors, uncovered specific driving transcription factors through protein-DNA footprints, and nominated long-range gene-regulatory interactions in tumors [82]. The findings by group of Corces and others reveal genetic risk loci of cancer predisposition as active DNA regulatory elements in cancer can identify gene-regulatory interactions underlying cancer immune evasion and pinpoint noncoding mutations that drive enhance activation and may affect patient survival. These results suggest a systematic approach to understanding the noncoding genome in cancer to advance diagnosis

and therapy. In their study, K-means clustering was used, being one of the simplest and most popular unsupervised machine learning algorithms [82]. Meanwhile, given that whole genome sequencing (WGS) analysis using a large number of cancer tissues is being actively conducted worldwide, the development of AI-based platforms that can perform integrated analyses of large-scale multiomics data must be critical to finding useful information for the diagnosis and therapy of cancer.

As mentioned above, the Hi-C technique emerged as a powerful tool for studying the spatial organization of chromosomes, as it measures all pair-wise interaction frequencies across the entire genome [127]. During recent years, this method facilitated a number of significant discoveries like A/B compartment, topological associating domains (TADs), chromatin loops and frequently interacting regions (FIREs), which significantly expanded understandings of three-dimensional (3D) gene organization and gene regulation machinery [61,147–151]. However, the Hi-C technology usually requires an extremely deep-sequencing depth to achieve high resolution; this fact causes the remarkable rise of experimental costs, which makes it hard for researchers to apply it to a large number of cell lines [149,152,153]. In this regard, several computational methods have been reported to improve the resolution of Hi-C data and detect physiological interactions at the regulatory element level [152,154–157]. For example, Zhu et al. reported EpiTensor, which is a high-order tensor decomposition based algorithm to identify 3D spatial associations within TADs from 1D maps of histone modifications, chromatin accessibility and RNA-seq [155]; Whalen et al. presented a machine learning pipeline called TargetFinder, which integrates data for annotation, Cap Analysis of Gene Expression (CAGE), ChIP-seq, DNase I hypersensitive sites sequencing (DNase-seq), FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) and DNA methylation to predict individual promoter-enhancer interactions across the genome [157]. Additionally, Bkhetan et al. introduced a supervised learning pipeline called 3DEpiLoop, which uses random forest as a statistical learning algorithm, and this algorithm can predict three-dimensional chromatin looping interactions within TADs from one-dimensional epigenetics and transcription factor profiles using the statistical learning [156]. Zhang et al. also developed HiCPlus, which is a computational approach based on the deep convolutional neural network, to infer high-resolution Hi-C interaction matrices from low-resolution Hi-C data [154]. More recently, Li et al. developed a bootstrapping deep learning model called DeepTACT (deep neural networks for chromatin conTACTs predictions), which can predict chromatin contacts at individual regulatory element level using sequence features and chromatin accessibility information [152]. This model employed a bootstrapping strategy, which is based on the theory established in the paper reported by Wallace et al. in 2011 [158]. In essence, DeepTACT can predict not only promoter-enhancer interactions, but also promoter-promoter interactions, and DeepTACT fine-maps chromatin contacts of high-quality promoter capture Hi-C (PCHi-C) from the multiple regulatory element level (5–20 kb) to the individual regulatory element level [130]. In addition, DeepTACT can identify a set of hub promoters, which are active across cell lines, enriched in housekeeping genes, closely related to fundamental biological processes and capable of reflecting cell similarity [152]. The other important advantage of this model is that we can infer novel associations for coronary artery disease through integrative analysis of chromatin contacts predicted by DeepTACT and existing GWAS, which provides a powerful way to build a fine-scale chromatin connectivity map to explore the mechanisms of human diseases [152]. As noted above, because most of the non-coding variants are not well annotated linked to genes that they regulate, it is still difficult to evaluate the significance of these mutations. Hence, precise identification of interactions between promoters and their regulation is urgently needed; aforementioned integrated analysis of DeepTACT-based chromatin contacts and GWAS-based gene mutation data appears to be pretty important.

4.3. Issues of AI Technologies for Omics Analysis

Thus far we introduced a number of merits to use AI technologies for integrated analysis of omics data; but there are also several defects we need to overcome in them. One of the serious issues of AI technologies is a phenomenon called overfitting. In general, overfitting means that the production of an

analysis that corresponds too closely or exactly to a particular set of data, which sometimes causes the failure of fitting additional data or predicting future observations reliably [159]. Overfitting in neural networks shows poor performance on the test set compared to the training data set, signifying a loss of generalization. More specifically, the model learns the noise patterns present in the training data set, thereby causing a large gap between the training and test error [160]. Principally, deep neural networks are prone to overfitting because of the large number of parameters to be learned [160]; additionally, these networks are so flexible and overparameterized that they adjust the parameters in order to fit the training data even with labels randomized [160–162].

Meanwhile, in order to avoid overfitting, several methods have also been proposed, and we highlight some important methods below:

Cross-validation: Cross-validation is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. It is used to test the effectiveness of a machine learning model and is also a resampling procedure used to evaluate a model if we have a limited data. It was reported that cross-validation could reduce the risk of selecting models that suffer from overfitting to the observed data [163].

Regularization: An appropriate level of complexity is needed to avoid overfitting, and regularization is a method that controls a model's complexity by penalizing the magnitude of its parameters [164]. The common regularization methods to reduce overfitting are L1 regulation (a regression model that uses L1 regulation technique is called Lasso Regression), L2 regulation (a regression model that uses L2 regulation technique is called ridge regression), dropout regulation (reducing overfitting in neural networks by preventing complex co-adaptations on training data) and early stopping (stopping the model when model reaches a plateau) [165].

Train with more data: Even though it is not always available, training with more data can help algorithms detect the signal better. In the earlier example of modeling height vs. age in children, it is clear how sampling more schools can help our model. However, an important point we should be careful about is that this is not always the case because this method cannot help our model if we just add noisy data. Therefore, that is why we should always ensure our data are clean and relevant.

The other critical issue of using AI technologies for omics analysis is that omics data including genomic data and epigenetic data possess a large number of parameters (for example, the number of human genes are around 30,000), which are often much higher than that of sample number. Especially, in the case of rare diseases, the number of patients is critically small; but current aforementioned WGS technology enables researchers to interrogate all three billion base pairs of the human genome. This kind of problem is generally typified as the "curse of dimensionality"; the number of features characterizing the data are "too large" and "the curse of dataset sparsity"; the number of samples on which these features are measured is "too small" [166]. The curse of dataset sparsity refers to the scenario where the number of parameters like genomic factors is far larger than the number of samples, which results in model overfitting and computational inefficiency [167]. In order to overcome this "curse of dimensionality" issue for omics analysis, new techniques have also been developed. Recently, regularized logistic regression using the L1 regularization has successfully applied in high-dimensional cancer classification to tackle both the estimation of gene coefficients and the simultaneous performance of gene selection; but the L1 regularization has a biased gene selection and does not have the order property. Hence, Wu et al. investigated the L1/2 regularized logistic regression for gene selection in cancer classification, and experimental results on three DNA microarray database demonstrated the proposed method using sparse logistic regression with L1/2 regularization outperformed other commonly used sparse methods (L1 regulation and elastic net penalty) in terms of classification performance [168]. Furthermore, Romero et al. developed a novel deep learning-based technique called diet networks, which could considerably reduce the number of free parameters [169]. This model is based on the idea that we can first learn or provide a distributed representation for each input feature (e.g., for each position in the genome where variations are observed in data), and then learn (with another neural network called the parameter prediction network) how to map a feature's distributed

representation (on the basis of the feature's identity) to the vector of parameters specific to that feature in the classifier neural network (the weights which link the value of the feature to each of the hidden units) [169], which could deal with the issues of producing the parameters associated with each feature as a multitask learning model [169]. Given that the diet networks algorithm enables significant reduction of both the number of parameters and the error rate of the classifier, it must be useful to apply this method for analysis of various type of omics data, including epigenetic data.

5. Concluding Remarks and Future Perspectives

In this review, we discussed the current conditions and possibilities of omics analyses using AI for the realization of precision medicine. In particular, we focused on epigenetics analysis. As mentioned, omics analyses using AI technology have many possibilities; however, there are a number of issues we need to overcome. In this regard, we thought that AI-based techniques need to be improved for their successful application to realization of precision medicine, based on the efforts to solve current issues one by one. One important strategy seems to be that experts of biomedical science, and experts of information science and bioinformatics should collaborate deeply. In this way, both groups of experts can solve problems together based on highly merged knowledge because, definitely, this is an interdisciplinary research field. In addition, while the progress of AI algorithms is important, we thought that it is also critical to construct a database for the accumulation of a large quantity of omics data, where high-quality appropriate annotation is added with right clinical information. After all, even if a huge number of omics data is available, poor quality of big data would create misleading results. As it is thought that the practical use of AI is indispensable to the realization of precision medicine, we believe that continuing efforts to solve the issues we have mentioned herein surely and steadily will contribute to the realization of true precision medicine.

Author Contributions: R.H.: conceptualization, drafting of the manuscript writing and funding acquisition, M.K.: writing—original draft; K.T.: writing—original draft; K.A.: writing—original draft; S.K.: writing—original draft; all the co-authors have read and approved the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JST CREST [Grant Number JPMJCR1689] and JSPS Grant-in-Aid for Scientific Research on Innovative Areas [Grant Number JP18H04908].

Acknowledgments: We are grateful to Daisuke Okanohara, Jun Sese and Kazuma Kobayashi for helpful discussion. The authors show a great gratitude to the past and present members of Hamamoto Laboratory.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Hoosain, N.; Pearce, B.; Jacobs, C.; Benjeddou, M. Mapping SLCO1B1 Genetic Variation for Global Precision Medicine in Understudied Regions in Africa: A Focus on Zulu and Cape Admixed Populations. *OMICS* **2016**, *20*, 546–554. [[CrossRef](#)] [[PubMed](#)]
2. Goyal, M.R. *Scientific and Technical Terms in Bioengineering and Biological Engineering*; Apple Academic Press: Cambridge, MA, USA, 2018.
3. Kasztura, M.; Richard, A.; Bempong, N.E.; Loncar, D.; Flahault, A. Cost-effectiveness of precision medicine: A scoping review. *Int. J. Public Health* **2019**. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, X.; Yang, H.; Zhang, R. Challenges and future of precision medicine strategies for breast cancer based on a database on drug reactions. *Biosci. Rep.* **2019**, *39*. [[CrossRef](#)] [[PubMed](#)]
5. Prasad, V. Perspective: The precision-oncology illusion. *Nature* **2016**, *537*, S63. [[CrossRef](#)] [[PubMed](#)]
6. Meric-Bernstam, F.; Brusco, L.; Shaw, K.; Horombe, C.; Kopetz, S.; Davies, M.A.; Routbort, M.; Piha-Paul, S.A.; Janku, F.; Ueno, N.; et al. Feasibility of Large-Scale Genomic Testing to Facilitate Enrollment Onto Genomically Matched Clinical Trials. *J. Clin. Oncol.* **2015**, *33*, 2753–2762. [[CrossRef](#)]
7. Dupont, C.; Armant, D.R.; Brenner, C.A. Epigenetics: Definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* **2009**, *27*, 351–357. [[CrossRef](#)]

8. Rozek, L.S.; Dolinoy, D.C.; Sartor, M.A.; Omenn, G.S. Epigenetics: Relevance and implications for public health. *Annu. Rev. Public Health* **2014**, *35*, 105–122. [[CrossRef](#)]
9. Baylin, S.B. Resistance, epigenetics and the cancer ecosystem. *Nat. Med.* **2011**, *17*, 288–289. [[CrossRef](#)]
10. Mohammad, H.P.; Baylin, S.B. Linking cell signaling and the epigenetic machinery. *Nat. Biotechnol.* **2010**, *28*, 1033–1038. [[CrossRef](#)]
11. Ezponda, T.; Popovic, R.; Shah, M.Y.; Martinez-Garcia, E.; Zheng, Y.; Min, D.J.; Will, C.; Neri, A.; Kelleher, N.L.; Yu, J.; et al. The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene* **2013**, *32*, 2882–2890. [[CrossRef](#)]
12. Cho, H.S.; Hayami, S.; Toyokawa, G.; Maejima, K.; Yamane, Y.; Suzuki, T.; Dohmae, N.; Kogure, M.; Kang, D.; Neal, D.E.; et al. RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia* **2012**, *14*, 476–486. [[CrossRef](#)] [[PubMed](#)]
13. Cho, H.S.; Suzuki, T.; Dohmae, N.; Hayami, S.; Unoki, M.; Yoshimatsu, M.; Toyokawa, G.; Takawa, M.; Chen, T.; Kurash, J.K.; et al. Demethylation of RB regulator MYPT1 by histone demethylase LSD1 promotes cell cycle progression in cancer cells. *Cancer Res.* **2011**, *71*, 1–6. [[CrossRef](#)] [[PubMed](#)]
14. Hayami, S.; Yoshimatsu, M.; Veerakumarasivam, A.; Unoki, M.; Iwai, Y.; Tsunoda, T.; Field, H.I.; Kelly, J.D.; Neal, D.E.; Yamaue, H.; et al. Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: Involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol. Cancer* **2010**, *9*, 59. [[CrossRef](#)]
15. Saloura, V.; Cho, H.S.; Kyiotani, K.; Alachkar, H.; Zuo, Z.; Nakakido, M.; Tsunoda, T.; Seiwert, T.; Lingen, M.; Licht, J.; et al. WHSC1 Promotes Oncogenesis through Regulation of NIMA-related-kinase-7 in Squamous Cell Carcinoma of the Head and Neck. *Mol. Cancer Res.* **2015**, *13*, 293–304. [[CrossRef](#)]
16. Tomasi, T.B.; Magner, W.J.; Khan, A.N. Epigenetic regulation of immune escape genes in cancer. *Cancer Immunol. Immunother.* **2006**, *55*, 1159–1184. [[CrossRef](#)] [[PubMed](#)]
17. Cho, H.S.; Kelly, J.D.; Hayami, S.; Toyokawa, G.; Takawa, M.; Yoshimatsu, M.; Tsunoda, T.; Field, H.I.; Neal, D.E.; Ponder, B.A.; et al. Enhanced expression of EHMT2 is involved in the proliferation of cancer cells through negative regulation of SIAH1. *Neoplasia* **2011**, *13*, 676–684. [[CrossRef](#)] [[PubMed](#)]
18. Cho, H.S.; Toyokawa, G.; Daigo, Y.; Hayami, S.; Masuda, K.; Ikawa, N.; Yamane, Y.; Maejima, K.; Tsunoda, T.; Field, H.I.; et al. The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. *Int. J. Cancer* **2012**, *131*, E179–E189. [[CrossRef](#)]
19. Hamamoto, R.; Furukawa, Y.; Morita, M.; Iimura, Y.; Silva, F.P.; Li, M.; Yagyū, R.; Nakamura, Y. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat. Cell Biol.* **2004**, *6*, 731–740. [[CrossRef](#)]
20. Hamamoto, R.; Nakamura, Y. Dysregulation of protein methyltransferases in human cancer: An emerging target class for anticancer therapy. *Cancer Sci.* **2016**. [[CrossRef](#)]
21. Hamamoto, R.; Saloura, V.; Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer* **2015**, *15*, 110–124. [[CrossRef](#)]
22. Hamamoto, R.; Silva, F.P.; Tsuge, M.; Nishidate, T.; Katagiri, T.; Nakamura, Y.; Furukawa, Y. Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci.* **2006**, *97*, 113–118. [[CrossRef](#)] [[PubMed](#)]
23. Hayami, S.; Kelly, J.D.; Cho, H.S.; Yoshimatsu, M.; Unoki, M.; Tsunoda, T.; Field, H.I.; Neal, D.E.; Yamaue, H.; Ponder, B.A.; et al. Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *Int. J. Cancer* **2011**, *128*, 574–586. [[CrossRef](#)] [[PubMed](#)]
24. Kang, D.; Cho, H.S.; Toyokawa, G.; Kogure, M.; Yamane, Y.; Iwai, Y.; Hayami, S.; Tsunoda, T.; Field, H.I.; Matsuda, K.; et al. The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis. *Genes Chromosom. Cancer* **2013**, *52*, 126–139. [[CrossRef](#)] [[PubMed](#)]
25. Kogure, M.; Takawa, M.; Cho, H.S.; Toyokawa, G.; Hayashi, K.; Tsunoda, T.; Kobayashi, T.; Daigo, Y.; Sugiyama, M.; Atomi, Y.; et al. Deregulation of the histone demethylase JMJD2A is involved in human carcinogenesis through regulation of the G(1)/S transition. *Cancer Lett.* **2013**, *336*, 76–84. [[CrossRef](#)] [[PubMed](#)]
26. Kogure, M.; Takawa, M.; Saloura, V.; Sone, K.; Piao, L.; Ueda, K.; Ibrahim, R.; Tsunoda, T.; Sugiyama, M.; Atomi, Y.; et al. The oncogenic polycomb histone methyltransferase EZH2 methylates lysine 120 on histone H2B and competes ubiquitination. *Neoplasia* **2013**, *15*, 1251–1261. [[CrossRef](#)] [[PubMed](#)]

27. Piao, L.; Suzuki, T.; Dohmae, N.; Nakamura, Y.; Hamamoto, R. SUV39H2 methylates and stabilizes LSD1 by inhibiting polyubiquitination in human cancer cells. *Oncotarget* **2015**, *6*, 16939–16950. [[CrossRef](#)]
28. Silva, F.P.; Hamamoto, R.; Kunizaki, M.; Tsuge, M.; Nakamura, Y.; Furukawa, Y. Enhanced methyltransferase activity of SMYD3 by the cleavage of its N-terminal region in human cancer cells. *Oncogene* **2008**, *27*, 2686–2692. [[CrossRef](#)]
29. Takawa, M.; Masuda, K.; Kunizaki, M.; Daigo, Y.; Takagi, K.; Iwai, Y.; Cho, H.S.; Toyokawa, G.; Yamane, Y.; Maejima, K.; et al. Validation of the histone methyltransferase EZH2 as a therapeutic target for various types of human cancer and as a prognostic marker. *Cancer Sci.* **2011**, *102*, 1298–1305. [[CrossRef](#)]
30. Toyokawa, G.; Cho, H.S.; Iwai, Y.; Yoshimatsu, M.; Takawa, M.; Hayami, S.; Maejima, K.; Shimizu, N.; Tanaka, H.; Tsunoda, T.; et al. The histone demethylase JMJD2B plays an essential role in human carcinogenesis through positive regulation of cyclin-dependent kinase 6. *Cancer Prev. Res.* **2011**, *4*, 2051–2061. [[CrossRef](#)]
31. Toyokawa, G.; Cho, H.S.; Masuda, K.; Yamane, Y.; Yoshimatsu, M.; Hayami, S.; Takawa, M.; Iwai, Y.; Daigo, Y.; Tsuchiya, E.; et al. Histone Lysine Methyltransferase Wolf-Hirschhorn Syndrome Candidate 1 Is Involved in Human Carcinogenesis through Regulation of the Wnt Pathway. *Neoplasia* **2011**, *13*, 887–898. [[CrossRef](#)]
32. Tsuge, M.; Hamamoto, R.; Silva, F.P.; Ohnishi, Y.; Chayama, K.; Kamatani, N.; Furukawa, Y.; Nakamura, Y. A variable number of tandem repeats polymorphism in an E2F-1 binding element in the 5' flanking region of SMYD3 is a risk factor for human cancers. *Nat. Genet.* **2005**, *37*, 1104–1107. [[CrossRef](#)] [[PubMed](#)]
33. Yoshimatsu, M.; Toyokawa, G.; Hayami, S.; Unoki, M.; Tsunoda, T.; Field, H.I.; Kelly, J.D.; Neal, D.E.; Maehara, Y.; Ponder, B.A.; et al. Dysregulation of PRMT1 and PRMT6, Type I arginine methyltransferases, is involved in various types of human cancers. *Int. J. Cancer* **2011**, *128*, 562–573. [[CrossRef](#)] [[PubMed](#)]
34. Kojima, M.; Sone, K.; Oda, K.; Hamamoto, R.; Kaneko, S.; Oki, S.; Kukita, A.; Machino, H.; Honjoh, H.; Kawata, Y.; et al. The histone methyltransferase WHSC1 is regulated by EZH2 and is important for ovarian clear cell carcinoma cell proliferation. *BMC Cancer* **2019**, *19*, 455. [[CrossRef](#)] [[PubMed](#)]
35. Kukita, A.; Sone, K.; Oda, K.; Hamamoto, R.; Kaneko, S.; Komatsu, M.; Wada, M.; Honjoh, H.; Kawata, Y.; Kojima, M.; et al. Histone methyltransferase SMYD2 selective inhibitor LLY-507 in combination with poly ADP ribose polymerase inhibitor has therapeutic potential against high-grade serous ovarian carcinomas. *Biochem. Biophys. Res. Commun.* **2019**, *513*, 340–346. [[CrossRef](#)]
36. Kim, S.K.; Kim, K.; Ryu, J.W.; Ryu, T.Y.; Lim, J.H.; Oh, J.H.; Min, J.K.; Jung, C.R.; Hamamoto, R.; Son, M.Y.; et al. The novel prognostic marker, EHMT2, is involved in cell proliferation via HSPD1 regulation in breast cancer. *Int. J. Oncol.* **2019**, *54*, 65–76. [[CrossRef](#)]
37. Shigekawa, Y.; Hayami, S.; Ueno, M.; Miyamoto, A.; Suzaki, N.; Kawai, M.; Hirono, S.; Okada, K.I.; Hamamoto, R.; Yamaue, H. Overexpression of KDM5B/JARID1B is associated with poor prognosis in hepatocellular carcinoma. *Oncotarget* **2018**, *9*, 34320–34335. [[CrossRef](#)]
38. Ryu, J.W.; Kim, S.K.; Son, M.Y.; Jeon, S.J.; Oh, J.H.; Lim, J.H.; Cho, S.; Jung, C.R.; Hamamoto, R.; Kim, D.S.; et al. Novel prognostic marker PRMT1 regulates cell growth via downregulation of CDKN1A in HCC. *Oncotarget* **2017**, *8*, 115444–115455. [[CrossRef](#)] [[PubMed](#)]
39. Gilmour, D.S.; Lis, J.T. In vivo interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol. Cell Biol.* **1985**, *5*, 2009–2018. [[CrossRef](#)]
40. Collas, P. The current state of chromatin immunoprecipitation. *Mol. Biotechnol.* **2010**, *45*, 87–100. [[CrossRef](#)]
41. Frommer, M.; McDonald, L.E.; Millar, D.S.; Collis, C.M.; Watt, F.; Grigg, G.W.; Molloy, P.L.; Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831. [[CrossRef](#)]
42. Fcraga, M.F.; Esteller, M. DNA methylation: A profile of methods and applications. *Biotechniques* **2002**, *33*, 632–649. [[CrossRef](#)] [[PubMed](#)]
43. Brownell, J.E.; Allis, C.D. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6364–6368. [[CrossRef](#)]
44. Ogryzko, V.V.; Schiltz, R.L.; Russanova, V.; Howard, B.H.; Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **1996**, *87*, 953–959. [[CrossRef](#)]
45. Huang, T.H.; Perry, M.R.; Laux, D.E. Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* **1999**, *8*, 459–470. [[CrossRef](#)]
46. Zuo, T.; Tycko, B.; Liu, T.M.; Lin, J.J.; Huang, T.H. Methods in DNA methylation profiling. *Epigenomics* **2009**, *1*, 331–345. [[CrossRef](#)] [[PubMed](#)]

47. Blat, Y.; Kleckner, N. Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell* **1999**, *98*, 249–259. [[CrossRef](#)]
48. Lieb, J.D.; Liu, X.; Botstein, D.; Brown, P.O. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.* **2001**, *28*, 327–334. [[CrossRef](#)]
49. Rea, S.; Eisenhaber, F.; O’Carroll, D.; Strahl, B.D.; Sun, Z.W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C.P.; Allis, C.D.; et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000**, *406*, 593–599. [[CrossRef](#)]
50. Sone, K.; Piao, L.; Nakakido, M.; Ueda, K.; Jenuwein, T.; Nakamura, Y.; Hamamoto, R. Critical role of lysine 134 methylation on histone H2AX for gamma-H2AX production and DNA repair. *Nat. Commun.* **2014**, *5*, 5691. [[CrossRef](#)]
51. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstone, J.R.; Cole, P.A.; Casero, R.A. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **2004**, *119*, 941–953. [[CrossRef](#)]
52. Yamane, K.; Toumazou, C.; Tsukada, Y.; Erdjument-Bromage, H.; Tempst, P.; Wong, J.; Zhang, Y. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **2006**, *125*, 483–495. [[CrossRef](#)] [[PubMed](#)]
53. Meissner, A.; Gnirke, A.; Bell, G.W.; Ramsahoye, B.; Lander, E.S.; Jaenisch, R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **2005**, *33*, 5868–5877. [[CrossRef](#)] [[PubMed](#)]
54. Chatterjee, A.; Rodger, E.J.; Stockwell, P.A.; Weeks, R.J.; Morison, I.M. Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. *J. Biomed. Biotechnol.* **2012**, *2012*, 741542. [[CrossRef](#)] [[PubMed](#)]
55. Hakim, O.; Misteli, T. SnapShot: Chromosome conformation capture. *Cell* **2012**, *148*, 1068-e1. [[CrossRef](#)]
56. Gavrilov, A.; Eivazova, E.; Priozykova, I.; Lipinski, M.; Razin, S.; Vassetzky, Y. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol. Biol.* **2009**, *567*, 171–188.
57. Johnson, D.S.; Mortazavi, A.; Myers, R.M.; Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **2007**, *316*, 1497–1502. [[CrossRef](#)]
58. Schmid, C.D.; Bucher, P. ChIP-Seq data reveal nucleosome architecture of human promoters. *Cell* **2007**, *131*, 831–832. [[CrossRef](#)]
59. Lister, R.; Pelizzola, M.; Dowen, R.H.; Hawkins, R.D.; Hon, G.; Tonti-Filippini, J.; Nery, J.R.; Lee, L.; Ye, Z.; Ngo, Q.M.; et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **2009**, *462*, 315–322. [[CrossRef](#)]
60. Stevens, M.; Cheng, J.B.; Li, D.; Xie, M.; Hong, C.; Maire, C.L.; Ligon, K.L.; Hirst, M.; Marra, M.A.; Costello, J.F.; et al. Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods. *Genome Res.* **2013**, *23*, 1541–1553. [[CrossRef](#)]
61. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragozy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)]
62. Belton, J.M.; McCord, R.P.; Gibcus, J.H.; Naumova, N.; Zhan, Y.; Dekker, J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **2012**, *58*, 268–276. [[CrossRef](#)]
63. Fullwood, M.J.; Liu, M.H.; Pan, Y.F.; Liu, J.; Xu, H.; Mohamed, Y.B.; Orlov, Y.L.; Velkov, S.; Ho, A.; Mei, P.H.; et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **2009**, *462*, 58–64. [[CrossRef](#)] [[PubMed](#)]
64. Li, G.; Fullwood, M.J.; Xu, H.; Mulawadi, F.H.; Velkov, S.; Vega, V.; Ariyaratne, P.N.; Mohamed, Y.B.; Ooi, H.S.; Tennakoon, C.; et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **2010**, *11*, R22. [[CrossRef](#)] [[PubMed](#)]
65. Buenostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **2013**, *10*, 1213–1218. [[CrossRef](#)]
66. Kumasaka, N.; Knights, A.J.; Gaffney, D.J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **2016**, *48*, 206–213. [[CrossRef](#)]
67. Corces, M.R.; Trevino, A.E.; Hamilton, E.G.; Greenside, P.G.; Sinnott-Armstrong, N.A.; Vesuna, S.; Satpathy, A.T.; Rubin, A.J.; Montine, K.S.; Wu, B.; et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **2017**, *14*, 959–962. [[CrossRef](#)] [[PubMed](#)]

68. Ackermann, A.M.; Wang, Z.; Schug, J.; Naji, A.; Kaestner, K.H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **2016**, *5*, 233–244. [[CrossRef](#)] [[PubMed](#)]
69. Lu, Z.; Hofmeister, B.T.; Vollmers, C.; DuBois, R.M.; Schmitz, R.J. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res.* **2017**, *45*, e41. [[CrossRef](#)] [[PubMed](#)]
70. Scharer, C.D.; Blalock, E.L.; Barwick, B.G.; Haines, R.R.; Wei, C.; Sanz, I.; Boss, J.M. ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naive SLE B cells. *Sci. Rep.* **2016**, *6*, 27030. [[CrossRef](#)] [[PubMed](#)]
71. Pott, S.; Lieb, J.D. Single-cell ATAC-seq: Strength in numbers. *Genome Biol.* **2015**, *16*, 172. [[CrossRef](#)]
72. Satpathy, A.T.; Saligrama, N.; Buenrostro, J.D.; Wei, Y.; Wu, B.; Rubin, A.J.; Granja, J.M.; Lareau, C.A.; Li, R.; Qi, Y.; et al. Transcript-indexed ATAC-seq for precision immune profiling. *Nat. Med.* **2018**, *24*, 580–590. [[CrossRef](#)] [[PubMed](#)]
73. Wang, J.; Zibetti, C.; Shang, P.; Sripathi, S.R.; Zhang, P.; Cano, M.; Hoang, T.; Xia, S.; Ji, H.; Merbs, S.L.; et al. ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat. Commun.* **2018**, *9*, 1364. [[CrossRef](#)] [[PubMed](#)]
74. Jia, G.; Preussner, J.; Chen, X.; Guenther, S.; Yuan, X.; Yekelchik, M.; Kuenne, C.; Looso, M.; Zhou, Y.; Teichmann, S.; et al. Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat. Commun.* **2018**, *9*, 4877. [[CrossRef](#)] [[PubMed](#)]
75. Dryden, N.H.; Broome, L.R.; Dudbridge, F.; Johnson, N.; Orr, N.; Schoenfelder, S.; Nagano, T.; Andrews, S.; Wingett, S.; Kozarewa, I.; et al. Unbiased Analysis of Potential Targets of Breast Cancer Susceptibility Loci by Capture Hi-C. *Genome Res.* **2014**, *24*, 1854–1868.77. [[CrossRef](#)]
76. Bannister, A.J.; Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **1996**, *384*, 641–643. [[CrossRef](#)]
77. Levy, S.E.; Myers, R.M. Advancements in Next-Generation Sequencing. *Annu. Rev. Genom. Hum. Genet.* **2016**, *17*, 95–115. [[CrossRef](#)]
78. Zhou, L.; Ng, H.K.; Drautz-Moses, D.I.; Schuster, S.C.; Beck, S.; Kim, C.; Chambers, J.C.; Loh, M. Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. *Sci. Rep.* **2019**, *9*, 10383. [[CrossRef](#)]
79. Zhou, W.; Dinh, H.Q.; Ramjan, Z.; Weisenberger, D.J.; Nicolet, C.M.; Shen, H.; Laird, P.W.; Berman, B.P. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **2018**, *50*, 591–602. [[CrossRef](#)]
80. Yan, H.; Tian, S.; Slager, S.L.; Sun, Z. ChIP-seq in studying epigenetic mechanisms of disease and promoting precision medicine: Progresses and future directions. *Epigenomics* **2016**, *8*, 1239–1258. [[CrossRef](#)]
81. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **2015**, *109*, 21–29. [[CrossRef](#)]
82. Corces, M.R.; Granja, J.M.; Shams, S.; Louie, B.H.; Seoane, J.A.; Zhou, W.; Silva, T.C.; Groeneveld, C.; Wong, C.K.; Cho, S.W.; et al. The chromatin accessibility landscape of primary human cancers. *Science* **2018**, *362*, eaav1898. [[CrossRef](#)] [[PubMed](#)]
83. Li, Z.; Schulz, M.H.; Look, T.; Begemann, M.; Zenke, M.; Costa, I.G. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **2019**, *20*, 45. [[CrossRef](#)] [[PubMed](#)]
84. Hebb, D.O. *The Organization of Behavior: A Neuropsychological Theory*; Wiley & Sons: New York, NY, USA, 1949.
85. Liu, J.; Gong, M.; Miao, Q. Modeling Hebb Learning Rule for Unsupervised Learning. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), Melbourne, Australia, 19–25 August 2017; pp. 2315–2321.
86. Kurisak, E.; Marsalek, P.; Stroffek, J.; Toth, P. Biological context of Hebb learning in artificial neural networks, a review. *Neurocomputing* **2015**, *152*, 27–35. [[CrossRef](#)]
87. Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychol. Rev.* **1958**, *65*, 386–408. [[CrossRef](#)]
88. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202. [[CrossRef](#)]
89. Nebauer, C. Evaluation of convolutional neural networks for visual recognition. *IEEE Trans. Neural Netw.* **1998**, *9*, 685–696. [[CrossRef](#)]

90. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
91. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [[CrossRef](#)]
92. Watkins, C.J.C.H. Learning from Delayed Rewards. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 1989.
93. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92), Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
94. Vapnik, V.; Lerner, A. Pattern Recognition Using Generalized Portrait Method. *Autom. Remote Control* **1963**, *24*, 774–780.
95. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
96. Ben-Hur, A.; Horn, D.; Siegelmann, H.; Vapnik, V. Support vector clustering. *J. Mach. Learn. Res.* **2001**, *2*, 125–137. [[CrossRef](#)]
97. Ho, T.K. Random Decision Forests. In Proceedings of the IEEE Third International Conference on Document Analysis and Recognition, Montreal, QC, USA, 14–16 August 1995; Volume 1, pp. 278–282.
98. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: Berlin, Germany, 2008; pp. 587–588.
99. Lohr, S. *IBM Is Counting on Its Bet on Watson, and Paying Big Money for It*; The New York Times: New York, NY, USA, 2016.
100. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
101. Marblestone, A.H.; Wayne, G.; Kording, K.P. Toward an Integration of Deep Learning and Neuroscience. *Front. Comput. Neurosci.* **2016**, *10*, 94. [[CrossRef](#)]
102. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *1*, 1097–1105. [[CrossRef](#)]
103. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.; Asari, V.K. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* **2019**, *8*, 292. [[CrossRef](#)]
104. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
105. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359. [[CrossRef](#)] [[PubMed](#)]
106. Yamada, M.; Saito, Y.; Imaoka, H.; Saiko, M.; Yamada, S.; Kondo, H.; Takamaru, H.; Sakamoto, T.; Sese, J.; Kuchiba, A.; et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Sci. Rep.* **2019**, *9*, 14465. [[CrossRef](#)]
107. Yasutomi, S.; Arakaki, T.; Hamamoto, R. Shadow Detection for Ultrasound Images Using Unlabeled Data and Synthetic Shadows. *arXiv* **2019**, arXiv:1908.01439.
108. Yasutomi, S.; Sakai, A.; Komatsu, M.; Matsuoka, R.; Komatsu, R.; Arakaki, T.; Tokunaka, M.; Kobayashi, K.; Asada, K.; Kaneko, S.; et al. Unsupervised Shadow Detection for Ultrasound Images by Deep Learning. *IEICE Tech. Rep.* **2019**, *118*, 151–156.
109. Srivastava, N.; Salakhutdinov, R. Multimodal Learning with Deep Boltzmann Machines. *J. Mach. Learn. Res.* **2014**, *15*, 2949–2980.
110. Zhu, B.; Song, N.; Shen, R.; Arora, A.; Machiela, M.J.; Song, L.; Landi, M.T.; Ghosh, D.; Chatterjee, N.; Baladandayuthapani, V.; et al. Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Sci. Rep.* **2017**, *7*, 16954. [[CrossRef](#)] [[PubMed](#)]
111. Chaudhary, K.; Poirion, O.B.; Lu, L.; Garmire, L.X. Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin. Cancer Res.* **2018**, *24*, 1248–1259. [[CrossRef](#)] [[PubMed](#)]
112. Lee, S.I.; Celik, S.; Logsdon, B.A.; Lundberg, S.M.; Martins, T.J.; Oehler, V.G.; Estey, E.H.; Miller, C.P.; Chien, S.; Dai, J.; et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat. Commun.* **2018**, *9*, 42. [[CrossRef](#)] [[PubMed](#)]
113. Yuan, H.; Paskov, I.; Paskov, H.; Gonzalez, A.J.; Leslie, C.S. Multitask learning improves prediction of cancer drug sensitivity. *Sci. Rep.* **2016**, *6*, 31619. [[CrossRef](#)]

114. Xiao, Y.; Wu, J.; Lin, Z.; Zhao, X. A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data. *Comput. Methods Programs Biomed.* **2018**, *166*, 99–105. [[CrossRef](#)]
115. Strezoski, G.; Van Noord, N.; Worring, M. Learning Task Relatedness in Multi-Task Learning for Images in Context. *arXiv* **2019**, arXiv:1904.03011.
116. Baxter, J. A model of inductive bias learning. *J. Artif. Intell. Res.* **2000**, *12*, 149–198. [[CrossRef](#)]
117. Caruana, R. Multitask Learning. *Mach. Learn.* **1997**, *28*, 41–75. [[CrossRef](#)]
118. Zhang, Y.; Yang, Q. A Survey on Multi-Task Learning. *arXiv* **2018**, arXiv:1707.08114v2.119.
119. Costello, J.C.; Heiser, L.M.; Georgii, E.; Gonen, M.; Menden, M.P.; Wang, N.J.; Bansal, M.; Ammad-ud-din, M.; Hintsanen, P.; Khan, S.A.; et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **2014**, *32*, 1202–1212. [[CrossRef](#)]
120. Gonen, M.; Margolin, A.A. Drug susceptibility prediction against a panel of drugs using kernelized Bayesian multitask learning. *Bioinformatics* **2014**, *30*, i556–i563. [[CrossRef](#)] [[PubMed](#)]
121. Heider, D.; Senge, R.; Cheng, W.; Hullermeier, E. Multilabel classification for exploiting cross-resistance information in HIV-1 drug resistance prediction. *Bioinformatics* **2013**, *29*, 1946–1952. [[CrossRef](#)] [[PubMed](#)]
122. Wei, G.; Margolin, A.A.; Haery, L.; Brown, E.; Cucolo, L.; Julian, B.; Shehata, S.; Kung, A.L.; Beroukhim, R.; Golub, T.R. Chemical genomics identifies small-molecule MCL1 repressors and BCL-xL as a predictor of MCL1 dependency. *Cancer Cell* **2012**, *21*, 547–562. [[CrossRef](#)] [[PubMed](#)]
123. Zhang, N.; Wang, H.; Fang, Y.; Wang, J.; Zheng, X.; Liu, X.S. Predicting Anticancer Drug Responses Using a Dual-Layer Integrated Cell Line-Drug Network Model. *PLoS Comput. Biol.* **2015**, *11*, e1004498. [[CrossRef](#)]
124. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intel.* **2013**, *35*, 1798–1828. [[CrossRef](#)]
125. Oliver, A.; Odena, A.; Raffel, C.; Cubuk, E.D.; Goodfellow, I.J. Realistic Evaluation of Deep Semi-Supervised Learning Algorithms. *arXiv* **2018**, arXiv:1804.09170.
126. Shi, M.; Zhang, B. Semi-supervised learning improves gene expression-based prediction of cancer recurrence. *Bioinformatics* **2011**, *27*, 3017–3023. [[CrossRef](#)]
127. Chapelle, O.; Sindhvani, V.; Keerthi, S.S. Optimization Techniques for Semi-Supervised Support Vector Machines. *J. Mach. Learn. Res.* **2008**, *9*, 203–233.
128. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends@Mach. Learn.* **2009**, *2*, 1–127. [[CrossRef](#)]
129. Leygo, C.; Williams, M.; Jin, H.C.; Chan, M.W.Y.; Chu, W.K.; Grusch, M.; Cheng, Y.Y. DNA Methylation as a Noninvasive Epigenetic Biomarker for the Detection of Cancer. *Dis. Mark.* **2017**, *2017*, 3726595. [[CrossRef](#)]
130. Elliott, G.O.; Johnson, I.T.; Scarll, J.; Dainty, J.; Williams, E.A.; Garg, D.; Coupe, A.; Bradburn, D.M.; Mathers, J.C.; Belshaw, N.J. Quantitative profiling of CpG island methylation in human stool for colorectal cancer detection. *Int. J. Colorectal Dis.* **2013**, *28*, 35–42. [[CrossRef](#)]
131. Linton, A.; Cheng, Y.Y.; Griggs, K.; Schedlich, L.; Kirschner, M.B.; Gattani, S.; Srikanan, S.; Chuan-Hao Kao, S.; McCaughan, B.C.; Klebe, S.; et al. An RNAi-based screen reveals PLK1, CDK1 and NDC80 as potential therapeutic targets in malignant pleural mesothelioma. *Br. J. Cancer* **2014**, *110*, 510–519. [[CrossRef](#)]
132. Yang, X.; Dai, W.; Kwong, D.L.; Szeto, C.Y.; Wong, E.H.; Ng, W.T.; Lee, A.W.; Ngan, R.K.; Yau, C.C.; Tung, S.Y.; et al. Epigenetic markers for noninvasive early detection of nasopharyngeal carcinoma by methylation-sensitive high resolution melting. *Int. J. Cancer* **2015**, *136*, E127–E135. [[CrossRef](#)] [[PubMed](#)]
133. Capper, D.; Jones, D.T.W.; Sill, M.; Hovestadt, V.; Schrimpf, D.; Sturm, D.; Koelsche, C.; Sahm, F.; Chavez, L.; Reuss, D.E.; et al. DNA methylation-based classification of central nervous system tumours. *Nature* **2018**, *555*, 469–474. [[CrossRef](#)] [[PubMed](#)]
134. Merve, A.; Millner, T.O.; Marino, S. Integrated phenotype-genotype approach in diagnosis and classification of common central nervous system tumours. *Histopathology* **2019**, *75*, 299–311. [[CrossRef](#)]
135. Van den Bent, M.J. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: A clinician’s perspective. *Acta Neuropathol.* **2010**, *120*, 297–304. [[CrossRef](#)]
136. Ellison, D.W.; Kocak, M.; Figarella-Branger, D.; Felice, G.; Catherine, G.; Pietsch, T.; Frappaz, D.; Massimino, M.; Grill, J.; Boyett, J.M.; et al. Histopathological grading of pediatric ependymoma: Reproducibility and clinical relevance in European trial cohorts. *J. Negat. Results Biomed.* **2011**, *10*, 7. [[CrossRef](#)]
137. Sturm, D.; Orr, B.A.; Toprak, U.H.; Hovestadt, V.; Jones, D.T.W.; Capper, D.; Sill, M.; Buchhalter, I.; Northcott, P.A.; Leis, I.; et al. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. *Cell* **2016**, *164*, 1060–1072. [[CrossRef](#)]

138. Hovestadt, V.; Remke, M.; Kool, M.; Pietsch, T.; Northcott, P.A.; Fischer, R.; Cavalli, F.M.; Ramaswamy, V.; Zapatka, M.; Reifemberger, G.; et al. Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **2013**, *125*, 913–916. [[CrossRef](#)]
139. Reuss, D.E.; Kratz, A.; Sahm, F.; Capper, D.; Schrimpf, D.; Koelsche, C.; Hovestadt, V.; Bewerunge-Hudler, M.; Jones, D.T.; Schittenhelm, J.; et al. Adult IDH wild type astrocytomas biologically and clinically resolve into other tumor entities. *Acta Neuropathol.* **2015**, *130*, 407–417. [[CrossRef](#)]
140. Pajtler, K.W.; Witt, H.; Sill, M.; Jones, D.T.; Hovestadt, V.; Kratochwil, F.; Wani, K.; Tatevossian, R.; Punchihewa, C.; Johann, P.; et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell* **2015**, *27*, 728–743. [[CrossRef](#)] [[PubMed](#)]
141. Lambert, S.R.; Witt, H.; Hovestadt, V.; Zucknick, M.; Kool, M.; Pearson, D.M.; Korshunov, A.; Ryzhova, M.; Ichimura, K.; Jabado, N.; et al. Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma. *Acta Neuropathol.* **2013**, *126*, 291–301. [[CrossRef](#)] [[PubMed](#)]
142. Mack, S.C.; Witt, H.; Piro, R.M.; Gu, L.; Zuyderduyn, S.; Stutz, A.M.; Wang, X.; Gallo, M.; Garzia, L.; Zayne, K.; et al. Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **2014**, *506*, 445–450. [[CrossRef](#)] [[PubMed](#)]
143. Johann, P.D.; Erkek, S.; Zapatka, M.; Kerl, K.; Buchhalter, I.; Hovestadt, V.; Jones, D.T.W.; Sturm, D.; Hermann, C.; Segura Wang, M.; et al. Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **2016**, *29*, 379–393. [[CrossRef](#)] [[PubMed](#)]
144. Wiestler, B.; Capper, D.; Sill, M.; Jones, D.T.; Hovestadt, V.; Sturm, D.; Koelsche, C.; Bertoni, A.; Schweizer, L.; Korshunov, A.; et al. Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma. *Acta Neuropathol.* **2014**, *128*, 561–571. [[CrossRef](#)] [[PubMed](#)]
145. Nishizaki, S.S.; Boyle, A.P. Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms. *Trends Genet.* **2017**, *33*, 34–45. [[CrossRef](#)]
146. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9362–9367. [[CrossRef](#)]
147. Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J.S.; Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380. [[CrossRef](#)]
148. Nora, E.P.; Lajoie, B.R.; Schulz, E.G.; Giorgetti, L.; Okamoto, I.; Servant, N.; Piolot, T.; van Berkum, N.L.; Meisig, J.; Sedat, J.; et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **2012**, *485*, 381–385. [[CrossRef](#)]
149. Rao, S.S.; Huntley, M.H.; Durand, N.C.; Stamenova, E.K.; Bochkov, I.D.; Robinson, J.T.; Sanborn, A.L.; Machol, I.; Omer, A.D.; Lander, E.S.; et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **2014**, *159*, 1665–1680. [[CrossRef](#)]
150. Schmitt, A.D.; Hu, M.; Jung, I.; Xu, Z.; Qiu, Y.; Tan, C.L.; Li, Y.; Lin, S.; Lin, Y.; Barr, C.L.; et al. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep.* **2016**, *17*, 2042–2059. [[CrossRef](#)] [[PubMed](#)]
151. Schmitt, A.D.; Hu, M.; Ren, B. Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 743–755. [[CrossRef](#)] [[PubMed](#)]
152. Li, W.; Wong, W.H.; Jiang, R. DeepTACT: Predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic Acids Res.* **2019**, *47*, e60. [[CrossRef](#)]
153. Mifsud, B.; Tavares-Cadete, F.; Young, A.N.; Sugar, R.; Schoenfelder, S.; Ferreira, L.; Wingett, S.W.; Andrews, S.; Grey, W.; Ewels, P.A.; et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* **2015**, *47*, 598–606. [[CrossRef](#)]
154. Zhang, Y.; An, L.; Xu, J.; Zhang, B.; Zheng, W.J.; Hu, M.; Tang, J.; Yue, F. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat. Commun.* **2018**, *9*, 750. [[CrossRef](#)]
155. Zhu, Y.; Chen, Z.; Zhang, K.; Wang, M.; Medovoy, D.; Whitaker, J.W.; Ding, B.; Li, N.; Zheng, L.; Wang, W. Constructing 3D interaction maps from 1D epigenomes. *Nat. Commun.* **2016**, *7*, 10812. [[CrossRef](#)]
156. Al Bkhetan, Z.; Plewczynski, D. Three-dimensional Epigenome Statistical Model: Genome-wide Chromatin Looping Prediction. *Sci. Rep.* **2018**, *8*, 5217. [[CrossRef](#)]

157. Whalen, S.; Truty, R.M.; Pollard, K.S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* **2016**, *48*, 488–496. [[CrossRef](#)]
158. Wallace, B.C.; Small, K.; Brodley, C.E.; Trikalinos, T.A. Class Imbalance, Redux. In Proceedings of the 2011 IEEE ICDM 11th International Conference on Data Mining, Vancouver, BC, Canada, 11–14 December 2011; pp. 754–763.
159. Sirmacek, B.; Kivits, M. Semantic Segmentation of Skin Lesions using a Small Data Set. *arXiv* **2019**, arXiv:1910.10534.
160. Salman, S.; Liu, X. Overfitting Mechanism and Avoidance in Deep Neural Networks. *arXiv* **2019**, arXiv:1901.06566.
161. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. *arXiv* **2017**, arXiv:1611.03530.
162. Arpit, D.; Jastrzebski, S.; Ballas, N.; Krueger, D.; Bengio, E.; Kanwal, M.S.; Maharaj, T.; Fischer, A.; Courville, A.; Bengio, Y.; et al. A Closer Look at Memorization in Deep Networks. *arXiv* **2017**, arXiv:1706.05394.
163. Rand-Hendriksen, K.; Ramos-Goni, J.M.; Augestad, L.A.; Luo, N. Less Is More: Cross-Validation Testing of Simplified Nonlinear Regression Model Specifications for EQ-5D-5L Health State Values. *Value Health* **2017**, *20*, 945–952. [[CrossRef](#)] [[PubMed](#)]
164. Lever, J.; Krzywinski, M.; Altman, N. Regularization. *Nat. Methods* **2016**, *13*, 803–804. [[CrossRef](#)]
165. Murugan, P.; Durairaj, S. Regularization and Optimization strategies in Deep Convolutional Neural Network. *arXiv* **2017**, arXiv:1712.04711.
166. Collins, A.; Yao, Y. Machine Learning Approaches: Data Integration for Disease Prediction and Prognosis. In *Applied Computational Genomics*; Springer: Berlin, Germany, 2018.
167. Wu, Q.; Boueiz, A.; Bozkurt, A.; Masoomi, A.; Wang, A.; DeMeo, D.L.; Weiss, S.T.; Qiu, W. Deep Learning Methods for Predicting Disease Status Using Genomic Data. *J. Biometr. Biostat.* **2018**, *9*, 417.
168. Wu, S.; Jiang, H.; Shen, H.; Yang, Z. Gene Selection in Cancer Classification Using Sparse Logistic Regression with L1/2 Regularization. *Appl. Sci.* **2018**, *8*, 1569. [[CrossRef](#)]
169. Romero, A.; Carrier, P.L.; Erraqabi, A.; Sylvain, T.; Auvolat, A.; Dejoie, E.; Legault, M.A.; Dubé, M.P.; Hussin, J.G.; Bengio, Y. Diet Networks: Thin Parameters for Fat Genomics. *arXiv* **2016**, arXiv:1611.09340.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).