CLAIMS

What is claimed:

1. A method of spatial analysis of analytes comprising:

A) placing a sample on a substrate, wherein the substrate comprises a plurality of fiducial markers and a set of capture spots, wherein the set of capture spots comprises at least 1000 capture spots;

B) obtaining one or more images of the sample on the substrate, wherein each respective image of the one or more images comprises a corresponding plurality of pixels in the form of a corresponding array of pixel values, wherein the corresponding array of pixel values comprises at least 100,000 pixel values;

C) obtaining a plurality of sequence reads, in electronic form, from the set of capture spots after the A) placing, wherein:

each respective capture probe plurality in a set of capture probe pluralities is (i) at a different capture spot in the set of capture spots and (ii) directly or indirectly associates with one or more analytes from the sample,

each respective capture probe plurality in the set of capture probe pluralities is characterized by at least one unique spatial barcode in a plurality of spatial barcodes,

the plurality of sequence reads comprises sequence reads corresponding to all or portions of the one or more analytes, and

the plurality of sequence reads comprises at least 10,000 sequence reads, and

each respective sequence read in the plurality of sequence reads includes a spatial barcode of the corresponding capture probe plurality in the set of capture probe pluralities or a complement thereof;

D) using all or a subset of the plurality of spatial barcodes to localize respective sequence reads in the plurality of sequence reads to corresponding capture spots in the set of capture spots, thereby dividing the plurality of sequence reads into a plurality of subsets of sequence reads, each respective subset of sequence reads corresponding to a different capture spot in the plurality of capture spots; and

E) using the plurality of fiducial markers to provide a composite representation comprising (i) the one or more images aligned to the set of capture spots on the substrate and (ii) a representation of all or a portion of each subset of sequence reads at each respective position within the one or more images that maps to a respective capture spot corresponding to the respective position of the one or

more analytes in the sample by aligning a first image in the one or more images to the set of capture spots by a procedure that comprises:

analyzing the corresponding array of pixel values of the first image to identify a plurality of derived fiducial markers of the first image;

using a substrate identifier uniquely associated with the substrate to select a first template in a plurality of templates, wherein each template in the plurality of templates comprises reference positions for a corresponding plurality of reference fiducial markers and a corresponding coordinate system;

aligning the plurality of derived fiducial markers of the first image with the corresponding plurality of reference fiducial markers of the first template using an alignment algorithm to obtain a transformation between the plurality of derived fiducial markers of the first image and the corresponding plurality of reference fiducial markers of the first template; and

using the transformation and the coordinate system of the first template to locate a corresponding position in the first image of each capture spot in the set of capture spots

by running a plurality of heuristic classifiers on the plurality of pixels, wherein, for each pixel in the plurality of pixels, each respective heuristic classifier in the plurality of heuristic classifiers casts a vote for the respective pixel between a first class and the second class, thereby forming a corresponding aggregated score for each respective pixel in the plurality of pixels, and applying the aggregated score and intensity of each respective pixel in the plurality of pixels to a segmentation algorithm to independently assign a probability to each respective pixel in the plurality of pixels of being sample or background.

2. The method of claim 1, wherein the composite representation provides a relative abundance of nucleic acid fragments mapping to each analyte in a plurality of analytes at each capture spot in the plurality of capture spots.

3. The method of claim 1, wherein each corresponding aggregated score is a class in a set of classes comprising obvious first class, likely first class, likely second class, and obvious second class.

4. The method of claim 1, wherein the method further comprises, for each respective locus in a plurality of loci, performing a procedure that comprises:

i) performing an alignment of each respective sequence read in the plurality of sequence reads that maps to the respective locus thereby determining a haplotype identity for the respective

sequence read from among a corresponding set of haplotypes for the respective locus, and

ii) categorizing each respective sequence read in the plurality of sequence reads that maps to the respective locus by the spatial barcode of the respective sequence read and by the haplotype identity,

thereby determining a spatial distribution of each haplotype in each corresponding set of haplotypes in the sample, wherein the spatial distribution includes, for each capture spot in the set of capture spots on the substrate, an abundance of each haplotype in the set of haplotypes for the respective locus.

5. The method of claim 4, the method further comprises using the spatial distribution to characterize a biological condition in a subject.

6. The method of claim 2, the method further comprising:

overlaying a mask on the first image, wherein the mask causes each respective pixel in the plurality of pixels of the first image that has been assigned a greater probability of being sample to be assigned a first attribute and each respective pixel in the plurality of pixels that has been assigned a greater probability of being background to be assigned a second attribute.

7. The method of claim 6, wherein the first attribute is a first color and the second attribute is a second color.

8. The method of claim 7, wherein the first color is one of red and blue and the second color is the other of red and blue.

9. The method of claim 6, wherein the first attribute is a first level of brightness or opacity and the second attribute is a second level of brightness or opacity.

10. The method of claim 6, the method further comprising:

assigning each respective representation, of a capture spot in the plurality of capture spots in the composite representation, the first attribute or the second attribute based upon the independent assignment of pixels in the vicinity of the respective representation of the capture spot in the composite representation.

11. The method of claim 1, wherein a capture spot in the set of capture spots comprises a capture domain.

12. The method of claim 1, wherein a capture spot in the set of capture spots comprises a cleavage domain.

13. The method of claim 1, wherein each capture spot in the set of capture spots is attached directly or is attached indirectly to the substrate.

14. The method of claim 1, wherein the one or more analytes comprises five or more analytes, ten or more analytes, fifty or more analytes, one hundred or more analytes, five hundred or more analytes, 1000 or more analytes, 2000 or more analytes, or between 2000 and 100,000 analytes.

15. The method of claim 1, wherein the unique spatial barcode encodes a unique predetermined value selected from the set {1, …, 1024}, {1, …, 4096}, {1, …, 16384}, {1, …, 65536}, {1, …, 262144}, {1, …, 1048576}, {1, …, 4194304}, {1, …, 16777216}, {1, …, 67108864}, or {1, …, 1×10^{12}}.

16. The method of any one of claims 1-15, wherein a respective capture probe plurality in the set of capture probe pluralities includes 1000 or more capture probes, 2000 or more capture probes, 10,000 or more capture probes, 100,000 or more capture probes, 1×10^6 or more capture probes, 2×10^6 or more capture probes, or 5×10^6 or more capture probes.

17. The method of claim 16, wherein each capture probe in the respective capture probe plurality includes a poly-T sequence and the unique spatial barcode that characterizes the different capture spot.

18. The method of claim 16, wherein each capture probe in the respective capture probe plurality includes the same spatial barcode from the plurality of spatial barcodes.

19. The method of claim 16, wherein each capture probe in the respective capture probe plurality includes a different spatial barcode from the plurality of spatial barcodes.

20. The method of claim 1, wherein the sample is a sectioned tissue sample having a depth of 100 microns or less.

21. The method of claim 20, wherein the one or more images comprises a plurality of images and a first image in the plurality of images is obtained using a first section of the sample and a second image in the plurality of images is obtained using a second section of the sample.

22. The method of claim 1, wherein

the one or more analytes is a plurality of analytes,

a respective capture probe plurality in the set of capture probe pluralities includes a plurality of capture probes, each capture probe in the plurality of capture probes including a capture domain that is characterized by a capture domain type in a plurality of capture domain types, and

each respective capture domain type in the plurality of capture domain types is configured to bind to a different analyte in the plurality of analytes.

23. The method of claim 22, wherein the plurality of capture domain types comprises between 2 and 15,000 capture domain types and the respective capture probe plurality includes at least five, at least 10, at least 100, or at least 1000 capture probes for each capture domain type in the plurality of capture domain types.

24. The method of claim 1, wherein

the one or more analytes is a plurality of analytes,

a respective capture spot in the set of capture spots includes a plurality of capture probes, each capture probe in the plurality of capture probes including a capture domain that is characterized by a single capture domain type configured to bind to each analyte in the plurality of analytes in an unbiased manner.

25. The method of claim 1, wherein each capture spot in at least 30 percent, at least forty percent, at least fifty percent, at least sixty percent, at least seventy percent, at least eighty percent, or at least ninety percent of the capture spots in the set of capture spots is contained within a corresponding 100 micron by 100 micron square on the substrate.

26. The method of claim 1, wherein a distance between a center of each respective capture spot to a

neighboring capture spot in the set of capture spots on the substrate is between 50 microns and 300 microns.

27. The method of claim 1, wherein at least 30 percent, at least forty percent, at least fifty percent, at least sixty percent, at least seventy percent, at least eighty percent, or at least ninety percent of the capture spots in the set of capture spots has a diameter of 80 microns or less.

28. The method of claim 1, wherein at least 30 percent, at least forty percent, at least fifty percent, at least sixty percent, at least seventy percent, at least eighty percent, or at least ninety percent of the capture spots in the set of capture spots has a diameter of between 25 microns and 65 microns.

29. The method of claim 1, wherein a distance between a center of each respective capture spot to a neighboring capture spot in the set of capture spots on the substrate is between 40 microns and 100 microns.

30. The method of claim 2, wherein the plurality of heuristic classifier comprises a first heuristic classifier that identifies a single intensity threshold that divides the plurality of pixels into the first class and the second class, thereby causing the first heuristic classifier to cast a vote for each respective pixel in the plurality of pixels for either the first class or the second class, and wherein the single intensity threshold represents a minimization of intra-class intensity variance between the first and second class or a maximization of inter-class variance between the first class and the second class.

31. The method of claim 30, wherein the plurality of heuristic classifier comprises a second heuristic classifier that identifies local neighborhoods of pixels with the same class identified using the first heuristic classifier and applies a smoothed measure of maximum difference in intensity between pixels in the local neighborhood thereby causing the second heuristic classifier to cast a vote for each respective pixel in the plurality of pixels for either the first class or the second class.

32. The method of claim 31, wherein the plurality of heuristic classifier comprises a third heuristic classifier that performs edge detection on the plurality of pixels to form a plurality of edges in the image, morphologically closes the plurality of edges to form a plurality of morphologically closed regions in the image and assigns pixels in the morphologically closed regions to the first class and

pixels outside the morphologically closed regions to the second class, thereby causing the third heuristic classifier to cast a vote for each respective pixel in the plurality of pixels for either the first class or the second class.

33. The method of claim 32, wherein:

each respective pixel assigned by each of the heuristic classifiers in the plurality of classifiers to the second class is labelled as obvious second class, and

each respective pixel assigned by each of the plurality of heuristic classifiers as the first class is labelled as obvious first class.

34. The method of claim 33, wherein the segmentation algorithm is a GrabCut segmentation algorithm.

35. The method of claim 12, wherein the cleavage domain comprises a sequence recognized and cleaved by a uracil-DNA glycosylase and/or an endonuclease VIII.

36. The method of claim 1, wherein a capture probe plurality in the set of capture probe pluralities does not comprise a cleavage domain and each capture probe in the capture probe plurality is not cleaved from the substrate.

37. The method of claim 1, wherein the one or more analytes comprises DNA or RNA.

38. The method of claim 1, wherein the one or more analytes comprises a protein.

39. The method of claim 1, wherein each respective capture probe plurality in the set of capture probe pluralities is attached directly or attached indirectly to the substrate.

40. The method of claim 1, wherein the C) obtaining comprises *in-situ* sequencing of the set of capture spots on the substrate.

41. The method of claim 1, wherein the C) obtaining comprises high-throughput sequencing.

42. The method of claim 4, wherein a respective locus in the plurality of loci is biallelic and the

corresponding set of haplotypes for the respective locus consists of a first allele and a second allele.

43. The method of claim 42, wherein the respective locus includes a heterozygous single nucleotide polymorphism (SNP), a heterozygous insert, or a heterozygous deletion.

44. The method of any one of claims 1-43, wherein the plurality of sequence reads comprises 50,000 or more sequence reads, 100,000 or more sequence reads, or 1×10^6 or more sequence reads.

45. The method of claim 4, wherein the plurality of loci comprises between two and 100 loci, more than 10 loci, more than 100 loci, or more than 500 loci.

46. The method of claim 1, wherein the unique spatial barcode in the respective sequence read is localized to a contiguous set of nucleotides within the respective sequence read.

47. The method of claim 46, wherein the contiguous set of nucleotides is an N-mer, wherein N is an integer selected from the set {4, …, 20}.

48. The method of claim 4, the method further comprising retrieving the plurality of loci from a lookup table, file or data structure.

49. The method of claim 4, wherein the alignment algorithm is a local alignment that aligns the respective sequence read to a reference sequence using a scoring system that (i) penalizes a mismatch between a nucleotide in the respective sequence read and a corresponding nucleotide in the reference sequence in accordance with a substitution matrix and (ii) penalizes a gap introduced into an alignment of the sequence read and the reference sequence.

50. The method of claim 49, wherein the local alignment is a Smith-Waterman alignment.

51. The method of claim 49, wherein the reference sequence is all or portion of a reference genome.

52. The method of claim 4, the method further comprising removing from the plurality of sequence reads one or more sequence reads that do not overlay any loci in the plurality of loci.

53. The method of claim 52, wherein the plurality of sequence reads are RNA-sequence reads and wherein the removing comprises removing one or more sequences reads in the plurality of sequence reads that overlap a splice site in the reference sequence.

54. The method of claim 4, wherein the plurality of loci include one or more loci on a first chromosome and one or more loci on a second chromosome other than the first chromosome.

55. The method of any one of claims 1-54, wherein the plurality of sequence reads include 3'-end or 5'-end paired sequence reads.

56. The method of claim 1, wherein

the one or more analytes is a plurality of analytes,

a respective capture probe plurality in the set of capture probe pluralities includes a plurality of probes, each capture probe in the plurality of capture probes including a capture domain that is characterized by a single capture domain type configured to bind to each analyte in the plurality of analytes in an unbiased manner.

57. The method of claim 5, wherein the biological condition is absence or presence of a disease.

58. The method of claim 5, wherein the biological condition is a type of a cancer.

59. The method of claim 5, wherein the biological condition is a stage of a disease.

60. The method of claim 5, wherein the biological condition is a stage of a cancer.

61. The method of any one of claims 1-60, wherein the one or more images includes a brightfield image or a fluorescence image of the sample.

62. The method of any one of claims 1-60, wherein the one or more images is a plurality of images.

63. The method of claim 62, wherein:

a first image in the plurality of images is a brightfield image of the biological sample, and

a second image in the plurality of images is a fluorescence image of the biological sample.

64. The method of claim 66, wherein the fluorescence image is an immunofluorescence image.

65. The method of any one of claims 1-64, wherein the one or more images is a plurality of images and the plurality of images comprises two or more fluorescence images.

66. The method of any one of claims 1-65, wherein the representation of all or a portion of each subset of sequence reads at a respective position within the one or more images communicates a number of unique molecules that map to a particular analyte or combination of analytes in the sample represented by the subset of sequence reads that, in turn, map to the respective capture spot.

67. The method of claim 66, wherein the number of unique molecules that map to a particular analyte or combination of analytes in the sample represented by the subset of sequence reads that, in turn, map to the respective capture spot is communicated on a color scale or an intensity scale.

68. The method of any one of claims 1-67, wherein a respective capture probe plurality in the set of capture probe pluralities directly associates with an analyte from the sample.

69. The method of any one of claims 1-67, wherein a respective capture probe plurality in the set of capture probe pluralities indirectly associates with an analyte from the sample through an analyte capture agent.

70. A computer system comprising:

one or more processors;

memory; and

one or more programs, wherein the one or more programs are stored in the memory and configured to be executed by the one or more processors, the one or more programs for spatial analysis of analytes, the one or more programs including instructions for:

A) placing a sample on a substrate, wherein the substrate comprises a plurality of fiducial markers and a set of capture spots, wherein the set of capture spots comprises at least 1000 capture spots;

B) obtaining one or more images of the sample on the substrate, wherein each respective image of the one or more images comprises a corresponding plurality of pixels in the form of a

corresponding array of pixel values, and wherein the corresponding array of pixel values comprises at least 100,000 pixel values;

C) obtaining a plurality of sequence reads, in electronic form, from the set of capture spots after the A) placing, wherein:

each respective capture probe plurality in a set of capture probe pluralities is (i) at a different capture spot in the set of capture spots and (ii) directly or indirectly associates with one or more analytes from the sample,

each respective capture probe plurality in the set of capture probe pluralities is characterized by at least one unique spatial barcode in a plurality of spatial barcodes,

the plurality of sequence reads comprises sequence reads corresponding to all or portions of the one or more analytes,

the plurality of sequence reads comprises at least 10,000 sequence reads, and

each respective sequence read in the plurality of sequence reads includes a spatial barcode of the corresponding capture probe plurality in the set of capture probe pluralities or a complement thereof;

D) using all or a subset of the plurality of spatial barcodes to localize respective sequence reads in the plurality of sequence reads to corresponding capture spots in the set of capture spots, thereby dividing the plurality of sequence reads into a plurality of subsets of sequence reads, each respective subset of sequence reads corresponding to a different capture spot in the corresponding plurality of capture spots; and

E) using the plurality of fiducial markers to provide a composite representation comprising (i) the one or more images aligned to the set of capture spots on the substrate and (ii) a representation of all or a portion of each subset of sequence reads at each respective position within the one or more images that maps to the capture spot corresponding to the respective position of the one or more analytes in the sample by aligning a first image in the one or more images to the set of capture spots by a procedure that comprises:

analyzing the corresponding array of pixel values of the first image to identify a plurality of derived fiducial markers of the first image;

using a substrate identifier uniquely associated with the substrate to select a first template in a plurality of templates, wherein each template in the plurality of templates comprises reference positions for a corresponding plurality of reference fiducial markers and a corresponding coordinate system;

aligning the plurality of derived fiducial markers of the first image with the corresponding

plurality of reference fiducial markers of the first template using an alignment algorithm to obtain a transformation between the plurality of derived fiducial markers of the first image and the corresponding plurality of reference fiducial markers of the first template; and

using the transformation and the coordinate system of the first template to locate a corresponding position in the first image of each capture spot in the set of capture spots by running a plurality of heuristic classifiers on the plurality of pixels, wherein, for each pixel in the plurality of pixels, each respective heuristic classifier in the plurality of heuristic classifiers casts a vote for the respective pixel between a first class and the second class, thereby forming a corresponding aggregated score for each respective pixel in the plurality of pixels, and applying the aggregated score and intensity of each respective pixel in the plurality of pixels to a segmentation algorithm to independently assign a probability to each respective pixel in the plurality of pixels of being sample or background.

71. A computer readable storage medium storing one or more programs, the one or more programs comprising instructions, which when executed by an electronic device with one or more processors and a memory cause the electronic device to perform spatial analysis of analytes by a method comprising:

A) placing a sample on a substrate, wherein the substrate comprises a plurality of fiducial markers and a set of capture spots, wherein the set of capture spots comprises at least 1000 capture spots;

B) obtaining one or more images of the sample on the substrate, wherein each respective image of the one or more images comprises a corresponding plurality of pixels in the form of a corresponding array of pixel values, and wherein the corresponding array of pixel values comprises at least 100,000 pixel values;

C) obtaining a plurality of sequence reads, in electronic form, from the set of capture spots after the A) placing, wherein:

each respective capture probe plurality in a set of capture probe pluralities is (i) at a different capture spot in the set of capture spots and (ii) directly or indirectly associates with one or more analytes from the sample,

each respective capture probe plurality in the set of capture probe pluralities is characterized by at least one unique spatial barcode in a plurality of spatial barcodes,

the plurality of sequence reads comprises sequence reads corresponding to all or portions of the one or more analytes,

the plurality of sequence reads comprises at least 10,000 sequence reads, and

each respective sequence read in the plurality of sequence reads includes a spatial barcode of the corresponding capture probe plurality in the set of capture probe pluralities or a complement thereof;

D) using all or a subset of the plurality of spatial barcodes to localize respective sequence reads in the plurality of sequence reads to corresponding capture spots in the set of capture spots, thereby dividing the plurality of sequence reads into a plurality of subsets of sequence reads, each respective subset of sequence reads corresponding to a different capture spot in the corresponding plurality of capture spots; and

E) using the plurality of fiducial markers to provide a composite representation comprising (i) the one or more images aligned to the set of capture spots on the substrate and (ii) a representation of all or a portion of each subset of sequence reads at each respective position within the one or more images that maps to the capture spot corresponding to the respective position of the one or more analytes in the sample by aligning a first image in the one or more images to the set of capture spots by a procedure that comprises:

analyzing the corresponding array of pixel values of the first image to identify a plurality of derived fiducial markers of the first image;

using a substrate identifier uniquely associated with the substrate to select a first template in a plurality of templates, wherein each template in the plurality of templates comprises reference positions for a corresponding plurality of reference fiducial markers and a corresponding coordinate system;

aligning the plurality of derived fiducial markers of the first image with the corresponding plurality of reference fiducial markers of the first template using an alignment algorithm to obtain a transformation between the plurality of derived fiducial markers of the first image and the corresponding plurality of reference fiducial markers of the first template; and

using the transformation and the coordinate system of the first template to locate a corresponding position in the first image of each capture spot in the set of capture spots

by running a plurality of heuristic classifiers on the plurality of pixels, wherein, for each pixel in the plurality of pixels, each respective heuristic classifier in the plurality of heuristic classifiers casts a vote for the respective pixel between a first class and the second class, thereby forming a corresponding aggregated score for each respective pixel in the plurality of pixels, and applying the aggregated score and intensity of each respective pixel in the plurality of pixels to a segmentation algorithm to independently assign a probability to each respective pixel in the plurality of pixels of

being sample or background.