

## REVIEW ARTICLE

OPEN

# Genome analysis through image processing with deep learning models

Yao-zhong Zhang<sup>1</sup>  and Seiya Imoto<sup>1</sup> 

© The Author(s) 2024



Genomic sequences are traditionally represented as strings of characters: A (adenine), C (cytosine), G (guanine), and T (thymine). However, an alternative approach involves depicting sequence-related information through image representations, such as Chaos Game Representation (CGR) and read pileup images. With rapid advancements in deep learning (DL) methods within computer vision and natural language processing, there is growing interest in applying image-based DL methods to genomic sequence analysis. These methods involve encoding genomic information as images or integrating spatial information from images into the analytical process. In this review, we summarize three typical applications that use image processing with DL models for genome analysis. We examine the utilization and advantages of these image-based approaches.

*Journal of Human Genetics* (2024) 69:519–525; <https://doi.org/10.1038/s10038-024-01275-0>

## INTRODUCTION

Genome analysis has traditionally been conducted through string analysis [1], involving processes such as alignment [2–5], assembly [6–10], and structural variation detection [11, 12]. With the rapid advancement of high-throughput sequencing technologies, genomic data are expanding significantly across various dimensions, including data quality (e.g., read depth and length) and diversity of biological contexts and samples. These advancements present computational challenges in analyzing the increasingly complex and diverse genomic data, particularly at the whole-genome scale.

Meanwhile, deep learning models, such as Convolutional Neural Networks (CNNs), originally developed for applications in computer vision [13], have become powerful tools with potential applications in genome analysis. Deep learning models facilitate the learning of data representations across multiple levels of abstraction and enable the discovery of intricate patterns within large datasets [14]. In this article, we focus on the approach that performs genome analysis through image processing using deep learning models. This approach represents genomic data as image or image-like tensors and utilizes image-based deep learning methods for genome analysis. In the following sections, we begin with data representations for image and genomic data, then revisit three typical applications utilizing this approach. These include converting genomic sequences into image representations, using image recognition algorithms to identify patterns in genomes, and integrating both image and gene expression profiles.

## IMAGE, GENOME, AND DEEP LEARNING MODEL

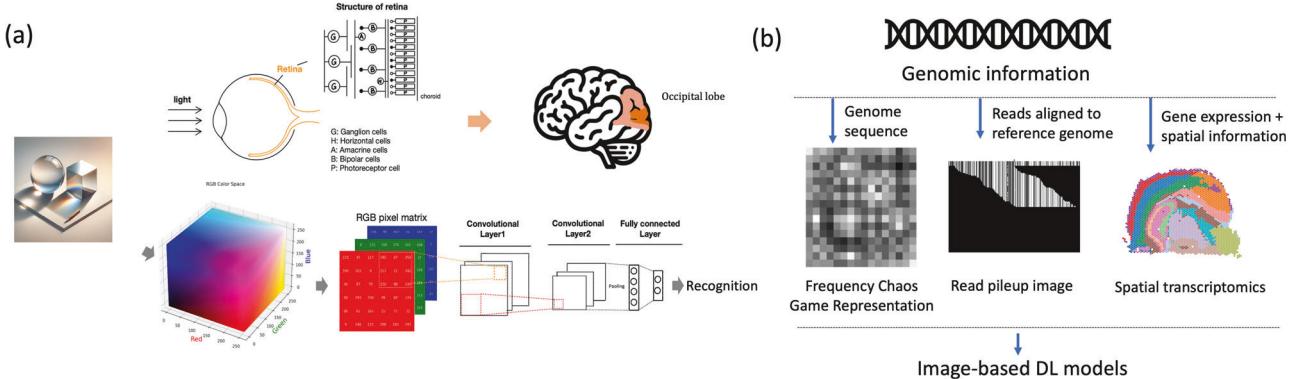
In the context of computer systems, an image is typically represented as a two-dimensional (2D) grid of pixels. Each pixel

captures a portion of the visible light spectrum to recreate the appearance of a scene or object. The color and intensity of each pixel are quantified using color models, which define how colors are represented in the digital space. Common color models include RGB (Red, Green, and Blue), HSV/HSB (Hue, Saturation, and Value/Black), and CMYK (Cyan, Magenta, Yellow, and Key) [15–17]. The RGB model is one of the most common color models used in electronic displays, cameras, and scanners. It represents colors through a combination of red, green, and blue light, where the different intensities of these colors create the wide spectrum of colors that the human eye can perceive, as shown in the color cube of Fig. 1a. Consequently, image manipulations of filtering, transformation, and enhancement are performed based on the pixel matrix [18–20].

To the human eye, an image is perceived by the process of light being captured and focused by the eye's optical system onto the retina. Cells in the retina convert the light into electrical signals and send them to the brain for processing and interpretation [21, 22], as shown in Fig. 1a. The retina is structured into three primary layers: the photoreceptor layer, the inner nuclear layer, and the ganglion cell layer [23]. Photoreceptor cells, known as rods and cones in the photoreceptor layer, convert the light into electrical signals [24]. Rods are responsible for vision in low light conditions and do not detect color, while cones are active in brighter light and enable color vision. The inner nuclear layer is located in the middle and contains several types of cells, including bipolar cells, horizontal cells, and amacrine cells. Bipolar cells receive signals from the photoreceptors and transmit them to the ganglion cells, acting as a direct relay [25]. Horizontal cells connect to multiple photoreceptor cells and are involved in integrating and regulating input to adjust for varying light levels, enhancing contrast. Amacrine cells interact with bipolar and ganglion cells, playing a key role in the timing and shaping of the signal as it passes

<sup>1</sup>Division of Health Medical Intelligence, Human Genome Center, the Institute of Medical Science, the University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.  
✉email: [yaozhong@hgc.jp](mailto:yaozhong@hgc.jp); [imoto@hgc.jp](mailto:imoto@hgc.jp)

Received: 8 January 2024 Revised: 8 July 2024 Accepted: 8 July 2024  
Published online: 31 July 2024



**Fig. 1** Overview of image processing by humans, computer systems, and genome analysis using image-based deep learning techniques. **a** The process of image processing by both humans and computer systems. In computer systems, colors defined by the RGB model are used for image representation and processed by a CNN model. **b** Three applications of genome analysis utilizing image processing with DL models, including genome representation, pattern extraction from read pileup images, and spatial transcriptomics. The pileup image is generated according to DeepVariant (r1.6). The spatial transcriptomics image is generated using the publicly available 10x Genomics Visium mouse brain dataset using squidpy package [80]

through the retina. The integrated signals are then transmitted by the ganglion cells through their axons, which collectively form the optic nerve, to the occipital lobe of the brain [26].

Modern deep learning works by adding more layers and more units within a layer, providing a powerful framework for machine learning [27]. Deep learning models for images have been developed to train on large-scale image data [28–30] for various tasks, including classification [31], recognition [32], segmentation [33], and generation [34, 35]. Taking CNN [13] for instance, the success of AlexNet, a CNN-based model, in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) marked the beginning of an era dominated by modern deep learning models for image analysis [31]. This milestone in artificial intelligence demonstrated that CNNs could achieve remarkable accuracy in classifying and interpreting image data [36]. Convolutional layers, the core components of CNNs, operate through filters or kernels that detect specific features within an image, such as edges, textures, or shapes. This mechanism is analogous to the way the retina's photoreceptors and subsequent neuronal layers respond to specific visual stimuli within their receptive fields. The structure similarity between CNN and retina is demonstrated in Fig. 1a. Based on the stacked convolutional layers, CNN can learn spatial hierarchies of features from input images.

Different from images represented in 2D-pixel matrices, genome sequences are typically represented as one-dimensional (1D) strings of characters. String-based algorithms are extensively used for tasks, such as alignment, searching, and comparison. BLAST (Basic Local Alignment Search Tool) [37] is an example of the string-based method. To leverage the capabilities of image-based deep learning models for analyzing genomic strings, two strategies are commonly used: one is adapting the architecture of models to process 1D sequences. For example, DeepBind adapts CNN for 1D sequence inputs to learn and predict protein-DNA/RNA binding preferences [38]. It utilizes one-hot encoding for nucleotide sequences, representing the nucleotides or k-mers as distinct binary vectors. The encoded sequence is then input into the tailored CNN model (e.g., the sliding window size is  $4 \times K$ , where  $K$  is the window length) to learn and predict binding preferences. Through one-hot encoding, a 1D sequence can be transformed into a 2D matrix, but computational operations and pattern recognition are primarily performed across the sequence's length. This is different from traditional 2D image processing, where operations are applied both horizontally and vertically across the image's spatial dimensions. Distinguished from the above approach, the other is through the representation of genomic information as image or image-like tensors and using image-

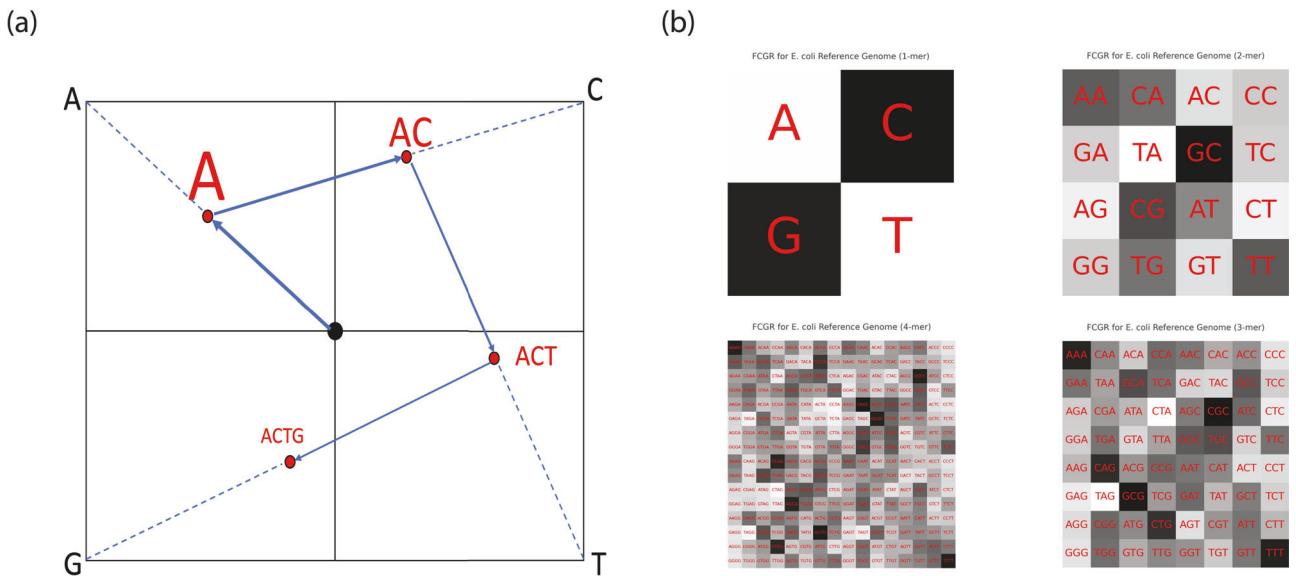
based deep learning models. Here, we focus on the latter approach for genome analysis, which has demonstrated promising results in recent works but is less systematically discussed.

## GENOME ANALYSIS THROUGH IMAGE PROCESSING WITH DEEP LEARNING MODELS

### Converting genome sequences into images

Chaos Game Representation (CGR) [39] emerges as a powerful technique offering a way of portraying genome sequences through fractal images. Originating from the field of mathematical chaos theory, CGR translates the sequential information of a genome into a spatial context, enabling the visualization of nucleotide arrangements and their patterns across entire genomes. This transformation is achieved by representing nucleotides as fixed points in a multi-dimensional space and iteratively plotting the genomic sequence to unravel hidden structures. The efficacy of CGR lies in its ability to condense extensive genomic data into compact, visually interpretable images, preserving the sequential context and highlighting peculiarities that might be elusive in traditional sequential representations. As a result, CGR facilitates a holistic understanding of the genome's structural intricacies, promoting the discovery of functional elements, genomic variations, and evolutionary relationships.

The Frequency Chaos Game Representation (FCGR) extends the CGR by incorporating k-mer frequencies [40]. In FCGR, the bit depth of the CGR image encodes the frequency information of k-mers. Generating an FCGR image involves several key steps. Initially, the length of the k-mer is selected, which influences both the granularity and the complexity of the resulting FCGR image. This selection strikes a balance between resolution and computational efficiency, with longer k-mers providing more detail at the expense of increased computational requirements. After determining the k-mer length, the next step is to calculate the frequency of each k-mer within the target genome. This process involves scanning the entire genomic sequence to count the occurrences of each unique k-mer. These frequencies are then used to construct the FCGR image, where each k-mer corresponds to a specific pixel or a group of pixels. The placement of these pixels in the FCGR image follows the rules of the Chaos Game, which assigns each k-mer a unique location based on its nucleotide composition. This approach transforms the linear, sequential genomic information into a two-dimensional, spatial representation. The result is a fractal-like image that visually encodes the distribution and relationships of k-mers throughout the genome. As demonstrated in Fig. 2b, the visual complexity of



**Fig. 2** Frequency chaos game representation for genome sequences. **a** Position finding algorithm for chaos game representation: the algorithm starts at a fixed point (indicated by a black dot) and progresses to determine the positions for k-mers. These positions, represented by red dots, are calculated as the midpoint between the current position and the position corresponding to the nucleotide of the k-mer. The blue arrows illustrate the sequence of steps to locate the specific position of a k-mer in the FCGR image, with progression from 1-mer to 4-mer. **b** Frequency chaos game representation of the *E. coli* reference genome: this figure depicts the FCGRs for the *E. coli* reference genome across different k-mer lengths, ranging from 1-mer to 4-mer. The gradient from white to black signifies the frequency of each k-mer within the genome. The size of each box in the representation correlates with the resolution of the FCGR image or the length of the k-mer

the FCGR image increases with the k-mer length. More specifically, a 1-mer FCGR might appear relatively simple and uniform, whereas a 4-mer FCGR exhibits a more intricate and detailed pattern. This increase in resolution reveals more nuanced aspects of the genomic sequence, such as the prevalence of certain k-mers or the presence of repetitive elements.

As an image representation, the FCGR can be seamlessly integrated with image-based deep learning models for genome analysis [31, 41, 42]. Mill'an Arias et al. [43] explored the use of deep learning algorithms in conjunction with FCGR for genome clustering, which can effectively process DNA sequences over 1 billion bp and outperform two classic clustering methods (K-means ++ and Gaussian Mixture Models) for unlabelled data. Hammad et al. [44] used a pre-trained convolutional neural network to extract features from FCGR images and selected the most significant features for downstream classifiers for COVID-19 detection. Besides using deep learning models with FCGR representation, Zhang et al. proposed to use contrastive learning to further learn the representation that integrates phage-host interactions based on FCGR [45]. In addition, they conducted an ablation study to demonstrate the performance gains achieved through the 2D representation of the k-mer vectors used in conjunction with the CNN. Compared to the use of k-mer frequency features as a 1D vector, FCGR introduces an additional dimension by arranging k-mers such that higher-order features can be extracted. For example, k-mers sharing the same suffix are positioned in close proximity (e.g., 3-mers of {A/C/G/T}AA as illustrated in the top left corner of the 3-mer FCGR in Fig. 2(b)). Consequently, when applying a CNN to FCGR, the convolutional module can effectively extract features associated with k-mers that have identical suffixes. The synergy between FCGR and advanced deep learning methods potentially leads to more powerful tools for genome analysis.

Besides their application using deep learning models, CGR and FCGR can be employed with alignment-free algorithms for genome comparison and phylogenetic analysis [39, 46, 47]. For instance, Jeffrey demonstrated how CGR could be used to visually differentiate between the genomic sequences of various

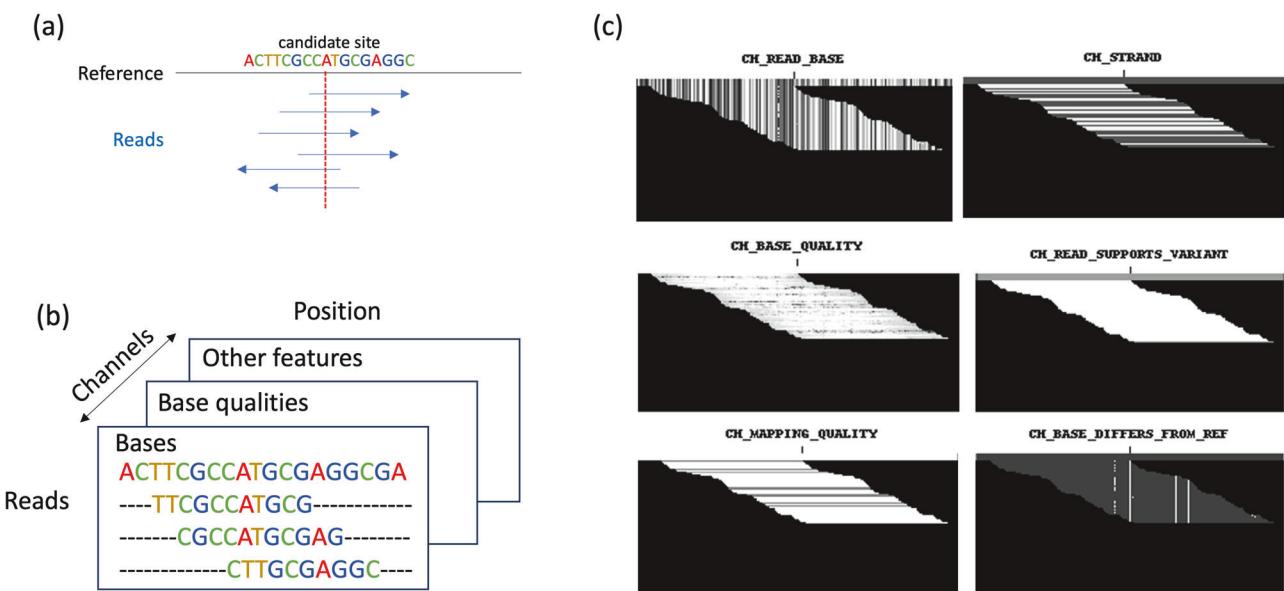
organisms, offering an effective approach to studying phylogenetic relationships [39]. Local and global patterns between genome sequences can be compared without alignment to reference genomes. The work by Deschavanne et al. [40] showed the effectiveness of using FCGR as signature patterns for genomic sequences, and the utility of image distances as a measure of phylogenetic proximity.

When comparing string-based and image-based methods, it is evident that each has its unique advantages and limitations. The string-based analysis is well-established and efficient for raw sequence data processing, but it may lack the intuitive insight provided by a visual representation. Image-based analysis, on the other hand, offers a more intuitive and detailed view of complex genomic patterns but requires more computational resources and expertise in image processing and deep learning. In practice, the choice between these methods depends on the specific requirements of the research. For tasks requiring direct manipulation and comparison of raw sequences, string-based methods are more suitable. Conversely, for applications where pattern recognition and complex relationship analysis are crucial, image-based methods may offer insights.

### Detecting patterns in read pileup images

Image-based deep learning methods can be utilized in processing information derived from sequencing data. Genome analysis is a complex and multifaceted discipline that involves the interpretation of vast amounts of sequencing data to uncover genetic variations and understand their implications. To identify genomic variants from sequencing data, the first step is to map sequencing reads to the reference genome (Fig. 3a). Alignment information is crucial for detecting genomic variants such as single nucleotide variants (SNVs), insertions, deletions, and more complex structural changes [48].

Traditionally, read alignments and variant calling rely on text-based representations and statistical approaches [49, 50]. With the advent of deep learning methods in bioinformatics, there has been a paradigm shift towards more sophisticated and data-driven methodologies. One representative method is DeepVariant



**Fig. 3** Read pileup images generated by DeepVariant. **a** Sequencing reads aligned to the reference genome. **b** Read pileup images derived from read alignment. **c** Pileup images of different feature channels generated by DeepVariant, with the example of Pileup chr20:10001627\_G→A (GRCh38.p13 build) provided by DeepVariant

[51], which converts the read alignment data into read pileup images for variant calling. The read alignment information is transformed into a multi-channel image-like tensor, in which each channel corresponds to a specific aspect of the data, including features of read base, base quality, mapping quality, strand information, read support variants and base differs from the reference genome, as shown in Fig. 3b, c. The information can be equivalent to those shown in the IGV [52] that are commonly used in human experts' manual evaluation of a putative variant. This image representation can capture the spatial relationships and patterns within the read alignments, providing a rich and contextual view of the candidate genomic region. For read pileup images, DeepVariant uses CNNs to predict the likelihood of different genotypes of homozygous reference (hom-ref), heterozygous (het), or homozygous alternate (hom-alt). The CNNs in DeepVariant are used to consider extensive contextual genomic alignment information, ensuring that the predictions are not solely based on the immediate read alignments. Additionally, these predictions are also influenced by the broader genomic context and overall aligned reads. More importantly, such a method can generalize well across different genomes and different sequencing platforms for variant calling [51, 53]. Not limited to the CNN model, other deep learning models, such as generative adversarial networks, have been applied for pileup image features [54]. Besides detecting SNVs and small-indel variants, extensions of read pileup images have been developed for calling structural variants (SVs) [55–57]. A recently proposed method named Cue [56] transforms read alignments to images for encoding SV abstractions and uses a specific CNN to predict SVs. Such a method does not rely on hand-crafted features, demonstrating enhanced generalization capabilities and competitive performance.

A similar image visualization technique has also been employed in nanopore sequencing. For nanopore basecalling, short-tandem repeats (STRs) tend to exhibit a higher error rate. To address this challenge, Fang et al. [58] developed DeepRepeat, a method that transforms ionic current signals of reads into RGB channels for predicting repeats. More specifically, the current signals of STR and its upstream and downstream units are converted into RGB channels of a color image. In the image, the height and width represent signal range and STR unit size, respectively. Images from

repeat and non-repeat regions are used to train a CNN model for identifying the presence of repeats. Their experimental results demonstrate that this approach can effectively detect STR regions, showing advantages over previous non-image-based methods.

#### Spatial transcriptomics: integration of spatial image information and gene expression data

Different from the previous two applications that transform genomic information into image representations, gene expression profiles can be analyzed in conjunction with image data. Spatial transcriptomic (ST) integrates spatial information from tissue organization with gene expression profiles, which extends the ability for the genomic understanding of tissue organization, cellular interactions, and disease [59]. For instance, in cancer research, ST can reveal how different regions of a tumor may express distinct gene profiles, potentially leading to variations in treatment response and prognosis [60]. In neuroscience, ST can help in mapping the diverse gene expression patterns in different brain regions, offering insights into the molecular mechanisms underlying brain function and disorders [61]. The spatial dimension introduces an additional layer of information not available through traditional transcriptomic analysis.

According to different visualization and transcript detection methods, ST technologies can be classified into two main categories: imaging-based ST and sequencing-based ST. Imaging-based ST techniques, such as seqFISH [62], MERFISH [63], and ISS [64], employ fluorescence microscopy to probe specific transcripts within a slice at single-cellular resolution. Sequencing-based ST techniques, such as 10X Visium and Stereo-seq [65], capture the entire transcriptome at spot resolution through spatial barcoding. Currently, most sequencing-based ST platforms provide a spot resolution that contains 10 to 15 cells per spot [66].

Given the multi-modal nature of ST data and variability in techniques, analyzing raw ST data is far from trivial [67]. On the one hand, accurate processing of image and transcript data is essential to obtain spatial coordinates and gene expression profiles. On the other hand, integrating image and transcript data adds another layer of complexity. For imaging-based ST data analysis, image processing techniques such as correction, registration, and segmentation are routinely employed. These

processing techniques are essential to enhance the quality and reliability of the data, facilitating more accurate mapping of gene expression within tissue sections [68]. Identifying cell boundaries is fraught with challenges due to the high variability in cell shapes, sizes, and densities, especially in heterogeneous tissue samples. Advanced image processing algorithms are required to overcome these challenges, ensuring precise and accurate segmentation. Even for ST techniques that require minimal image processing, additional steps are necessary after sequencing to map transcripts back to their spatial coordinates [69]. The accompanying tissue images can provide useful information for this mapping process, enhancing the accuracy and effectiveness of spatial localization. Upon acquiring a transcriptomically and spatially coherent mapping, the processed ST data can be transformed into image-like tensors for downstream analysis. Similar to RGB channels in an image, both color information and gene expression profiles can be encoded as channels to represent ST data. Deep learning models are effective at handling multi-modal data, making them suitable for analyzing both image and gene expression data in spatial transcriptomics.

Deep learning models have been employed in various applications, including cell/nuclei segmentation [70], alignment [71], spot deconvolution [72], and spatial clustering [73–75]. Various deep learning architectural frameworks, such as CNNs, autoencoders [76], U-nets [70], graph neural networks [77], and transformers [78] have been utilized either individually or in combination to address these challenges. For instance, Tangram [71] integrates molecular and anatomical features by combining a Siamese neural network with a segmentation U-net model to generate full segmentation masks of anatomical images. The U-net architecture is built based on a ResNet50 backbone. SpaCell [73] employs a pre-trained ResNet50 (trained on ImageNet) along with two separate autoencoders to generate a latent matrix representative for both image and gene-count data. SpaGCN [74] used a graph convolutional network for a constructed undirected weighted graph that integrates gene expression data, spatial location, and histological information for representing the spatial dependencies within the data. Chen et al. proposed sub-cellular spatial transcriptomics cell segmentation (SCS) to combine image data with sequencing data to improve cell segmentation accuracy based on a transformer model [66]. The proposed SCS method demonstrates a better performance when compared with traditional image-based segmentation methods.

## DISCUSSION AND CONCLUDING REMARKS

In this article, we reviewed the methodologies that conduct genome analysis through image processing with deep learning methods. The efficacy of image visualization techniques in genome analysis can be attributed to the effective information representation and the usage of advanced deep learning models. With the advancements in sequencing technologies, genomic profiling has become richer and more comprehensive. This expands the versatility of genomic data beyond the traditional string-based formats. Representing genomic information as image inputs provides a solution to represent various genomic data in 2D space with multiple channels. The image representation can be further processed with the advanced deep learning models for genome analysis. Furthermore, image processing is indispensable for processing image data in spatial transcriptomic analysis. Distinguished from the traditional one-hot encoding approach, using image-based deep learning models for genome analysis can be characterized by the addition of extra dimensions beyond the sequential level. Such an extension further leverages the capabilities of deep learning models to extract higher-order features or patterns in dimensions beyond 1D space.

Despite the promise of utilizing image-based deep learning models for genome analysis, there are challenges to be addressed. While genomic data can be converted into image input for analysis with deep learning models, it's important to be aware of the different data characteristics. Genomic data is typically categorical, and the commonly used categorical representation lacks smoothness, presenting challenges for training deep learning models [79]. On the other hand, the interpretation of results from deep learning models can be complex, and there is a need for more transparent and interpretable models. Additionally, the computational demands of training and applying these models, particularly on large genomic datasets, are significant.

Looking ahead, image visualization for genome analysis is moving towards the development of more efficient, interpretable, and scalable deep learning models for genome analysis. There is also an increasing focus on integrating multi-omics data, combining genomic, transcriptomic, and proteomic information, to provide a more holistic view of biological systems. Genome analysis through image processing can provide a promising solution to achieve this goal.

## REFERENCES

- Durbin R, Eddy SR, Krogh A, Mitchison G. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge university press; 1998.
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–3100.
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. Mummer4: A fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14:1005944.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. Star: ultrafast universal rna-seq aligner. *Bioinformatics*. 2013;29:15–21.
- Sohn J-i, Nam J-W. The present and future of de novo whole-genome assembly. *Brief Bioinforma*. 2018;19:23–40.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. Megahit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6.
- Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun*. 2021;12:60.
- Idury RM, Waterman MS. A new algorithm for dna sequence assembly. *J Comput Biol*. 1995;2:291–306.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. Abyss: a parallel assembler for short read sequence data. *Genome Res*. 2009;19:1117–23.
- Li Y, Zheng H, Luo R, Wu H, Zhu H, Li R, et al. Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. *Nat Biotechnol*. 2011;29:723–30.
- Consortium GP, Auton A, Brooks L, Durbin R, Garrison E, Kang H. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278–324.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Maxwell JC. On the theory of the three primary colours. *Trans R Soc Edinb*. 1860;21:275–98.
- Smith AR. Color gamut transform pairs. *ACM SIGGRAPH Computer Graph*. 1978;12:12–19.
- Kang HR. Digital color halftoning. SPIE/IEEE Series on Imaging Science and Engineering. 1999.
- Chandell R, Gupta G. Image filtering algorithms and techniques: A review. *Int J Adv Res Computer Sci Softw Eng*. 2013;3:198–202.
- Petrou MM, Petrou C. Image processing: the fundamentals. John Wiley & Sons; 2010, pp. 47–116.
- Singh G, Mittal A. Various image enhancement techniques-a critical review. *Int J Innov Sci Res*. 2014;10:267–74.
- London A, Benhar I, Schwartz M. The retina as a window to the brain—from eye research to cns disorders. *Nat Rev Neurol*. 2013;9:44–53.
- Hussey KA, Hadyniak SE, Johnston RJ Jr. Patterning and development of photoreceptors in the human retina. *Front Cell Developmental Biol*. 2022;10:878350.
- Cammalleri M, Bagnoli P, Bigiani A. Molecular and cellular mechanisms underlying somatostatin-based signaling in two model neural networks, the retina and the hippocampus. *Int J Mol Sci*. 2019;20:2506.

24. Schnapf JL, Baylor DA. How photoreceptor cells respond to light. *Sci Am*. 1987;256:40–47.
25. Shiells R. Photoreceptor-bipolar cell transmission. In: *Neurobiology and Clinical Aspects of the Outer Retina*. Springer; 1995, pp. 297–324.
26. Erskine L, Herrera E. The retinal ganglion cell axon's journey: insights into molecular mechanisms of axon guidance. *Developmental Biol*. 2007;308:1–14.
27. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT press, (2016).
28. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009. IEEE Conference on Computer Vision and Pattern Recognition. Ieee; 2009. pp. 248–55.
29. Krizhevsky A. Learning multiple layers of features from tiny images. 2009.
30. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: Common objects in context. In: Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13, pp. 740–55 (2014).
31. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Info Processing Syst*. 2012;25.
32. Uchida S. Image processing and recognition for biological images. *Dev, growth Differ*. 2013;55:523–49.
33. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–40 (2015).
34. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Commun ACM*. 2020;63:139–44.
35. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst*. 2020;33:6840–51.
36. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vision (IJCV)* 2015;115:211–52 <https://doi.org/10.1007/s11263-015-0816-y>.
37. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. Ncbi blast: a better web interface. *Nucleic acids Res*. 2008;36:5–9.
38. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nat Biotechnol*. 2015;33:831–8.
39. Jeffrey HJ. Chaos game representation of gene structure. *Nucleic acids Res*. 1990;18:2163–70.
40. Deschavanne PJ, Giron A, Vilain J, Fagot G, Fertil B. Genomic signature: characterization and classification of species assessed by chaos game representation of sequences. *Mol Biol Evol*. 1999;16:1391–9.
41. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Computer Vision-ECCV. 2016: 14th European Conference. Amsterdam, The Netherlands: Springer; 2016. p. 630–45. October 11–14, 2016, Proceedings, Part IV 14.
42. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. arXiv preprint. An image is worth 16x16 words: Transformers for image recognition at scale. 2020. arXiv:2010.11929.
43. Millán Arias P, Alipour F, Hill KA, Kari L. Delucs: Deep learning for unsupervised clustering of dna sequences. *Plos one*. 2022;17:0261531.
44. Hammad MS, Ghoneim VF, Mabrouk MS, Al-Atabany WI. A hybrid deep learning approach for covid-19 detection based on genomic image processing techniques. *Sci Rep*. 2023;13:4003.
45. Zhang Y-z, Liu Y, Bai Z, Fujimoto K, Uematsu S, Imoto S. Zero-shotcapable identification of phage-host relationships with whole-genome sequence representation by contrastive learning. *Brief Bioinforma*. 2023;24:239.
46. Joseph J, Sasikumar R. Chaos game representation for comparison of whole genomes. *BMC Bioinforma*. 2006;7:1–10.
47. Lichtblau D. Alignment-free genomic sequence comparison using fcgr and signal processing. *BMC Bioinforma*. 2019;20:1–17.
48. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and snp calling from next-generation sequencing data. *Nat Rev Genet*. 2011;12:443–51.
49. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 2018;201178. <https://www.biorxiv.org/content/10.1101/201178v3.abstract>.
50. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*. 2012.
51. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal snp and small-indel variant caller using deep neural networks. *Nat Biotechnol*. 2018;36:983–7.
52. Robinson JT, Thorvaldsd'ottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–26.
53. Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, et al. Haplotype-aware variant calling with pepper-margin-deepvariant enables high accuracy in nanopore long-reads. *Nat methods*. 2021;18:1322–32.
54. Yang H, Gu F, Zhang L, Hua X-S. Using generative adversarial networks for genome variant calling from low depth ont sequencing data. *Sci Rep*. 2022;12:8725.
55. Cai L, Wu Y, Gao J. Deepsv: accurate calling of genomic deletions from high-throughput sequencing data using deep convolutional neural network. *BMC Bioinforma*. 2019;20:1–17.
56. Popic V, Rohlicek C, Cunial F, Hajirasouliha I, Meleshko D, Garimella K, et al. Cue: a deep-learning framework for structural variant discovery and genotyping. *Nat Methods*. 2023;20:559–68.
57. Ye K, Wang S, Lin J, Jia P, Xu T, Meng D. Svision-pro: comparative sequence-to-image representation and instance segmentation for de novo and somatic structural variant discovery. 2023.
58. Fang L, Liu Q, Monteyns AM, Gonzalez-Alegre P, Davidson BL, Wang K. Deeprepeat: direct quantification of short tandem repeats on signal data from nanopore sequencing. *Genome Biol*. 2022;23:108.
59. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*. 2015;348:6090.
60. Yu Q, Jiang M, Wu L. Spatial transcriptomics technology in cancer research. *Front Oncol*. 2022;12:1019111.
61. Close JL, Long BR, Zeng H. Spatially resolved transcriptomics in neuroscience. *Nat methods*. 2021;18:23–25.
62. Shah S, Lubeck E, Zhou W, Cai L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron*. 2016;92:342–57.
63. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. Spatially resolved cell atlas of the mouse primary motor cortex by merfish. *Nature*. 2021;598:137–43.
64. Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, et al. In situ sequencing for rna analysis in preserved tissue and cells. *Nat methods*. 2013;10:857–60.
65. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*. 2022;185:1777–92.
66. Chen H, Li D, Bar-Joseph Z. Cell segmentation for high-resolution spatial transcriptomics. In: Research in Computational Molecular Biology: 27th Annual International Conference, RECOMB 2023, Istanbul, Turkey, April 16–19, 2023, Proceedings, vol. 13976, p. 251 (2023).
67. Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan G-C. Advances in spatial transcriptomic data analysis. *Genome Res*. 2021;31:1706–18.
68. Petukhov V, Xu RJ, Soldatov RA, Cadina P, Khodosevich K, Moffitt JR, et al. Cell segmentation in imaging-based spatial transcriptomics. *Nat Biotechnol*. 2022;40:345–54.
69. Palla G, Fischer DS, Regev A, Theis FJ. Spatial components of molecular tissue biology. *Nat Biotechnol*. 2022;40:308–18.
70. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Vol. 36 pp. 234–41 (2015).
71. Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat methods*. 2021;18:1352–62.
72. Lopez R, Li B, Keren-Shaul H, Boyeau P, Kedmi M, Pilzer D, et al. Destvi identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol*. 2022;40:1360–9.
73. Tan X, Su A, Tran M, Nguyen Q. Spacell: integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics*. 2020;36:2293–4.
74. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. Spagn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat methods*. 2021;18:1342–51.
75. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun*. 2022;13:1739.
76. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science*. 2006;313:504–7.
77. Scarselli F, Gorri M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. *IEEE Trans Neural Netw*. 2008;20:61–80.
78. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Info Processing Syst*. 2017;30:6000–10.
79. Xia J, Zhang L, Zhu X, Liu Y, Gao Z, Hu B, et al. Understanding the limitations of deep models for molecular property prediction: Insights and solutions. In: Thirty-seventh Conference on Neural Information Processing Systems, Vol. 36. 2024.
80. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. *Nat methods*. 2022;19:171–8.

**ACKNOWLEDGEMENTS**

We thank Yuxuan Pang for the discussion on the spatial transcriptome section.

**FUNDING**

Open Access funding provided by The University of Tokyo.

**COMPETING INTERESTS**

The authors declare no competing interests.

**ADDITIONAL INFORMATION**

**Correspondence** and requests for materials should be addressed to Yao-zhong Zhang or Seiya Imoto.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024