CLAIMS

1. A method for creating a transcriptome-wide expression profile, comprising:

(a) providing:

(i) a plurality of sample transcripts derived from a biological sample;

(ii) a plurality of centroid transcripts comprising at least a portion of the sample transcripts, the remaining sample transcripts being non-centroid transcripts;

(b) measuring the expression level of the plurality of centroid transcripts; and

(c) inferring the expression levels of the non-centroid transcripts from the centroid transcript expression levels, thereby creating a genome-wide expression profile.

2. A method according to Claim 1, in which the plurality of centroid transcripts is less than the plurality of sample transcripts.

3. A method according to Claim 1 or Claim 2, in which the plurality of centroid transcripts is provided by:

(a) performing computational analysis on a library of transcriptome-wide transcript expression data, such that a plurality of transcript clusters are created, in which the number of the clusters is less than the total number of transcripts in the library;

(b) identifying a centroid transcript within each of the transcript clusters thereby creating a plurality of centroid transcripts, the remaining transcripts being non-centroid transcripts;

(c) determining the ability of the measurements of the expression levels of the centroid transcripts to infer the levels of at least a portion of transcripts from the remaining transcripts, in which the portion is comprised of non-centroid transcripts; and

(d) selecting the centroid transcripts whose expression levels have the ability to infer the levels of the portion of non-centroid transcripts.

4. A method according to any preceding claim, in which expression levels are measured on a device selected from the group consisting of a microarray, a bead array, a liquid array, and a nucleic-acid sequencer.

5. A method according to Claim 4, comprising:

(a) performing steps (a) and (b) of Claim 3 on a first library of transcriptome-wide mRNA-expression data from a first collection of biological samples;

(b) measuring the expression levels of at least a portion of transcripts from a second collection of biological samples with the device, in which the portion of transcripts comprises transcripts identified as centroid transcripts from the first library;

(c) determining the ability of the measurements of the expression levels of the centroid transcripts to infer the levels of at least a portion of transcripts from the second library, in which the portion is comprised of non-centroid transcripts; and

(d) selecting the centroid transcripts whose expression levels have the ability to infer the levels of the portion of non-centroid transcripts.

6. A method according to any preceding claim, in which the plurality of centroid transcripts is approximately 1000 centroid transcripts.

7. A method according to any of Claims 3 to 6, in which the computational analysis comprises cluster analysis.

8. A method according to any preceding claim, comprising establishing a correlation between the expression levels of the centroid transcripts and the expression levels of the non-centroid transcripts.

9. A method according to Claim 3 or Claim 4, in which the method further comprises repeating the steps thereof until validated centroid transcripts for each of the plurality of transcript clusters are identified.

10. A method according to any preceding claim, in which the expression levels of non-centroid transcripts are inferred by:

(a) providing an algorithm capable of predicting the level of expression of transcripts within a first population which are not within a second population, the predicting based on the measured level of expression of transcripts within the second population;

(b) processing the plurality of sample transcripts under conditions such that a plurality of different templates representing only the plurality of centroid transcripts is created;

(c) measuring the amount of each of the different templates to create a plurality of measurements; and

(d) applying the algorithm to the plurality of measurements, thereby predicting the level of expression of transcripts within the plurality of sample transcripts which are not within the plurality of centroid transcripts.

11. A method according to Claim 10, in which the algorithm involves a dependency matrix.

12. A method according to any preceding claim, in which the transcriptome-wide expression profile identifies the biological sample as diseased or as healthy.

13. A method according to any preceding claim, in which the transcriptome -wide expression profile provides a functional readout of the action of a perturbagen.

14. A method according to any preceding claim, in which the transcriptome -wide expression profile comprises an expression profile suitable for use in a connectivity map.

15. A method according to any preceding claim, in which the transcriptome-wide expression profile is a genomic expression profile.

16. A method of assaying gene expression, comprising:

(a) providing:

(i) approximately 1000 different barcode sequences;

(ii) approximately 1000 beads, each bead comprising a homogeneous set of nucleic-acid probes, each set complementary to a different barcode sequence of the approximately 1000 barcode sequences;

(iii) a population of more than 1000 different transcripts, each transcript comprising a gene-specific sequence;

(iv) an algorithm capable of predicting the level of expression of unmeasured transcripts;

(b) processing the population of transcripts to create approximately 1000 different templates, each template comprising one of the approximately 1000 barcode sequences operably associated with a different gene-specific sequence, in which the approximately 1000 different templates represents less than the total number of transcripts within the population;

(c) measuring the amount of each of the approximately 1000 different templates to create a plurality of measurements; and

(d) applying the algorithm to the plurality of measurements, thereby predicting the level of expression of unmeasured transcripts within the population.

17. A method according to Claim 16, in which the method further comprises a device capable of measuring the amount of each of the approximately 1000 different templates.

18. A method according to Claim 16 or Claim 17, in which the beads are optically addressed.

19. A method according to any of Claims 16 to 18, in which the processing comprises ligation-mediated amplification.

20. A method according to Claim 18 or Claim 19, in which the measuring comprises detecting the optically addressed beads.

21. A method according to Claim 20, in which the measuring comprises hybridizing the approximately 1000 different templates to the approximately 1000 beads through the nucleic-acid probes complementary to the approximately 1000 barcode sequences.

22. A method according to any of Claim 18 to 20, in which the measuring comprises a flow cytometer.

23. A method according to any preceding claim, in which expression levels of a set of substantially invariant transcripts are additionally measured in the biological sample.

24. A method according to Claim 23, in which the expression levels of the centroid transcripts are normalized with respect to the expression levels of the invariant transcripts.

25. A kit, comprising:

(a) a first container comprising a plurality of centroid transcripts derived from a transcriptome;

(b) a second container comprising buffers and reagents compatible with measuring the expression level of the plurality of centroid transcripts within a biological sample; and

(c) a set of instructions for inferring the expression level of non-centroid transcripts within the biological sample, based upon the expression level of the plurality of centroid transcripts.

CLAIMS

1. A method for creating a transcriptome-wide expression profile, comprising:

(a) providing:

(i) a plurality of sample transcripts derived from a biological sample;

(ii) a plurality of centroid transcripts comprising at least a portion of the sample transcripts, the remaining sample transcripts being non-centroid transcripts;

(b) measuring the expression level of the plurality of centroid transcripts; and

(c) inferring the expression levels of the non-centroid transcripts from the centroid transcript expression levels, thereby creating a genome-wide expression profile.

2. A method according to Claim 1, in which the plurality of centroid transcripts is less than the plurality of sample transcripts.

3. A method according to Claim 1 or Claim 2, in which the plurality of centroid transcripts is provided by:

(a) performing computational analysis on a library of transcriptome-wide transcript expression data, such that a plurality of transcript clusters are created, in which the number of the clusters is less than the total number of transcripts in the library;

(b) identifying a centroid transcript within each of the transcript clusters thereby creating a plurality of centroid transcripts, the remaining transcripts being non-centroid transcripts;

(c) determining the ability of the measurements of the expression levels of the centroid transcripts to infer the levels of at least a portion of transcripts from the remaining transcripts, in which the portion is comprised of non-centroid transcripts; and

(d) selecting the centroid transcripts whose expression levels have the ability to infer the levels of the portion of non-centroid transcripts.

4. A method according to any preceding claim, in which expression levels are measured on a device selected from the group consisting of a microarray, a bead array, a liquid array, and a nucleic-acid sequencer.

5. A method according to Claim 4, comprising:

(a) performing steps (a) and (b) of Claim 3 on a first library of transcriptome-wide mRNA-expression data from a first collection of biological samples;

(b) measuring the expression levels of at least a portion of transcripts from a second collection of biological samples with the device, in which the portion of transcripts comprises transcripts identified as centroid transcripts from the first library;

(c) determining the ability of the measurements of the expression levels of the centroid transcripts to infer the levels of at least a portion of transcripts from the second library, in which the portion is comprised of non-centroid transcripts; and

(d) selecting the centroid transcripts whose expression levels have the ability to infer the levels of the portion of non-centroid transcripts.

6. A method according to any preceding claim, in which the plurality of centroid transcripts is approximately 1000 centroid transcripts.

7. A method according to any of Claims 3 to 6, in which the computational analysis comprises cluster analysis.

8. A method according to any preceding claim, comprising establishing a correlation between the expression levels of the centroid transcripts and the expression levels of the non-centroid transcripts.

9. A method according to Claim 3 or Claim 4, in which the method further comprises repeating the steps thereof until validated centroid transcripts for each of the plurality of transcript clusters are identified.

10. A method according to any preceding claim, in which the expression levels of non-centroid transcripts are inferred by:

(a) providing an algorithm capable of predicting the level of expression of transcripts within a first population which are not within a second population, the predicting based on the measured level of expression of transcripts within the second population;

(b) processing the plurality of sample transcripts under conditions such that a plurality of different templates representing only the plurality of centroid transcripts is created;

(c) measuring the amount of each of the different templates to create a plurality of measurements; and

(d) applying the algorithm to the plurality of measurements, thereby predicting the level of expression of transcripts within the plurality of sample transcripts which are not within the plurality of centroid transcripts.

11. A method according to Claim 10, in which the algorithm involves a dependency matrix.

12. A method according to any preceding claim, in which the transcriptome-wide expression profile identifies the biological sample as diseased or as healthy.

13. A method according to any preceding claim, in which the transcriptome -wide expression profile provides a functional readout of the action of a perturbagen.

14. A method according to any preceding claim, in which the transcriptome -wide expression profile comprises an expression profile suitable for use in a connectivity map.

15. A method according to any preceding claim, in which the transcriptome-wide expression profile is a genomic expression profile.

16. A method according to any preceding claim, in which expression levels of a set of substantially invariant transcripts are additionally measured in the biological sample.

17. A method according to Claim 16, in which the expression levels of the centroid transcripts are normalized with respect to the expression levels of the invariant transcripts.

18. A kit, comprising:

(a) a first container comprising a plurality of centroid transcripts derived from a transcriptome;

(b) a second container comprising buffers and reagents compatible with measuring the expression level of the plurality of centroid transcripts within a biological sample; and

(c) a set of instructions for inferring the expression level of non-centroid transcripts within the biological sample, based upon the expression level of the plurality of centroid transcripts.