human. Cancer cell line collections such as breast cancer have been shown to retain their subtype characteristics corresponding to those of the *in vivo* counterparts and these cell lines have been demonstrated as powerful tools to model heterogeneity in cancer *in vitro*.

[0020] It has been shown that the gene expression profiles and genomic signatures of NCI60 have been further utilised to identify phenotype-specific drugs. The gene expression signature of NCI60 incorporated with the drug sensitivity results from over 40,000 compound screens have effectively identified targets not only with selectivity to the RAS and PI3K pathways but also with disease specificity to breast cancer subtypes (basal vs. luminal). This shows that oncogenomic data derived from a mixed assembly of cancer cell lines can be robust to provide valid therapeutic leads which are disease or phenotype specific, which further supports the possibility to establish a diagnostic-therapeutic framework incorporating both genome-scale data for cancer subtype identification and experimental models for therapeutic target discovery.

[0021] It is an object of the present invention to facilitate the development of better prognostic and therapeutic strategies which will benefit cancer patients with novel treatment options and to improve the overall survival.

SUMMARY OF THE INVENTION

[0022] In a first aspect the present invention is directed to a method of generating a scheme allowing classification of a cancer of an individual patient for estimating a clinical outcome for said patient, wherein the method comprises:

a) identifying subtype clusters of a cancer type based on comparison of a set of gene expression data derived from clinical cancer samples of multiple test subjects suffering from said cancer type; wherein clinical samples are divided into subtype clusters of said cancer type based on the similarity of their gene expression data;

b) correlating the subtype clusters identified under a) to a specific clinical outcome of cancer;

c) determining an epithelial-mesenchymal transition (EMT) ~~score~~signature, comprising:

i) determining ~~the~~ a quantitative expression of at least two markers characteristics for each subtype clusters identified under a), wherein each of these markers is either characteristic for a mesenchymal cell phenotype or for a epithelial cell phenotype,

ii) subjecting the results of the quantitative expression of the at least two markers to a regression analysis ~~for~~ to thereby determin~~e~~ing said epithelial-mesenchymal transition (EMT) signature;

~~ii) determining the quantitative expression of the at least two markers in said patient sample and comparing it with the epithelial-mesenchymal transition (EMT) signature to assign an epithelial-mesenchymal transition (EMT) score for said patient sample;~~

iii) correlating the epithelial mesenchymal transition (EMT) signature to the specific clinical outcome of cancer identified in b);

d) estimating ~~an~~ the outcome of a patient suffering from said cancer by :

1. assigning the set of gene expression data of said patient sample to any one of the subtype clusters according to a) and b), and

2. ~~by determining~~ assigning the patients epithelial-mesenchymal transition (EMT) score by determining the quantitative expression of the at least two markers in said patient sample and comparing it with the epithelial-mesenchymal transition (EMT) signature under c) to assign an epithelial-mesenchymal transition (EMT) score for said patient sample to correlate the EMT score to the specific clinical outcome of cancer.

[0023] In a second aspect, the present invention refers to a method of estimating a clinical outcome of a patient suffering from epithelial ovarian cancer (EOC) wherein the method comprises:

a) providing a set of expression data in a patient sample obtained from a patient suffering from epithelial ovarian cancer (EOC);

b) assigning the set of expression data derived from said patient sample to one of five subtype clusters for epithelial ovarian cancer (EOC),

wherein the first of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 – Epi A;

wherein the second of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - EpiB;

wherein the third of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - Mes;

wherein the fourth of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - StemA;

wherein the fifth of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - StemB;

c) determining an EMT score for the patient sample by; ~~and~~

i) computing an enrichment score by integration of the difference between the empirical cumulative distribution functions of genes from Gene List 3 and genes not in Gene List 3 for each set of expression data from the individual patient samples to determine the epithelial rank or mesenchymal rank of a sample;

ii) determining the EMT score by substracting the value for the epithelial rank from the value for the mesenchymal rank

d) estimating the clinical outcome depending on any one of the five subtype clusters and the EMT score to which the patient sample is assigned.

[0024] In a third aspect, the present invention refers to a method of determining whether the epithelial mesenchymal score of a patient suffering from a cancer can be changed by administering an EMT reversal agent to increase patients susceptibility for an anti-cancer treatment, wherein the method comprises classifying a tumor sample of said patient using a method of the present invention to determine the cancer subtype and EMT score; after determining the cancer subtype and EMT score determining whether a treatment with an EMT reversal agent is necessary to sensitize said patient to anti-cancer treatment.

[0025] In a fourth aspect the present invention refers to a cluster characterized in that it comprises at least one subtype cluster selected from the group consisting of subtype clusters shown in Gene List

2 - Epi A, subtype cluster shown in Gene List 2 Epi-B, subtype cluster shown in Gene List 2 – Mes, subtype cluster shown in Gene List 2 – Stem A and subtype cluster shown in Gene List 2 –Stem B.

[0026] In a fifth aspect, the present invention refers to a computer readable medium having stored therein a computer program comprising a set of executable instructions, when executed by a computer processor, controls the processor to perform the method of the present invention.

CLAIMS

1. A method of generating a scheme allowing classification of a cancer of an individual patient for estimating a clinical outcome for said patient, wherein the method comprises:

a) identifying subtype clusters of a cancer type based on comparison of a set of gene expression data derived from clinical cancer samples of multiple test subjects suffering from said cancer type; wherein clinical samples are divided into subtype clusters of said cancer type based on the similarity of their gene expression data;

b) correlating the subtype clusters identified under a) to a specific clinical outcome of cancer;

c) determining an epithelial-mesenchymal transition (EMT) score, comprising:

i) determining the quantitative expression of at least two markers wherein each of these markers is either characteristic for a mesenchymal cell phenotype or for a epithelial cell phenotype, for determining said epithelial-mesenchymal transition (EMT) signature;

ii) determining the quantitative expression of the at least two markers in said patient sample and comparing it with the epithelial-mesenchymal transition (EMT) signature to assign an epithelial-mesenchymal transition (EMT) score for said patient sample;

d) estimating an outcome of a patient suffering from said cancer by assigning the set of gene expression data of said patient sample to any one of the subtype clusters according to a) and b), and by determining the patients epithelial-mesenchymal transition (EMT) score.

2. The method of claim 1, wherein identifying the subtype clusters according to a) comprises:

a') selecting a subset of genes which are most variably expressed across all set of gene expression data;

b') subjecting the most variably expressed genes determined under a') to clustering to identify said subtypes of said cancer type.

3. The method of claim 2, wherein a gene is considered most variably expressed across all sets of gene expression data of all clinical samples referred to under a') if the standard deviation of the gene expression is higher across all sets of gene expression data in comparison to the standard deviation of any other gene.

4. The method of claim 2, wherein a gene is considered most variably expressed across all sets of gene expression data of all clinical samples referred to under a') if said gene has a higher variability in expression among all sets of expression data obtained from the clinical samples.

5. The method of claim 3 or 4, wherein the most 500 or 600 or 700 or 800 or 900 or 1000 variably expressed genes across all sets of gene expression data will be selected.

6. The method of claim 2, wherein the clustering algorithm for clustering according to b') is a hierarchical clustering with agglomerative linkage.

7. The method of claim 2, wherein the clustering algorithm for consensus clustering according to b') is a hierarchical clustering with agglomerative linkage with Eucledian distance and with a sub-sampling ratio of between about 0.5 to 0.9, preferably 0.8, for 1000 iterations.

8. The method of any one of claims 1 to 7, wherein assigning the set of gene expression data of said patient sample to a subtype cluster referred to under d) is carried out by

- selecting a set of gene expression data from test subjects which correlate the best with the cancer phenotype assigned to each of the subtype clusters under b) for a given subtype cluster;
- subdividing said set of gene expression data of a clinical sample selected from all subtype clusters equally into at least two training groups;
- determining gene expression signatures within the at least two training groups which are characteristic for each of the subtype clusters previously determined; and
- comparing the set of gene expression data of a patient sample with the expression signatures for determining to which subtype cluster the patient sample is to be assigned.

9. The method of claim 8, wherein an at least 50% or 60% or 70% or 75% match of the gene expression data of the patient sample with the expression signature of one of the subtype clusters assigns the patient sample to the respective subtype cluster.

10. The method of claim 8, wherein silhouette analysis is used to select sets of gene expression data from test subjects which correlate the best with the cancer phenotype assigned to each of the subtype clusters for a given subtype cluster.

11. The method of claim 10, wherein expression profiles having the highest silhouette values are selected from all subtype clusters to be subdivided into the at least two training groups.

12. The method of claim 8, wherein a regression model is used to determine gene expression signatures within one or both of the at least two training groups which are characteristic for each of the subtype clusters previously determined.

13. The method of claim 8, wherein a Bayesian probit regression model is used to assign the patient sample to one of the subtype clusters.

14. The method of claim any one of claims 1 to 13, wherein the epithelial-mesenchymal transition (EMT) signature is determined by subjecting the results of the quantitative expression of the at least two markers characteristic for each subtype cluster identified under a) to a regression analysis.

15. The method of any one of claims 1 to 14, wherein correlating the subtype clusters identified under a) to a specific clinical outcome of cancer is carried out by comparing the single samples assigned to any one of the subtype clusters with the clinical development of the test subject from which the respective set of gene expression data is obtained.

16. The method of any one of claims 1 to 15, wherein the cancer is selected from the group consisting of epithelial ovarian cancer (EOC), breast carcinoma, hepatocarcinoma, pancreatic carcinoma, glioblastoma multiforme (GBM), lung carcinoma, colorectal cancer and bladder carcinoma.

17. The method of any one of claims 1 to 15, wherein the cancer is a solid tumor type cancer.

18. A method of estimating a clinical outcome of a patient suffering from epithelial ovarian cancer (EOC) wherein the method comprises:

a) providing a set of expression data in a patient sample obtained from a patient suffering from epithelial ovarian cancer (EOC);

b) assigning the set of expression data derived from said patient sample to one of five subtype clusters for epithelial ovarian cancer (EOC),

wherein the first of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 – Epi A;

wherein the second of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - EpiB;

wherein the third of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - Mes;

wherein the fourth of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - StemA;

wherein the fifth of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - StemB;

c) determining an EMT score for the patient sample; and

d) estimating the clinical outcome depending on any one of the five subtype clusters and the EMT score to which the patient sample is assigned.

19. The method of claim 18, wherein assigning the set of expression data derived from said patient sample to one of five subtype clusters for epithelial ovarian cancer (EOC) is carried out either

a') by clustering the expression data derived from the patient sample together with the expression data which make up said five different subtype clusters of epithelial ovarian cancer (EOC) to determine to which group the expression data of the patient sample belongs; or

b') by subjecting the expression data obtained from the patient sample together with the expression data which make up said five different subtype clusters of epithelial ovarian cancer (EOC) to a regression analysis.

20. The method of claim 18 or 19, wherein the EMT score referred to under c) is determined by:

e) computing an enrichment score by integration of the difference between the empirical cumulative distribution functions of genes from Gene List 3 and genes not in Gene List 3 for each set of expression data from the individual patient samples to determine the epithelial rank or mesenchymal rank of a sample;

f) determining the EMT score by substracting the value for the epithelial rank from the value for the mesenchymal rank.

21. The method of claim 18 or 19, wherein the EMT score referred to under c) for a patient cancer sample is determined by:

g) computing an enrichment score by integration of the difference between the empirical cumulative distribution functions of genes from Gene List 4 and genes not in Gene List 4 for each set of expression data from the individual patient cancer sample to determine the epithelial rank or mesenchymal rank of a sample;

h) determining an EMT score by substracting the value for the epithelial rank from the value for the mesenchymal rank.

22. The method of claim 20 or 21, wherein integration of the difference between the empirical cumulative distribution functions of genes from Gene List 3 or Gene List 4 and genes not in Gene List 3 or Gene List 4, respectively, for each set of expression data from the individual patient cancer sample to determine the epithelial rank or mesenchymal rank of a sample is carried out using single sample enrichment analysis.

23. A method of determining whether the epithelial mesenchymal score of a patient suffering from a cancer can be changed by administering an EMT reversal agent to increase patients susceptibility for an anti-cancer treatment, wherein the method comprises classifying a tumor sample of said patient using a method of any one of claims 1 to 22 to determine the cancer subtype and EMT score; after determining the cancer subtype and EMT score determining whether a treatment with an EMT reversal agent is necessary to sensitize said patient to anti-cancer treatment.

24. The method of claim 23, wherein the anti-cancer treatment is any one of administering an anti-cancer drug, surgery, radiation therapy, chemotherapy, or hormonal therapy.

25. A cluster characterized in that it comprises at least one subtype cluster selected from the group consisting of subtype clusters shown in Gene List 2 - Epi A, subtype cluster shown in Gene List 2 Epi-B, subtype cluster shown in Gene List 2 – Mes, subtype cluster shown in Gene List 2 – Stem A and subtype cluster shown in Gene List 2 –Stem B.

26. A computer readable medium having stored therein a computer program comprising a set of executable instructions, when executed by a computer processor, controls the processor to perform the method according to any one of claims 1 to 17 or 18 to 22.

27. A computer program comprising a set of executable instructions, when executed by a computer processor, controls the processor to perform the method according to any one of claims 1 to 17 or 18 to 22.

CLAIMS

1. A method of generating a scheme allowing classification of a cancer of an individual patient for estimating a clinical outcome for said patient, wherein the method comprises:

a) identifying subtype clusters of a cancer type based on comparison of a set of gene expression data derived from clinical cancer samples of multiple test subjects suffering from said cancer type; wherein clinical samples are divided into subtype clusters of said cancer type based on the similarity of their gene expression data;

b) correlating the subtype clusters identified under a) to a specific clinical outcome of cancer;

c) determining an epithelial-mesenchymal transition (EMT) signature, comprising:

i) determining a quantitative expression of at least two markers characteristics for each subtype clusters identified under a), wherein each of these markers is either characteristic for a mesenchymal cell phenotype or for a epithelial cell phenotype,

ii) subjecting the results of the quantitative expression of the at least two markers to a regression analysis to thereby determine said epithelial-mesenchymal transition (EMT) signature,

iii) correlating the epithelial mesenchymal transition (EMT) signature to the specific clinical outcome of cancer identified in b);

and

d) estimating the outcome of a patient suffering from said cancer by:

1. assigning the set of gene expression data of said patient sample to any one of the subtype clusters according to a) and b), and
2. assigning the patients epithelial-mesenchymal transition (EMT) score by determining the quantitative expression of the at least two markers in said patient sample and comparing it with the epithelial-mesenchymal transition (EMT) signature under c) to assign an epithelial-mesenchymal transition (EMT) score for said patient sample to correlate the EMT score to the specific clinical outcome of cancer.

2. The method of claim 1, wherein identifying the subtype clusters according to a) comprises:

a') selecting a subset of genes which are most variably expressed across all set of gene expression data;

b') subjecting the most variably expressed genes determined under a') to clustering to identify said subtypes of said cancer type.

3. The method of claim 2, wherein a gene is considered most variably expressed across all sets of gene expression data of all clinical samples referred to under a') if the standard deviation of the gene expression is higher across all sets of gene expression data in comparison to the standard deviation of any other gene.

4. The method of claim 2, wherein a gene is considered most variably expressed across all sets of gene expression data of all clinical samples referred to under a') if said gene has a higher variability in expression among all sets of expression data obtained from the clinical samples.

5. The method of claim 3 or 4, wherein the most 500 or 600 or 700 or 800 or 900 or 1000 variably expressed genes across all sets of gene expression data will be selected.

6. The method of claim 2, wherein the clustering algorithm for clustering according to b') is a hierarchical clustering with agglomerative linkage.

7. The method of claim 2, wherein the clustering algorithm for consensus clustering according to b') is a hierarchical clustering with agglomerative linkage with Eucledian distance and with a sub-sampling ratio of between about 0.5 to 0.9, preferably 0.8, for 1000 iterations.

8. The method of any one of claims 1 to 7, wherein assigning the set of gene expression data of said patient sample to a subtype cluster referred to under d) is carried out by

- selecting a set of gene expression data from test subjects which correlate the best with the cancer phenotype assigned to each of the subtype clusters under b) for a given subtype cluster;
- subdividing said set of gene expression data of a clinical sample selected from all subtype clusters equally into at least two training groups;
- determining gene expression signatures within the at least two training groups which are characteristic for each of the subtype clusters previously determined; and

- comparing the set of gene expression data of a patient sample with the expression signatures for determining to which subtype cluster the patient sample is to be assigned.

9. The method of claim 8, wherein an at least 50% or 60% or 70% or 75% match of the gene expression data of the patient sample with the expression signature of one of the subtype clusters assigns the patient sample to the respective subtype cluster.

10. The method of claim 8, wherein silhouette analysis is used to select sets of gene expression data from test subjects which correlate the best with the cancer phenotype assigned to each of the subtype clusters for a given subtype cluster.

11. The method of claim 10, wherein expression profiles having the highest silhouette values are selected from all subtype clusters to be subdivided into the at least two training groups.

12. The method of claim 8, wherein a regression model is used to determine gene expression signatures within one or both of the at least two training groups which are characteristic for each of the subtype clusters previously determined.

13. The method of claim 8, wherein a Bayesian probit regression model is used to assign the patient sample to one of the subtype clusters.

14. The method of any one of claims 1 to 13, wherein correlating the subtype clusters identified under a) to a specific clinical outcome of cancer is carried out by comparing the single samples assigned to any one of the subtype clusters with the clinical development of the test subject from which the respective set of gene expression data is obtained.

15. The method of any one of claims 1 to 14, wherein the cancer is selected from the group consisting of epithelial ovarian cancer (EOC), breast carcinoma, hepatocarcinoma, pancreatic carcinoma, glioblastoma multiforme (GBM), lung carcinoma, colorectal cancer and bladder carcinoma.

16. The method of any one of claims 1 to 14, wherein the cancer is a solid tumor type cancer.

17. A method of estimating a clinical outcome of a patient suffering from epithelial ovarian cancer (EOC) wherein the method comprises:

a) providing a set of expression data in a patient sample obtained from a patient suffering from epithelial ovarian cancer (EOC);

b) assigning the set of expression data derived from said patient sample to one of five subtype clusters for epithelial ovarian cancer (EOC),

wherein the first of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 – Epi A;

wherein the second of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - EpiB;

wherein the third of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - Mes;

wherein the fourth of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - StemA;

wherein the fifth of the five subtype clusters of epithelial ovarian cancer (EOC) is characterized by the genes referred to in Gene List 2 - StemB;

c) determining an EMT score for the patient sample by

i) computing an enrichment score by integration of the difference between the empirical cumulative distribution functions of genes from Gene List 3 and genes not in Gene List 3 for each set of expression data from the individual patient samples to determine the epithelial rank or mesenchymal rank of a sample;

ii) determining the EMT score by substracting the value for the epithelial rank from the value for the mesenchymal rank

d) estimating the clinical outcome depending on any one of the five subtype clusters and the EMT score to which the patient sample is assigned.

18. The method of claim 17, wherein assigning the set of expression data derived from said patient sample to one of five subtype clusters for epithelial ovarian cancer (EOC) is carried out either

a') by clustering the expression data derived from the patient sample together with the expression data which make up said five different subtype clusters of epithelial ovarian cancer (EOC) to determine to which group the expression data of the patient sample belongs; or

b') by subjecting the expression data obtained from the patient sample together with the expression data which make up said five different subtype clusters of epithelial ovarian cancer (EOC) to a regression analysis.

19. The method of claim 17 or 18, wherein the EMT score referred to under c) for a patient cancer sample is determined by:
g) computing an enrichment score by integration of the difference between the empirical cumulative distribution functions of genes from Gene List 4 and genes not in Gene List 4 for each set of expression data from the individual patient cancer sample to determine the epithelial rank or mesenchymal rank of a sample;
h) determining an EMT score by substracting the value for the epithelial rank from the value for the mesenchymal rank.

20. The method of claim 18 or 19, wherein integration of the difference between the empirical cumulative distribution functions of genes from Gene List 3 or Gene List 4 and genes not in Gene List 3 or Gene List 4, respectively, for each set of expression data from the individual patient cancer sample to determine the epithelial rank or mesenchymal rank of a sample is carried out using single sample enrichment analysis.

21. A method of determining whether the epithelial mesenchymal score of a patient suffering from a cancer can be changed by administering an EMT reversal agent to increase patients susceptibility for an anti-cancer treatment, wherein the method comprises classifying a tumor sample of said patient using a method of any one of claims 1 to 20 to determine the cancer subtype and EMT score; after determining the cancer subtype and EMT score determining whether a treatment with an EMT reversal agent is necessary to sensitize said patient to anti-cancer treatment.

22. The method of claim 21, wherein the anti-cancer treatment is any one of administering an anti-cancer drug, surgery, radiation therapy, chemotherapy, or hormonal therapy.

23. A cluster characterized in that it comprises at least one subtype cluster selected from the group consisting of subtype clusters shown in Gene List 2 - Epi A, subtype cluster shown in Gene List 2 Epi-B, subtype cluster shown in Gene List 2 – Mes, subtype cluster shown in Gene List 2 – Stem A and subtype cluster shown in Gene List 2 –Stem B.

24. A computer readable medium having stored therein a computer program comprising a set of executable instructions, when executed by a computer processor, controls the processor to perform the method according to any one of claims 1 to 16 or 17 to 20.

25. A computer program comprising a set of executable instructions, when executed by a computer processor, controls the processor to perform the method according to any one of claims 1 to 17 or 18 to 22.