

and gene expression profiles. Furthermore, above methods cannot be applied retrospectively to existing large-scale cancer genomics bulk tumour data, representing a vast and mostly unexplored resource for studying cross-talk in the tumour microenvironment.

[0004] One major branch of oncology drug development is focusing on development of antibodies (or antibody-conjugated-drugs) that specifically target antigens/proteins inside or on the surfaces of cancer cells. It is therefore critical in the early phases of drug development to have access to accurate molecular profiles of cancer cells. While experimental models (cell lines and animal models) can provide an approximation, such models are generally limited by how cancer cells rapidly adapt physiology outside their natural environment. For example, EGFR expression (and EGFR gene copies) in glioblastoma cancer cells is greatly reduced immediately upon culturing of the cancer cells *in vitro*.

[0005] Cancer cell gene expression can currently also be estimated with single cell profiling or laser micro-dissection. However, these approaches have limitations: the molecular profiles are biased after cell disassociation, the techniques require lots of work and are expensive, they cannot easily separate, for example, non-malignant from malignant (cancer) epithelial cells, and they cannot readily be applied to standard frozen or Formalin-Fixed Paraffin-Embedded (FFPE) tumour samples, nor are these methods scalable.

[0006] There is therefore a need for technologies that allow high-throughput profiling of cancer cells *ex vivo*. Furthermore, other desirable features and characteristics will become apparent from the subsequent detailed description and the appended claims, taken in conjunction with the accompanying drawings and this background of the disclosure.

SUMMARY OF THE INVENTION

[0007] In one aspect, the present invention refers to a method of predicting expression profiles of cancerous and non-cancerous cells, respectively, based on multiple sets of expression profiles, wherein each set of the multiple sets of expression profiles is obtained from tumour-derived samples comprising a mixture of cancerous and non-cancerous cells of one tumour type, wherein the method comprises: a. determining tumour purity values for the one or more tumour-derived samples; b. providing different sets of expression profiles, wherein the sets of expression profiles comprise combined expression data for multiple or all molecules expressed by cancerous and non-cancerous cells comprised in the one or more tumour-derived samples; c. deconvoluting each combined expression data referred under b. by extrapolating expression profiles of the multiple or all molecules expressed in the different tumour samples with different tumour purity values to a tumour purity value at least substantially equal to 1 or 0; thereby predicting the

expression profiles of the cancerous and non-cancerous cells respectively from the sets of expression profiles; wherein the tumour purity value is a mean tumour purity value.

BRIEF DESCRIPTION OF THE DRAWINGS

- 5 [0008] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:
- [0009] FIG. 1 shows an illustration comparing conventional clinical sequencing and TUMERIC-solo sequencing in accordance with a present embodiment.
- 10 [0010] FIG. 2 shows an overview illustration of the TUMERIC sequencing process in accordance with the present embodiment.
- [0011] FIG. 3 shows a flow diagram of the overall TUMERIC-solo process in accordance with the present embodiment.
- [0012] FIG. 4 shows a flow diagram of the TUMERIC-solo tumour purity estimation process
15 of FIG. 3 in accordance with the present embodiment.
- [0013] FIG. 5 shows a flow diagram of the TUMERIC-solo transcriptome deconvolution of FIG. 3 in accordance with the present embodiment.
- [0014] FIG. 6 shows a working example of tumour transcriptome deconvolution in accordance with the present embodiment wherein FIG. 6a shows estimated tumour purity values
20 for around 8000 bulk tumour samples across 20 solid tumour types; FIG. 6b shows genes specifically expressed in cancer and stromal cells across cancer types; as expected, only cancer-cell specific genes are affected by DNA copy number alterations in the corresponding tumours; FIG. 6c shows the inferred cancer and stroma compartment expression levels for 280 known stromal-specific genes; FIG. 6d shows the inferred cancer and stroma compartment expression
25 levels in melanoma (skin cutaneous melanoma - SKCM), as well as bulk tumour measurements, for cancer and stroma specific genes previously identified with melanoma tumour single cell RNA sequencing (scRNA-sequencing); FIG. 6e shows genes and pathways which are ordered by inferred expression difference between cancer and stroma compartments in each tumour type; FIG. 6f shows protein expression inferred for cancer and stroma compartments in (OV) and
30 breast (BRCA) cancer cohorts using iTRAQ protein quantification data and compared to RNA sequencing data from the same tumors; and FIG. 6g shows genes with highly variable cancer vs. stroma mRNA expression differences across cancer types as identified, where immunohistochemistry (IHC) staining data was compared to deconvoluted RNA sequencing data for the gene (S100A6) with highest mRNA abundance.

CLAIMS

1. A method of predicting expression profiles of cancerous and non-cancerous cells, respectively, based on multiple sets of expression profiles, wherein each set of the multiple sets of expression profiles is obtained from tumour-derived samples comprising a mixture of cancerous and non-cancerous cells of one tumour type, wherein the method comprises:

 - a. determining tumour purity values for the one or more tumour-derived samples;
 - b. providing different sets of expression profiles, wherein the sets of expression profiles comprise combined expression data for multiple or all molecules expressed by cancerous and non-cancerous cells comprised in the one or more tumour-derived samples;
 - c. deconvoluting each combined expression data referred under b. by extrapolating expression profiles of the multiple or all molecules expressed in the different tumour samples with different tumour purity values to a tumour purity value at least substantially equal to 1 or 0; thereby predicting the expression profiles of the cancerous and non-cancerous cells respectively from the sets of expression profiles; wherein the tumour purity value is a mean tumour purity value.
2. The method of claim 1, wherein the tumour-derived sample is obtained from a single subject.
3. The method of claim 2, wherein the tumour-derived sample is partitioned into 2 or more sections, and wherein one set of expression profiles is generated for each section.
4. The method of any one of claims 1 to 3, wherein providing different sets of expression profile comprises ~~the~~ use of existing data sets of expression profiles.
5. The method of claim 4, wherein the existing data sets of expression profiles are from TCGA and ICGC databases.
6. The method of any one of the preceding claims, wherein the tumour type is selected from the group consisting of BLCA (Bladder Urothelial Carcinoma), BRCA (Breast Invasive Carcinoma), CESC (Cervical Squamous Cell Carcinoma), CRC (Colon and Rectum

Adenocarcinoma) (COAD (Colon adenocarcinoma) and READ (Rectum adenocarcinoma) combined), ESCA (Esophageal Carcinoma), GBM (Glioblastoma Multiforme), HNSC (Head and Neck Squamous Cell Carcinoma), KIRC (Kidney Renal Clear Cell Carcinoma), KIRP (Kidney Renal Papillary Cell Carcinoma), LGG (Brain Lower Grade Glioma), LIHC (Liver Hepatocellular Carcinoma), LUAD (Lung Adenocarcinoma), LUSC (Lung Squamous Cell Carcinoma), OV (Ovarian Serous Cystadenocarcinoma), PAAD (Pancreatic Adenocarcinoma), PRAD (Prostate Adenocarcinoma), SKCM (Skin Cutaneous Melanoma), STAD (Stomach Adenocarcinoma), THCA (Thyroid Carcinoma) and UCEC (Uterine Corpus Endometrial Carcinoma).

7. The method of any one of the preceding claims, wherein the expression profiles are selected from the group consisting of gene expression, RNA expression, epigenetic expression, protein expression, proteomic expression, and combinations thereof, for example, RNA and epigenetic expression, and RNA and protein expression.

8. The method of any one of the preceding claims, wherein the method for determining tumour purity is selected from the group consisting of distribution of somatic DNA variant allele frequencies, somatic DNA copy number alteration amplitudes, germline B-allele frequencies, gene expression signatures or patterns, protein expression signatures or patterns, and DNA methylation signatures or patterns, and combinations thereof.

9. The method of claim 8, wherein at least two, or at least three, or at least four, or at least five, or two, or three, or four, or five or all of the methods of claim 8 are used together to determine mean tumour purity.

~~10. The method of any one of claims 1 to 8, wherein tumour purity value is a mean tumour purity value.~~

~~11.~~ 10. The method of claim 1, further comprising scoring molecules of step c. based on the level of up-regulation or down-regulation in cancer tissue versus stromal tissue; and/or scoring molecules of step c. based on the level of up-regulation or down-regulation in cancer tissue versus healthy tissue.

~~12~~.11. The method of claim ~~11~~10, further comprising assigning the up- and down-regulated molecules to genes or transcript isoforms of known data sets of membrane associated proteins or receptors; and/or assigning the up- and down-regulated molecules to genes or transcript isoforms of known data sets of HLA-binding peptides and T-cell antigen binding peptides.

~~13~~.12. The method of claim ~~12~~11, wherein the known data sets for assigning genes or transcript isoforms originates from Gene Ontology and/or TANTIGEN.

~~14~~.13. The method of claim 11 and 12~~and 13~~, further comprising selecting genes or transcript isoforms for antibody based therapy and / or T-cell based therapy.

~~15~~.14. The method of any one of ~~the preceding claims~~claims 11 to 13, wherein the gene or transcript isoform is a membrane associated protein, membrane associated receptor, antigen peptide, target protein, peptide, and/or is targetable by an antibody.

~~16~~.15. The method of any one of the preceding claims, wherein the molecules are selected from the group of gene, DNA, RNA or protein molecules, or combinations thereof.