CLAIMS

[Claim 1]. A computer-implemented method of determining metadata about clusters on a tile of a flow cell, the method comprising:

accessing a series of image sets generated during a sequencing run, each image set in the series generated during a respective sequencing cycle of the sequencing run, each image in the series depicting clusters and their surrounding background, each image in the series having pixels in a pixel domain, and each of the pixels is divided into a plurality of subpixels in a subpixel domain;

obtaining, from a base caller, a base call classifying each of the subpixels as one of four bases (A, C, T, and G), thereby producing a base call sequence for each of the subpixels across a plurality of sequencing cycles of the sequencing run;

generating a cluster map that identifies the clusters as disjointed regions of contiguous subpixels which share a substantially matching base call sequence; and

determining cluster metadata based on the disjointed regions in the cluster map,

wherein the cluster metadata includes cluster centers, cluster shapes, cluster sizes, cluster background, and/or cluster boundaries.

[Claim 2]. The computer-implemented method of claim 1, further including:

using the cluster metadata to generate ground truth training data for training a neural network-based template generator for cluster metadata determination task,

wherein the ground truth training data comprises a decay map, a ternary map, or a binary map,

wherein the neural network-based template generator is trained to produce the decay map, the ternary map, or the binary map as output based on the ground truth training data, and

wherein, upon execution of the cluster metadata determination task during inference, the cluster metadata is in turn determined from the decay map, the ternary map, or the binary map that are produced as the output by the trained neural network-based template generator.

[Claim 3]. The computer-implemented method of claim 2, further including:

using the cluster metadata derived from the decay map, the ternary map, or the binary map produced as the output by the neural network-based template generator for base calling by a neural network-based base caller, in order to increase throughput in high-throughput nucleic acid sequencing technologies.

[Claim 4]. The computer-implemented method of any of claims 1-3, further including:

generating the cluster map by identifying as background those subpixels that do not belong to any of the disjointed regions.

[Claim 5]. The computer-implemented method of any of claims 1-4, wherein the cluster map identifies cluster boundary portions between two contiguous subpixels whose base call sequences do not substantially match.

[Claim 6]. The computer-implemented method of any of claims 1-5, wherein the cluster map is generated based on:

identifying origin subpixels at preliminary center coordinates of the clusters determined by the base caller; and

breadth-first searching for substantially matching base call sequences by beginning with the origin

subpixels and continuing with successively contiguous non-origin subpixels.

[Claim 7]. The computer-implemented method of any of claims 2-6, further including:

determining hyperlocated center coordinates of the clusters by calculating centers of mass of the disjointed regions of the cluster map as an average of coordinates of respective contiguous subpixels forming the disjointed regions; and

storing the hyperlocated center coordinates of the clusters in the memory for use as the ground truth training data for training the neural network-based template generator.

[Claim 8]. The computer-implemented method of any of claims 2-7, further including:

identifying centers of mass subpixels in the disjointed regions of the cluster map at the hyperlocated center coordinates of the clusters;

upsampling the cluster map using interpolation and storing the upsampled cluster map in the memory for use as the ground truth training data for training the neural network-based template generator; and

in the upsampled cluster map, assigning a value to each contiguous subpixel in the disjointed regions based on a decay factor that is proportional to distance of a contiguous subpixel from a center of mass subpixel in a disjointed region to which the contiguous subpixel belongs.

[Claim 9]. The computer-implemented method of any of claims 2-8, further including:

generating the decay map from the upsampled cluster map that expresses the contiguous subpixels in the disjointed regions and the subpixels identified as the background based on their assigned values; and

storing the decay map in the memory for use as the ground truth training data for training the neural network-based template generator.

[Claim 10]. The computer-implemented method of claim 8 or 9, further including:

in the upsampled cluster map, categorizing, on the cluster-by-cluster basis, the contiguous subpixels in the disjointed regions as cluster interior subpixels belonging to a same cluster, the centers of mass subpixels as cluster center subpixels, subpixels containing the cluster boundary portions as boundary subpixels, and the subpixels identified as the background as background subpixels; and

storing the categorizations in the memory for use as the ground truth training data for training the neural network-based template generator.

[Claim 11]. The computer-implemented method of claim 10, further including:

storing, on the cluster-by-cluster basis, coordinates of the cluster interior subpixels, the cluster center subpixels, the boundary subpixels, and the background subpixels in the memory for use as the ground truth training data for training the neural network-based template generator;

downscaling the coordinates by a factor used to upsample the cluster map; and

storing, on the cluster-by-cluster basis, the downscaled coordinates in the memory for use as the ground truth training data for training the neural network-based template generator.

[Claim 12]. The computer-implemented method of any of claims 1-11, further including:

generating cluster maps for a plurality of tiles of the flow cell;

storing the cluster maps in memory and determining the cluster metadata of clusters in the tiles based on the cluster maps, including the cluster centers, the cluster shapes, the cluster sizes, the cluster background, and/or the cluster boundaries;

in the upsampled cluster maps of the clusters in the tiles, categorizing, on a cluster-by-cluster basis, subpixels as cluster interior subpixels belonging to a same cluster, cluster center subpixels, boundary subpixels, and background subpixels;

storing the categorizations in the memory for use as the ground truth training data for training the neural network-based template generator;

storing, on the cluster-by-cluster basis across the tiles, coordinates of the cluster interior subpixels, the cluster center subpixels, the boundary subpixels, and the background subpixels in the memory for use as the ground truth training data for training the neural network-based template generator;

downscaling the coordinates by the factor used to upsample the cluster map; and

storing, on the cluster-by-cluster basis across the tiles, the downscaled coordinates in the memory for use as the ground truth training data for training the neural network-based template generator.

[Claim 13]. The computer-implemented method of any of claims 1-12, wherein the base call sequences are substantially matching when a predetermined portion of base calls match on an ordinal position-wise basis.

[Claim 14]. The computer-implemented method of any of claims 1-13, wherein the cluster map is generated based upon a predetermined minimum number of subpixels for a disjointed region.

[Claim 15]. The computer-implemented method of any of claims 1-14, wherein the flow cell has at least one patterned surface with an array of wells that occupy the clusters, further including:

based on the determined shapes and sizes of the clusters, determining

which ones of the wells are substantially occupied by at least one cluster,

which ones of the wells are minimally occupied, and

which ones of the wells are co-occupied by multiple clusters.

[Claim 16]. A non-transitory computer-readable medium loaded with instructions that, when executed on a processor, cause the processor to implement the method of any of claims 1-15.

[Claim 17]. A computer system, including: a processor and memory, the loaded with instructions that, when executed on the processor, cause the processor to implement the method of any of claims 1-15.